

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Han, Bingyi

Title:

Ethics of Artificial Intelligence for Analysing Student Learning Behaviour: A Multi-Stakeholder Perspective

Date:

2025

Persistent Link:

<https://hdl.handle.net/11343/361195>

Terms and Conditions:

Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.

ETHICS OF ARTIFICIAL INTELLIGENCE FOR ANALYSING STUDENT LEARNING BEHAVIOUR: A MULTI-STAKEHOLDER PERSPECTIVE

Bingyi Han

<https://orcid.org/0000-0001-9044-1755>

Submitted in total fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Computing and Information Systems
Faculty of Engineering and Information Technology
THE UNIVERSITY OF MELBOURNE

March 2025

Abstract

The integration of Artificial Intelligence into Education (AIEd) promises enhanced personalisation, educational efficiency, and learning outcomes, yet raises complex ethical challenges that warrant scrutiny. One important challenge is the growing use of AI sensing technologies that monitor and analyse student learning behaviour and emotional states to inform teaching interventions. However, the rapid rise of these AI-enabled sensing and intervention technologies has outpaced a comprehensive understanding of their broader implications. For instance, the adoption of these technologies potentially impacts learners' autonomy and educational interactions. Existing ethical discussions and frameworks predominantly focus on computational concerns such as data privacy, algorithmic bias and fairness. However, they often neglect the lived experiences of people affected by the system, particularly students, the primary recipients of its impacts. Students' perspectives and concerns remain underexplored. Moreover, AIEd adoption involves multiple stakeholder groups whose competing values and power dynamics shape design and implementation processes. How these actors negotiate ethical considerations, and how such decisions may contribute to or overlook ethical risks, remains unknown. Without insights into both the consequences of AIEd adoption and the dynamics behind decision-making, current approaches risk failing to safeguard students from unintended harm.

This thesis addresses these issues through a bottom-up exploration of how different stakeholder groups perceive and respond to AIEd and its impacts. It investigates three core questions: What are the ethical challenges of AIEd from students' perspectives? How do stakeholder dynamics shape ethical risks? And what values do students prioritise in AIEd adoption? This thesis comprises three empirical studies involving 71 student-generated narratives, 30 semi-structured interviews, and 131 survey responses collected through an online experiment.

Study 1 employs a speculative story-completion method to explore students' perceptions of AIEd. It reveals concerns about psychological discomfort, behavioural change, diminished learning autonomy, altered pedagogical landscape, and changing social dynamics in AI-mediated classrooms. It underscores the complex, interwoven impacts of AIEd beyond technical or functional considerations.

Study 2 uses semi-structured interviews with developers, parents, and school decision-makers to examine their ethical considerations in AIEd design and adoption. It uncovers critical ethical gaps rooted in power imbalances and divergent priorities amongst stakeholders. It reveals the systematic exclusion of students' perspectives and a significant accountability gap that fails to protect them from potential harm.

Study 3 employs a mixed-methods online experiment to investigate students' affective responses, beliefs about learning, and ethical priorities in relation to specific AIEd features. It validates the findings of Study 1 about emotional discomfort regarding monitoring features and shows a strong student preference for autonomy and privacy over presumed learning benefits. This misalignment with institutional priorities identified in Study 2 is not just a difference in preference but a critical ethical fault line that, if unaddressed, may erode students' trust, engagement, and well-being.

Together, these studies contribute novel empirical evidence that advances ethical, human-centred, and context-sensitive AIEd design and governance. The thesis extends current AIEd ethics discourse by foregrounding students' lived experiences and advocating for emotional well-being and user agency as central to responsible AIEd development. It calls for governance frameworks that are inclusive, reflexive, and responsive to those most affected yet least empowered in technology adoption.

Declaration

This is to certify that

- i. This thesis comprises only my original work toward the degree of Doctor of Philosophy, except where indicated in the Preface.
- ii. Due acknowledgment has been made in the text to all sources and materials used.
- iii. The total length of this thesis does not exceed 100,000 words, excluding tables, bibliographies and appendices.

Bingyi Han

March 2025

Preface

This PhD project has generated nine peer-reviewed outputs of which I am the first author. As thesis with publication, five of these papers (published or under review) are included as the core chapters of this thesis; the remaining four also stem from this project but are not reproduced in the main body.

Peer-reviewed publications included in this thesis

The following five papers are presented in their author-accepted versions, with only minor formatting amendments to ensure consistency across the thesis. Their inclusion complies with the University of Melbourne's policy for thesis with publication and has been approved by the advisory committee.

The declaration for publications was signed for this thesis submission, indicating that I have contributed at least 80% of the content in these articles. In all listed works, I led all stages of the research and writing process. I designed the studies, conducted literature reviews, collected and analysed the data, and wrote full drafts of all the papers. Throughout this process, I consulted my supervisors, George Buchanan, Dana McKay, Simon Coghlan, and Wally Smith, who provided feedback and suggestions regarding direction and writing. In the second paper presented in Chapter 4, my research fellow, Dr Sadia Nawaz, contributed to paper restructuring and editing on the final paper drafts. In the paper presented in Chapter 6, my fellow PhD candidate, Ying Ma, contributed to quantitative data analysis.

Listed below are the publication status and full references for each article, structured according to the thesis chapters in which they are presented.

Chapter 4

Paper 1 - Published by the Association for Computing Machinery (ACM) in Dec 2022.

Bingyi Han, George Buchanan, and Dana McKay. 2023. Learning in the Panopticon: Examining the Potential Impacts of AI Monitoring on Students. In Proceedings of the 34th Australian Conference on Human-Computer Interaction (OzCHI '22). Association for Computing Machinery, New York, NY, USA, 9–21. <https://doi.org/10.1145/3572921.3572937>

Paper 2 - Published by Springer Nature in Jan 2025.

Bingyi Han, Sadia Nawaz, George Buchanan, and Dana McKay. 2025. Students' Perceptions: Exploring the Interplay of Ethical and Pedagogical Impacts for Adopting AI in Higher Education.

International Journal of Artificial Intelligence in Education. Jan 2025:1-26.
<https://doi.org/10.1007/s40593-024-00456-4>

Paper 3 - Published by the Association for Computing Machinery (ACM) in Apr 2025.

Bingyi Han, Simon Coghlan, George Buchanan, and Dana McKay. 2025. Who is Helping Whom? Student Concerns about AI-Teacher Collaboration in Higher Education Classrooms. *Proceedings of the ACM on Human-Computer Interaction* 9.2, 1-32. <https://doi.org/10.1145/3711104>

Chapter 5

Paper 4 - Submitted for publication to Elsevier - International Journal of Human-Computer Studies (IJHCS) in Mar 2025.

Bingyi Han, Simon Coghlan, George Buchanan, and Dana McKay. Ethical Gaps and Power Dynamics in Decision-Making about AI Adoption: The Case of AI for Monitoring Student Learning in Education. [under review]

Chapter 6

Paper 5 - Submitted for publication to the Association for Computing Machinery (ACM) CHI 2026 Conference in Sep 2025.

Bingyi Han, Ying Ma, Dana McKay, George Buchanan, Simon Coghlan, Wally Smith. AI Sensing Technologies in Higher Education: Students' Learning Perceptions, Affective Responses, and Ethical Concerns. [under review]

Peer-reviewed article based on this thesis research (not included in this thesis)

Four additional works (a book chapter, a conference paper, and two journal manuscripts currently under review) also stem from this thesis research but are not included in the main thesis body. As noted above, I led all stages of the research and writing process for these articles in consultation with my supervisors. These works are listed below.

Paper 6 - Published by Springer Nature in Jun 2023.

Bingyi Han, Sadia Nawaz, George Buchanan, and Dana McKay. 2023. Ethical and Pedagogical Impacts of AI in Education. *International Conference on Artificial Intelligence in Education*. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36272-9_54

Paper 7 - Published by Springer Nature in Dec 2025. The book chapter draft was included in Appendix A.

Bingyi Han, Simon Coghlan, George Buchanan, and Dana McKay. 2025. The Values of Academic Integrity in the Era of Surveillance AI in Classrooms. Chapter 8 on Artificial Intelligence and Academic Integrity - Navigating Ethical Challenges of AI in Education. Springer Nature Singapore 2025. <https://books.google.com.au/books?id=arOG0QEACAAJ>

Paper 8 - Submitted for publication to Springer - the Journal of AI and Ethics in Apr 2025.

Bingyi Han and Simon Coghlan. AI Sensing Technology in Classrooms: An Ethical Analysis. [under review]

Paper 9 - Submitted for publication to Emerald Publishing - Journal of Artificial Intelligence in Education in Aug 2025.

Bingyi Han and Simon Coghlan. Fairness in AIEd Beyond the Technical: Considering Unintended Consequences, Stakeholder Power, and Procedural Justice. [under review]

Ethics Declaration

This work was conducted following the University of Melbourne benchmark of Human Ethics. Ethics approval was obtained from the Human Research Ethics Committees before starting each research study. Study 1 Ethics Approval ID 22256, Study 2 Ethics Approval ID 24525, Study 3 Ethics Approval ID 30375.

Acknowledgment of Funding

This thesis was funded through the Australian Government's Research and Training program. Participant incentives were funded through the Human-Computer Interaction research group.

I was awarded two travel bursaries through the Human-Computer Interaction group and the Faculty of Engineering and Information Technology. The first was to travel to Canberra, Australia, to attend the OzCHI conference in 2021. The second was to attend the AIEd conference in Tokyo, Japan in 2023. This funding has offered me great opportunities, and I am deeply grateful for this support.

I was honoured to be selected as a PhD Thinker and received funding from the Kathleen Fitzpatrick Australian Research Council Laureate Fellowship to attend the PhD Thinkers Program 2023 at the University of South Australia, Adelaide, including support for travel and participation.

I was also awarded the DAAD the AINeT Fellow 2023 - Human-Centred AI by the German Academic Exchange Service (DAAD - Deutscher Akademischer Austauschdienst), which funded my two-week research visit to the Max Planck Institute for Human Development in Berlin in 2024.

Acknowledgments

This thesis would not have been completed without the encouragement, inspiration, support, and love of so many people, in all kinds of forms. From the beginning, I realised that a PhD is not just an academic journey; it's an odyssey of self-discovery. It's about understanding myself better, recognising the challenges I face in how I approach research, and learning how to grow through them. The act of doing research becomes a mirror of who I am as a person. This journey, though painful at times, has been a profound experience. It involved many struggles, tears, and moments of doubt. But I could never have made it this far without the incredible people around me – my family, friends, supervisors, mentors, colleagues, and the nurturing environment I was lucky enough to be a part of. Yes, the journey was painful, but the joy, satisfaction, and fulfilment I've gained far outweigh the hardships. I am forever grateful to those who listened to my excitement, celebrated every little success, and offered unwavering encouragement. It's because of them that I've made it here, and for that, I am truly blessed.

So, this acknowledgment is for the beautiful souls who have accompanied me along the way. Words can hardly express how thankful I am, but please know that my gratitude comes from the deepest part of my heart.

First and foremost, I would like to thank my supervisors, **George Buchanan, Dana McKay, Simon Coghlan, and Wally Smith**. You are the ones who offered me the ship and allowed me to steer it. Without your guidance, I could never have embarked on this voyage across the vast ocean of knowledge.

George, thank you for encouraging me to follow the topic I was truly passionate about, when four years ago I came to your office with a long list of possibilities, full of the belief that I could make the world a better place through my research. You gave me the freedom to explore ideas in my own way and believed in me from day one. There was a time I felt like I might be the happiest PhD student in the world, because I genuinely loved my research. That was thanks to your support.

Dana, thank you for your unwavering encouragement and the sparkling ideas you always brought, like twinkling stars lighting up my PhD journey. You made my research so much fun. I especially appreciated the way you protected me through times of rejection. You also protected me from doubting myself. You were the first to tell me I

had an academic voice of my own, that was when I began to believe in myself. Thank you for shaping me into a confident researcher.

Simon, you reminded me to take things slow and “sharpen the sword” – to refine my work carefully and thoughtfully. I learned so much from your meticulous attention to detail, like carving a beautiful piece of jade. I always looked forward to our “workshops,” where we would digest and explore the meaning of concepts or how to build them, thank you so much for teaching me this. Your emphasis on clarity, precision and patience has deeply influenced me, and I now recognise how essential these qualities are – qualities I didn’t always possess at the beginning. Thank you also for your warmth; I would have struggled much more without your support.

Wally, thank you for joining later in my journey and bringing not only kindness but also sharp insights. I’m especially grateful for your structural guidance and critical lens that challenged me to elevate my work. I learned so much from your approach, which is both rigorous and supportive, and your influence will continue to guide my future research endeavours.

I must thank **Ryan Kelly**, who encouraged me and helped me to begin this journey. When I started, we made a bet on whether I would still have the same excitement and passion for research by the end of it. Well, I can now tell you that I’ve won the bet! My excitement has only grown as I realise how much I truly love research. Even now, my heart races when a new idea sparks. Thank you, Ryan, for your generous support and thoughtful suggestions throughout this journey.

I’m deeply grateful to the amazing research group at the University of Melbourne. My fellow PhD buddies, mentors, and friends offered countless support and genuine advice. Being part of the Human-Computer Interaction group has been a highlight of my journey. A heartfelt thank you to **Xanthe Lowe-Brown** (thank you for your companionship and simply being there when I needed it most); **Elsy Garcia** (thank you for supporting me during my hardest times and witnessing my growth); the Human-Robot Interaction team – **Yan Zhang**, **Ramtin Tabatabaei**, **Tharaka Ratnayake**, **Jian Zhang**, **Seongwon Jeong** (thank you, my friends, for adopting me into your group and making me feel a true sense of belonging. I’m not sure I could’ve survived the final stretch of my PhD without our weekly hiking trips, group dinners, movie nights, and everything else we shared); **Joshua Newn** (a very important friend, thank you for your warmth and being the “adult” on all our road trips); **Ying Ma** (thank you for helping me exploring in the

quantitative world, our conversations on methodologies were invaluable); **Sadia Nawaz** (thank you for your warm support and ongoing encouragement). There are so many others whose energy and companionship I appreciate deeply: **Wei Zhao, Mo Zhang, Piumi Perera, Sophie Freeman, Madeliene Antonellos, Yushan Xing, Matthew Sidji, Stella Shuai Yuan, Lucy Sparrow, Ulan Kelesbekov, Songyan Teng, Jarod Govers, Joe Brailsford, Le Fang, Saumya Pareek, Samangi Arachchi, and Audrey Balaska** - Thank you.

I'm also thankful to friends from the Information Systems group and the Centre for AI and Digital Ethics (CAIDE) – thank you for the peer support during every coffee break: **Rashika Bahl, Davor Petreskei, Susie Sheldrick** (and **Basil Sheldrick**, the dogtrotal candidate!), **Ashley Anderson, Stella Peng, Mingxin Zhang, Dilang Tan, Mike Zhuang, Qianyi Li, Jiawei Tong, Joshua Rose** and **Brian Martin**. To my lovely office neighbour in FEIT, **Nikoo Mirhosseini**, thank you for sharing your good vibes, especially during our late-night work sessions. And to my closest research companions, **Rae Liu, Wei Yi, and Xian Wen**, thank you for all the deep conversations and philosophical reflections about research and life. Those moments were truly inspiring and always left me energised and excited!

A heartfelt thanks also goes to the senior staff members of the HCI group, IS group, and CAIDE. Your passion and dedication to research have been a constant source of inspiration for me. I am grateful for the kind advice you've shared and the enriching intellectual exchanges during seminars, research workshops, and group retreats – all of which have shaped my academic journey: **Jorge Goncalves, Wafa Johal, Eduardo Velloso, Adelaide Genay, Jenny Waycott, Sarah Webber, Matin Gibbs, Melissa Rogerson, Arzoo Atiq, Shanton Chang, Reeva Lederman, Rod Dilnutt, Ofir Turel, Tim Miller, Simon D'Alfonso, Marc Cheong, and Shaanon Cohnen**. Thank you!

To my best friends in Melbourne – **Chris, Carina, Mengliu, Gary, Zee, Quan, Mas, Ignacio, Rohit**, and in China – 杨柳青青, 刘琳娜, 张艺赢, 高则诚, 郭啸东, 余潇, 刘蕊, 兰茜, 张司典, 刘笑帆, 刘学 – You have given me so much strength. I'm not sure how many times I've cried in front of you, but thank you for always being there for me. My love and gratitude for you all need no more words!

To my family. 我最亲爱的爸爸妈妈: 言语如此单薄, 我无法用这些简单的字句向你们表达我的感谢。我的爸爸**韩水清**, 你总是为我提供最好的, 支持我去追

求我想要的，你说你想飞多高就去飞多高，我能做的只有尽我所能把梯子为你搭好，剩下就要靠自己展翅。我无比知足你为我提供的和我所拥有的。我的妈妈谷颖虹，你总想尽办法让我开心健康快乐，这是最无价的支持，在我难过时你总说“不开心就不学了回家吧做妈妈的快乐女儿”。我如此地被爱着，我是最幸运的女儿。没有你们的爱与支持，我无法完成这个博士论文，谢谢你们。Words feel so think, I cannot truly convey my gratitude to you with these simple sentences. My dad, **Han Shuiqing**, you always offer me the best and support me in pursuing my goals in every way. You've always told me, "Just spread your wings and fly, and I'll build the ladder". I feel so lucky for everything that I have. My mom, **Gu Yinghong**, your care and unconditional love gave me strength. Thank you for always reminding me that happiness should come first. You always told me, "If you're unhappy, quit PhD and come home to be my happy daughter." I am so beloved. I am truly the luckiest daughter. I can't reach here without your love. Thank you!

To my dearest grandmother, Guifen Wu (吴桂芬), who left me during the middle of my PhD – thank you for seeing how it all began and for celebrating with me when I first embarked on this journey. I wish more than anything that you could be here to celebrate its completion, to hear me called "Dr." for the first time. But I know you are still with me, just in a different way. Your immense kindness, gentle strength, and quiet wisdom have always guided me. You've inspired me to be more empathetic, more grounded, and a better person. Your love stays with me always. 我的外婆姥姥，我很难过你见证了我读博的开始却不能见证我顺利完成学业毕业，尽管我知道你会一直坚定地相信我并为我感到骄傲。你的善良，包容，坚韧一直影响着我去做更好的人。我知道你一直还在我身边，只是以不同的方式。谢谢你。

Lastly, I want to thank **myself** – thank you for believing in yourself, for taking care of yourself, for being tough and strong, for setting high expectations and for being willing to challenge yourself. I am so proud of you!

This has been a memorable journey – one I will never regret.

Table of Contents

Abstract.....	i
Declaration.....	iii
Preface.....	iv
Acknowledgments.....	vii
Table of Contents.....	xi
Chapter 1: Introduction.....	1
1.1 Background.....	2
1.2 Problems	5
1.3 Aims and Scope	10
1.4 Significance	12
1.5 Thesis Outline	15
Chapter 2: Literature Review	21
2.1 AI in Education.....	22
2.2 User Perceptions & Experiences of AIEd	28
2.3 Ethical Discussions and Frameworks	31
2.4 Summary – The Need for Human-Centred Ethics.....	35
Chapter 3: Research Approach.....	37
3.1 Overall Strategic Design Thinking	37
3.2 Overall Epistemological Stance: Constructionism	39
3.3 Theoretical Underpinning – Symbolic Interactionism	41
3.4 Study 1 – Story Completion Research.....	42
3.5 Study 2 – Interview Study	47
3.6 Study 3 – Experimental Study	52
3.7 Summary of Research Approach	54
Chapter 4: Students’ Perspective: Ethical and Sociotechnical Consequences of AIEd Adoption.....	55
4.1 Approach for Study 1.....	56
4.2 Findings and Contribution to the Thesis.....	56
4.3 Publication Information	58
4.4 Impacts on Individuals: Emotional Discomfort, Behavioural Change, and Cognitive Disruption.....	59
4.5 Impacts on Pedagogical Landscape: Learning Autonomy, Educational Processes, and Roles	73
4.6 Impacts on Classroom Dynamics: Interactions among Students, Teachers, and Systems	100

4.7	Additional Reflections	133
Chapter 5: Powerful Stakeholders' Perspectives: Ethical Gaps and Power Dynamics in AIED Decision-Making.....		135
5.1	Approach for Study 2.....	136
5.2	Findings and Contribution to the Thesis.....	136
5.3	Ethical Gaps in AIED Design and Adoption: Students' Systematic Exclusion, Developers' Pragmatism, Schools' Limited AI Literacy	138
5.4	Additional Reflections	171
Chapter 6: Students' Responses: Affective Responses, Learning Beliefs, and Ethical Priorities to AIED Design.....		172
6.1	Approach for Study 3.....	173
6.2	Findings and Contribution to the Thesis.....	174
6.3	Students' Affective Responses, Beliefs about Learning, and Prioritised Ethical Values to AIED Features	178
6.4	Additional Reflections	208
Chapter 7: Discussion.....		209
7.1	Synthesising Perspectives: Identifying Unique Ethical Challenges	209
7.2	Implications for AIED Ethics: Human-Centred Approach	213
7.3	Implications for the Future of AIED	218
7.4	Methodological Reflections on Story Completion	228
Chapter 8: Conclusion		230
8.1	Contributions	230
8.2	Limitations and Future Directions	231
8.3	Concluding Remarks	234
References.....		239
Appendices		249
Appendix A Book Chapter – AI and Academic Integrity.....		249

Chapter 1: Introduction

AI adoption into education (AIEd) presents both transformative opportunities and critical challenges. On one hand, it promises to enhance teaching, learning, and education management (Miao et al., 2021) through ‘intelligent’, ‘adaptive’, and ‘personalised’ systems (Holmes & Porayska-Pomsta, 2023). On the other hand, concerns arise over its computational issues, such as data privacy – ensuring student data is securely handled (Drachsler & Greller, 2016); algorithm bias – where AI may unintentionally favour certain groups over others (Baker & Hawn, 2022); and fairness – whether AI-driven decisions treat all students equitably (Madaio et al., 2022). Beyond these technical risks, sociotechnical implications include autonomy – how much control students have over AI-driven decisions (Sclater, 2016); and inclusion – ensuring AI benefits all learners rather than reinforcing existing inequalities (Holmes et al., 2021). Without careful design, implementation, and governance, AIEd risks unintended harm, particularly to vulnerable users – students.

AIEd adoption is intricate, impacting and being impacted by multiple stakeholders, including students, teachers, admins and institutions within the educational ecosystem. While ensuring its ethical and effective implementation requires cross-disciplinary effort among computer science, learning sciences, ethics, and public policy (Luckin et al., 2016), current discussions often lack empirical insights into how AIEd is experienced, adopted, and governed. There is limited understanding of its real-world consequences, how adoption decisions are made, and the broader sociotechnical considerations that shape its impact. This thesis addresses these problems by providing a human-centred, empirically grounded perspective on AIEd design, adoption and its ethical challenges.

This thesis, rooted in the discipline of human-computer interaction (HCI), examines the lived experiences¹ and decision-making processes of key stakeholders – students, parents, school governors, teacher representatives and developers. Rather than focusing

¹ In this thesis, the term “lived experiences” is used in the HCI tradition to denote participants’ authentic, situated perspectives on technologies, rather than in a phenomenological sense. While phenomenology (Van Manen, 2016) seeks to uncover the essential structures of lived experiences, this thesis adopts a symbolic interactionist stance (see Chapter 3 Methodology), examining how meanings are constructed in social interactions and situated context.

solely on computational concerns, it explores how different stakeholder groups interpret, prioritise, and negotiate their concerns about AIED adoption, particularly in relation to AI-powered classroom sensing and intervention systems. By adopting a constructionist paradigm, this thesis unpacks the sociotechnical dynamics that shape AIED's adoption and impact, providing empirical evidence on unanticipated ethical consequences and power imbalances in decision-making. It challenges technology-centric ethical discussions by foregrounding students' lived experiences, well-being, and agency, filling gaps in existing AI Ethics and AIED governance frameworks. Furthermore, this thesis bridges HCI, AIED, and AI ethics by advocating for context-sensitive, human-centred ethical principles that account for the complex realities of technology adoption. By taking a bottom-up approach, it supplements normative ethical frameworks – policy-driven ethical guidelines – ensuring that AIED adoption aligns with human-centred values, particularly for those most vulnerable and with the least agency in these decisions.

This introductory chapter outlines the background of this thesis and my motivation for this research topic (Section 1.1), presents the problem of the research (Section 1.2), articulates its aims and scope (Section 1.3), and discusses the research significance (Section 1.4). Finally, Section 1.5 provides an overview of the subsequent chapters.

1.1 BACKGROUND

Artificial Intelligence in Education (AIED) refers to the application of artificial intelligence (AI) technologies to enhance learning experiences, teaching practices, and educational decision-making. It integrates techniques such as machine learning, natural language processing, and computer vision to support personalised learning, automate assessment, and improve administrative efficiency (UNESCO, 2019). AI-powered solutions hold transformative potential, contributing to UNESCO's Sustainable Development Goals (SDG) (Vivar & Peñalvo, 2023) on quality and inclusive education (Miao & Holmes, 2023). AI adoption in education has been marked by the development of 'Intelligent', 'adaptive', and 'personalised' learning systems, primarily driven by the private sector for global deployment in schools and higher education institutions (Miao & Holmes, 2023).

The introduction of AIED dates back to the 1970s, with early systems using rule-based techniques to tailor learning experiences for individual students (Carbonell, 1970). Over time, AIED has evolved into three key domains - student-facing, teacher-facing, and

system-facing applications (Baker & Smith, 2019) - which now involve multiple stakeholders in more dynamic educational interactions. Today, advanced AIED leverages *machine learning*, *learning analytics*, and *educational data mining* to model students' cognitive and affective states, generating deeper insights into learning processes. These insights drive applications such as *intelligent tutoring systems*, *virtual pedagogical agents*, and *smart classroom systems* to work with teachers and students (detailed in section 2.1.1). Current AIED efforts focus on two salient challenges identified by Luckin et al. (2016): developing reliable and valid indicators to track learner progress across various skills and capabilities and gaining a deeper understanding of the most effective teaching approaches and learning contexts.

To address these challenges, recent AIED development has focused on deploying *classroom sensing techniques*, such as facial recognition, gaze-tracking, and systems designed to infer students' emotional and attentional engagement in real time. This trend reflects a growing reliance on collecting extensive data from students, encompassing their academic interactions to biometric information, to facilitate classroom observations, real-time analysis, and performance assessments. These systems can then provide teachers with actionable teaching strategies and timely intervention suggestions, and offer students personalised learning feedback and support. Many schools and universities have eagerly embraced such technologies to enhance teaching and learning in diverse settings, from general university classrooms (Alexander et al., 2019; Smolaks, 2019) to more specialised fields such as engineering education (Alzoubi et al., 2021) and nursing education (Martinez-Maldonado et al., 2023). This type of AI-powered classroom sensing and intervention system is the main focus of this thesis, which I will detail in Section 2.1.2.

These rapid technological developments and increasing deployment of AIED have inevitably introduced numerous risks and challenges, which “have so far outpaced policy debates and regulatory frameworks” ((Miao & Holmes, 2023), p.1). Key concerns include data privacy, algorithmic bias, fairness, and transparency (I will detail these in section 2.3). These challenges are compounded by the vast volumes of sensitive student data collected, especially inferred emotional states and behavioural patterns. If left unchecked, for instance, biases inherent in the data or algorithm risk perpetuating inequities, potentially disadvantaging underrepresented or marginalised student groups (Holmes et al., 2021). Furthermore, the opacity of AI decision-making processes complicates

accountability and trust, especially when these systems directly influence educational outcomes or classroom dynamics.

To address these concerns, the AIED community has introduced governance frameworks and ethical guidelines emphasising informed consent, data security, and algorithmic fairness (I will detail these in section 2.3.1). However, these efforts predominantly focus on technical and computational issues, especially on data use and management (Holmes et al., 2021). While essential, these frameworks fail to account for broader sociotechnical implications of AIED adoption, such as its impacts on the autonomy, agency, and well-being of students and teachers.

Currently, no ethical frameworks specifically address the human-centred implications of AIED adoption, particularly its long-term consequences (Holmes et al., 2021) on learning experiences, pedagogy, and social dynamics in education. There are no formal principles - such as codified rules or established guidelines - to regulate ethical AIED design and adoption practices. Likewise, no informal principles - such as culturally embedded norms and socially agreed-upon standards - exist to guide ethical decision-making in AIED implementation. This absence creates a critical oversight, risking the misalignment of AIED adoption with fundamental educational and ethical values.

Yet, as Holmes et al. (2021) observed, there remains a “disappointingly low level of interest in the (ethics) topic from the broader AIED community”, and a lack of “meaningful ethical reflection necessary for innovation in AIED”. Without deeper ethical engagement, AIED risks being designed and adopted in ways that compromise students’ well-being. From a student’s perspective, addressing only technical safeguards, such as data security and algorithmic regulation, is insufficient. AIED’s introduction in classrooms raises concerns about well-being, autonomy, and comfort, which are often sidelined in governance-focused technical solutions. This gap in ethical reflection and human-centred considerations forms the foundation of my research motivation.

In 2019, even before the rise of generative AI, I came across a news report about a university piloting AI facial and emotion recognition technology in classrooms (Smolaks, 2019). This system tracked and quantified students’ attention, posture, and participation in class (how many times a student put their head on the desk, fell asleep, browsed their phone, raised their hands, etc). As a student, I felt personally unsettled by the thought of such invasive monitoring, and I wanted to know if the potential benefits of these systems outweigh the ethical concerns. My conversations with peers revealed widespread

resistance to such technologies. My initial research confirmed that this trial was not an isolated case; rather, advancing these technologies was a mainstream focus in the AIED community.

This realisation led me to question whether designers, developers, and educational decision makers understand how students perceive such technologies. Are students' voices being heard in the design and deployment of these systems? How are stakeholders, including students, involved in the decision-making process, and to what extent are students' opinions being taken into account? If students' perspectives are neglected, it raises critical ethical and practical concerns, as they are the primary recipients of both the positive and negative impacts of AIED adoption. Students represent a vulnerable and essential stakeholder group whose well-being must be prioritised. These unanswered questions and the broader implications for multi-stakeholder engagement in AIED design ultimately shaped the focus of my PhD research. It also became one of my idealistic missions in pursuing this PhD: If AIED developers and researchers do not know how students perceive and respond to these technologies, they should. If they do not think they need to know, they absolutely should.

1.2 PROBLEMS

As discussed in the previous section, AIED advancements are becoming increasingly pervasive and potentially more intrusive. Technologies such as facial recognition, behavioural tracking, and emotion detection are pushing the boundaries of what is possible to transform the learning environment. While these innovations hold promise, they also introduce a host of challenges and complexities that remain largely unaddressed. The rapid adoption of AIED raises critical concerns regarding its impact on students, its ethical implications, and the governance of AIED design and use. At the heart of these challenges lie four key problems that must be addressed to ensure the responsible and inclusive integration of AIED in education.

Problem 1: Lack of Understanding of Unintended Consequences of AIED

The most visible and urgent issue is that, as AIED technologies become more integrated into education, their unanticipated and broader consequences remain poorly understood. Much of the existing research focuses on intended benefits, such as personalised learning, adaptive feedback, improved administrative efficiency, and enhanced educational outcomes. Much effort is put into technological advancement to

track students' in-class behaviours and thus more accurately capture students' learning performances and processes. However, there is little empirical evidence of the real-world effects of AIED adoption on students, particularly its potential risks to individuals' autonomy, agency, and learning experiences.

More than affecting individual students, AIED technologies, particularly more intrusive ones such as those incorporating facial recognition, behavioural tracking, and emotion detection, may reshape learning environments in unforeseen ways. For instance, these systems might inadvertently introduce new forms of student monitoring and behavioural control, raising concerns about surveillance-oriented environments and students' motivation. However, these potential impacts are less discussed and remain largely unexplored (I will explore them in Chapter 2 of the literature review).

This lack of empirical understanding of potential negative consequences of AIED limits the ability to design proactive safeguards. The ethical risks of AIED adoption often emerge in subtle, context-dependent ways, making it difficult to identify problems before they manifest at scale. If negative consequences are only recognised after harm has already occurred, it may be too late to mitigate their impact effectively. Knowing about impacts is a foundation for ethical action. We need a thorough understanding of the full spectrum of when, where, and how harm can occur and an understanding of who might be harmed and who causes harm (Nunes Vilaza et al., 2022). This understanding is required to ensure that ethical considerations are in place to guide the entire process of technology from creation to adoption before it becomes deeply entrenched in educational systems.

Problem 2: Limited Understanding of Stakeholder Dynamics in AIED Adoption

Why do unintended consequences occur? We lacked an understanding of how decisions of AIED design and adoption are made. AIED adoption does not happen in isolation; the educational ecosystem involves multiple stakeholders. It includes students, teachers, educational institutions, policymakers, developers, and researchers, each with different interests, priorities, and ethical considerations. These considerations can vary according to their specific roles, backgrounds and contexts. Designing, implementing, and governing AIED technologies in a way that addresses these diverse needs is inherently complex.

There are unresolved questions about how different stakeholder groups perceive AIED design and adoption, how their needs intersect in practice, and how relationships and potential tensions among these stakeholders complicate decision-making processes. For example, it remains unclear to what extent students' needs are considered in these processes and how ethical decisions account for students' concerns. Without a comprehensive understanding of these dynamics, attempts to foster ethical and inclusive AIED systems risk being uninformed and ineffective. Moreover, understanding the practical obstacles and challenges involved in implementing ethics-related decisions in AIED is critical to making productive and actionable recommendations for ethical design and governance.

Problem 3: Neglecting Students' Perspectives in AIED Design

An even more critical problem stemming from the unknown stakeholder dynamics is the neglect of students' perspectives in the design and development of AIED systems (Problem 4). As the primary recipients of these technologies, students are directly impacted by how AI shapes their learning experiences and are more vulnerable to its risks. Despite AIED's intention to influence student behaviour, little is known about how students perceive these interventions and whether they consider the technology useful, effective, ethical, or acceptable.

Designing technologies purportedly to benefit students without a clear understanding of their genuine needs, values, and ethical concerns is inherently problematic. This lack of insight risks creating AIED systems that are misaligned with students' expectations and educational goals, and could undermine their engagement, trust, and willingness to adopt these technologies. This misalignment can undermine the educational benefits AIED aims to deliver and ultimately hinder its effectiveness.

Furthermore, this issue of misalignment is exacerbated by the absence of user-centred design principles in AIED development (Holmes et al., 2021). Without integrating students' perspectives into the design process, AIED systems are more likely to overlook critical considerations related to their needs and ethical concerns. For AIED technologies to be effective and ethically responsible, it is essential to centre students' voices from the outset of development, ensuring that these systems align with their values and contribute positively to their learning experiences and goals.

Problem 4: Limited Understanding of Ethics of AIEd, Lacking a Human-Centred Approach

There is no universally agreed-upon understanding of the ethics of AIEd yet, though scholars have explored it from different perspectives (detailed in Chapter 2). For instance, some emphasise that AIEd ethics extends beyond technical concerns - data and computational aspects such as data privacy, algorithmic fairness, and transparency - to encompass broader educational and social values (e.g., Holmes and Porayska-Pomsta (2023), Holmes et al. (2021)). Similarly, Nguyen et al. (2023) seek to establish a global consensus on ethical AIEd through seven core principles to ensure AIEd remains transparent, fair, accountable, inclusive, and aligned with human-centred values. From these discussions, the ethics of AIEd can be understood as the principles and values that guide the responsible design, deployment, and governance of AI technologies in education to ensure that AI enhances education without compromising human dignity and causing harm. Ethical AIEd should encompass both technical safeguards and sociotechnical considerations (e.g., ensure AI aligns with pedagogical and human-centred values).

However, our current understanding of the ethics of AIEd remains insufficient, as it primarily addresses technical dimensions (Holmes et al., 2021), overlooking the sociotechnical consequences of AIEd adoption. It includes how these systems interact with and reshape social dynamics in education, influencing students, teachers, and learning environments in nuanced and often unpredictable ways. We currently lack the depth of understanding needed to capture the complexities of these technologies in educational settings. Ethical risks extend beyond “known unknowns” - anticipated technical challenges - to “unknown unknowns” (Holmes et al., 2021), the unintended consequences that emerge only through real-world use. For instance, while algorithmic bias and fairness are well-recognised problems, how AIEd shapes teacher-student relationships, influences educational inequalities, or alters the nature of learning itself remains underexplored. Yet, as emphasised in Problem 1, we lack empirical evidence to identify and address these unforeseen risks. Without this foundation, ethical guidelines risk being abstract ideals rather than actionable frameworks that protect students and educators from unintended harm. Problem 1 also forms this fundamental issue that we do not yet fully understand the scope of ethical issues at play.

Another challenge is that a lot of AIED researchers and developers are not equipped to tackle emerging ethical questions (Holmes et al., 2021), leading to a disconnect between technological aspirations and ethical considerations. Disappointingly, the AIED research community (as reflected in discussions and presentations at the International Conference on Artificial Intelligence in Education, a leading venue for AI and education research) has shown a low level of engagement with ethical issues (Holmes et al., 2021), even as education is recognised as a high-stakes domain for integrating AI (Campolo et al., 2017). Many ethical concerns arise not because of intentional harm but due to oversights in design and deployment. This lack of ethical attention and expertise (knowledge, skills, resources) from the AIED community may be because of how AIED interacts with complex educational settings and because the potential unintended consequences are not yet well-identified.

To address this insufficient understanding of the ethical landscape of AIED design, adoption and governance, we must move beyond technical and compliance-driven discussions within the AIED community. We need a human-centred approach to determine what ethical considerations should be included, drawing from the perspective of those directly affected by these systems. A more holistic human-centred ethical focus accounts for user experiences and socio-technical dynamics, to ensure AIED systems align with educational values and user well-being. Without this shift in ethical frameworks, issues in Problems 1, 2 and 3 will persist. To inform human-centred ethical principles that reflect the realities of AI-driven educational environments, we need to engage with key stakeholders and users of AIED.

Summary

This section has highlighted four critical challenges in the ethical and responsible adoption of AI in educational contexts. Central to these challenges is the need for a robust empirical understanding of the potential ethical consequences of AIED, the complex interactions among key stakeholders in educational ecosystems, and the perspectives of students as the primary beneficiaries. Addressing these issues will not only inform the design of future AIED systems but also provide a theoretical framework for the AIED community to meaningfully engage with ethical concerns. Ultimately, these insights will help inform user-centred design principles that prioritise ethical considerations throughout the development and deployment of AIED technologies. Additionally, they

will enhance the existing AI and AIEd ethical frameworks by incorporating human-centred approaches that foreground users' lived experiences and well-being.

1.3 AIMS AND SCOPE

This thesis aims to address the problems identified in the previous section by providing a more comprehensive empirical understanding of ethical AIEd design and adoption. It focuses on examining how different stakeholder groups perceive and respond to AIEd adoption and its impacts (Aim 1). By filling the gap in understanding the sociotechnical consequences of AIEd, this thesis seeks to expand the ethical discourse of AIEd through a human-centred approach (Aim 2). Building on these insights, this thesis aims to offer actionable implications and recommendations for ethical AIEd design, adoption, and governance that are grounded in stakeholder concerns and ethical considerations (Aim 3). The overarching objective is to inform the development of human-centred, ethically responsible AIEd systems that align with educational needs and human values. To achieve this, the aims of this thesis have been translated into the following research questions:

Overarching Research Question: What are the ethical challenges in adopting artificial intelligence in education (AIEd) from the perspectives of different stakeholder groups?

RQ1: What ethical challenges do students perceive in the adoption of AIEd?

RQ2: How do stakeholder dynamics shape ethical risks in AIEd adoption?

RQ3: What are students' affective responses, beliefs about learning, and prioritised ethical values respecting AIEd design and implementation?

These three sub-questions are designed to address the overarching research question, with the thesis structured correspondingly around them through three studies. To answer these questions, this PhD project focuses on various key stakeholder groups, categorised by their roles in interacting with AIEd:

Students: the primary target and proposed beneficiary of AIEd technologies, whose experiences and needs are central to understanding the ethical implications of their adoption.

Decision Makers: This group includes more influential stakeholders involved in the design, implementation, and oversight of AIED, such as parents, school governors and teacher representatives, and AIED developers and practitioners.

In the studies presented in this thesis, I engaged with multiple stakeholder groups to explore their responses to ethical AIED design and adoption. The primary focus is on students – their perceptions and concerns of AIED adoption (study 1), and whether other stakeholders, such as parents, developers and school decision makers, have adequately considered their concerns during the design and implementation (study 2). Finally, this thesis investigates whether students agree with the values prioritised by decision-makers in the AIED ecosystem (study 3). Importantly, this study does not aim to critique specific technologies but seeks to genuinely uncover stakeholder perspectives to inform future design and encourage ethical reflection within the AIED community.

As a thesis with publications, there are five papers (three published, one under review, and one manuscript) presented in this dissertation, demonstrating findings from three studies across Chapters 4-6. Each of the studies has evolved alongside and informed how I developed the key research question and its associated aims. While each of the five papers attends to the *Overarching Research Question* in various ways, they must be taken as a whole to answer it in a comprehensive manner. As such, the answers to the research question will be reflected on throughout the thesis, in the introductions to the study publications and in the comprehensive discussion in Chapter 7.

Regarding AIED technologies, this thesis focuses on AI-enabled classroom sensing and intervention systems. This focus is for several reasons: First, these technologies align with the systems that initially inspired this research, such as AI facial and emotion recognition systems. Second, these systems represent some of the more intrusive applications of AIED, involving real-time monitoring of students in the classroom. Also, the classroom dynamic is more complex, involving interactions among students, teachers and AI systems. Third, they are among the most popular research streams in the current AIED community. I will detail this research trend of AIED in Chapter 2, the literature review. By concentrating on the domain of AI-enabled classroom sensing and intervention systems, the study seeks to provide rich, actionable insights that can spark broader discussions and improvements in AIED design. For conciseness, the term ‘AIED’ will sometimes be used throughout the thesis to refer to these technologies, even though it is a broader concept.

The context of this PhD project has evolved alongside the specific research questions addressed in each study. Initially, the focus was on understanding students' perspectives within higher education institutions. However, ethical issues identified in higher education - where students tend to have greater autonomy - are likely to be even more acute for younger, more vulnerable students in compulsory education settings like K12. As the research progressed to explore the dynamics of multiple stakeholders for Aim 1b, I shifted the focus to K-12 education. This shift allowed for the examination of a more complex context, involving not only younger students but also parents as key additional stakeholders. This shift in attention was done to capture a broader spectrum of ethical and sociotechnical challenges, reflecting the varying vulnerabilities and stakeholder interactions present in different educational contexts.

This study is fundamentally centred on ethics, with a specific focus on the ethical concerns expressed by stakeholders regarding AIED adoption and design. As discussed in the previous section, one of the key challenges lies in the limited scope and application of current ethical frameworks (which I will explore further in Chapter 2, Literature Review). Normative ethical frameworks often remain abstract and detached from the practical realities of AI adoption in specific contexts. For instance, they do not sufficiently address how AIED impacts social dynamics and pedagogical interactions. Hence, rather than adopting a top-down normative approach, this research takes a grounded, human-centred approach, informed by stakeholders' lived experiences and perspectives. This approach explores the meaning of ethics by examining how individuals understand AIED and perceive its rightness or wrongness. This aligns with my constructionist approach, which I will detail in Chapter 3. Ultimately, this thesis aims to expand the current ethical discourse about AI in education by incorporating critical human-centred considerations, enriching the development of future human-centred ethical principles for AIED.

1.4 SIGNIFICANCE

This thesis research is significant in its contribution to empirical knowledge surrounding the ethical challenges of adopting AI in education. Specifically, from a constructionist approach, it presents how stakeholders understand AI adoption and perceive its ethical risks, and any challenges derived from those multistakeholder ecosystems of AIED design and adoption.

Study 1 adopts a future-oriented speculative research method – Story Completion – to comprehensively examine the potential consequences of AIED adoption from students’ perspectives. Study 1 addresses Problems 1, 2, and 4. This study contributes five publications (three of which are included in this thesis in Chapter 4), exploring students’ perceived impacts across multiple dimensions. Paper 1 presents individual-level consequences, such as psychological and behavioural changes resulting from direct interactions with AI monitoring systems in the classroom. Paper 2 presents pedagogical-level impacts, such as influences on learner autonomy, learning processes and approaches, teacher-student relationships, and pedagogical roles. Paper 3 (along with Paper 2) demonstrates how these impacts interact dynamically, influencing one another rather than existing in isolation. These findings highlight the far-reaching implications of AIED, extending beyond interactions among students and teachers to reshape the broader educational ecosystem. This interdependence challenges conventional AIED design approaches that focus narrowly on technological advancement and system performance with fixed intended outcomes. Instead, it advocates for more holistic, context-sensitive perspectives, recognising AI, teachers, students and the educational environment as an integrated ecosystem. It needs to better account for nuanced social dynamics within education that are mediated by AIED. It also shifts the ethical discussion from technology-centric concerns to a more student-centred sociotechnical perspective that respects students’ lived experiences. These insights emphasise the importance of embedding ethical considerations into AIED development, ensuring that systems align with the complex realities of educational environments rather than simply optimising for efficiency or automation.

Study 2 examines the ethical considerations and power dynamics in the AIED design and adoption decision-making processes among key stakeholders, including developers, school representatives, and parents. Study 2 addresses Problems 2 and 3. This study contributes an under-review paper (included in Chapter 5) that critically analyses the problems in decision-making processes that may lead to unethical outcomes. Through semi-structured interviews, Study 2 reveals how power imbalances, institutional pressures, and limited AI literacy can lead to skewed value priorities, flawed adoption decisions, and accountability gaps. This study shifts the ethical focus from AIED technology itself to the decision-making structures and dynamics. It reconfigures the understanding of ethical risks by highlighting how stakeholder interactions and power

imbalances shape ethical and practical challenges. The findings offer valuable insights for fostering more equitable and inclusive AIED adoption and a more effective accountability mechanism to oversee AIED adoption.

Study 3 examines a critical yet underexplored aspect of AIED adoption: students' emotional well-being, beliefs about learning, and ethical values, particularly in the context of intrusive technologies. Study 3 addresses Problems 1, 2, 3, and 4, and contributes a manuscript for future submission (included in Chapter 6). Study 3 uses an online experiment with technology probes – video mock-ups of AI-driven classroom sensing and intervention systems. The results reveal a consistently negative reception to AI monitoring. It extends prior findings by showing that students' resistance stems not only from privacy concerns (Drachsler & Greller, 2016) but also from discomfort, loss of autonomy, and disruptions to the social dynamics of learning environments. It further highlights the misalignment between the value priorities of students (e.g., autonomy and privacy) and more powerful stakeholders (as identified in Study 2). This study underscores the need for non-intrusive, customisable, and socially sensitive AI systems, offering critical insights for AIED developers. By synthesising and validating findings from Study 1 and Study 2, Study 3 strengthens the thesis's overarching argument: AIED adoption must respect students' values and well-being to ensure ethical and effective implementation.

Across these three studies, this multidisciplinary thesis makes various contributions to the fields of AIED, HCI, and Ethics:

To AIED, this thesis provides empirical evidence, ethical insights, and actionable design recommendations to inform more student-centred and ethically responsible AIED development. It foregrounds students' needs, values, and emotional well-being – dimensions often overlooked in current design practices – and highlights unanticipated consequences of AIED adoption that may arise in real-world adoption. The findings challenge system-centric approaches for AIED design and development, advocating for a holistic, context-sensitive view of AIED systems embedded within complex educational environments. This approach, rooted in HCI, considers the interconnected relationships between AIED systems, students, teachers, and the broader education ecosystem. In doing so, this work shifts the discourse from a technology-driven mindset toward a human-centred paradigm that prioritises meaningful, inclusive, and ethical integration of AI in education.

To HCI, this research builds on its foundational perspective of technology's situated nature (Suchman, 1987), reinforcing its relevance in educational contexts through an empirical case study of AI adoption in education. This thesis also critiques conventional HCI user studies that evaluate innovative technology, which often rely on user acceptance metrics focused on system functionality and task performance. Instead, it advocates for a richer examination of users' lived experiences and ethical considerations when interacting with the technology. This thesis also contributes to HCI's understanding of designing and adopting multi-user technologies by exposing how stakeholder power dynamics complicate participatory design and inclusive innovation in practice and increase the risk of ethical challenges. These insights inform broader HCI discussions on the socio-technical complexities of designing and governing emerging technologies in sensitive settings, particularly where multiple stakeholder groups are involved within imbalanced power structures.

To AI Ethics, this thesis broadens the ethical discourse beyond technical safeguards by centring human-centred values, lived experiences, and well-being. It demonstrates the need to supplement existing ethical frameworks and general principles with empirical insights grounded in specific sociotechnical contexts. In particular, it contributes concrete empirical details to the principles such as beneficence, well-being and justice, by showing how students perceive and potentially experience the lived consequences of AIED adoption. By highlighting the emotional and relational dimensions of AIED use – specifically, how AI systems shape, influence, and mediate human relationships within the learning environment – this thesis lays the foundation for more human-centred ethical approaches that better reflect the complexity of real-world educational settings.

Collectively, these contributions reframe ethical AIED design and governance as a fundamentally human issue – one that requires attention not only to systems and principles, but to the everyday realities, values, and vulnerability of those most affected.

1.5 THESIS OUTLINE

This thesis is structured into 7 main chapters. It starts with the present introduction (Chapter 1), followed by the literature review (Chapter 2), research approach (Chapter 3), findings from three studies (Chapters 4-6), a discussion that synthesises the findings and reflections (Chapter 7) and the conclusion (Chapter 8). As a thesis by publication, Chapters 4-6 present three peer-reviewed papers, and two under-review manuscripts,

authored primarily by me, with introductory sections and discussions situating the publications within the overall thesis.

Chapter 2: Literature Review. This chapter reviews the current state of AIED technologies (more specifically, classroom sensing technology that this thesis focuses on), user perceptions and experiences with AIED, and existing ethics discussions and frameworks related to AIED. It synthesises literature across disciplines from AIED, HCI, and ethics, highlighting gaps in the existing research and the limited focus of current ethical approaches. This chapter establishes the foundation for the later discussion on the findings of this thesis (Chapter 7), which demonstrates how these findings address the identified gaps and expand or challenge existing ethical approaches.

Chapter 3: Research Approach. This chapter outlines the epistemology (constructionism), theoretical underpinning (symbolic interactionism and pragmatism), and specific methods of data collection and analysis used for each study: story completion with lite-version grounded theory for Study 1, semi-structured interviews with reflexive thematic analysis for Study 2, and an online experiment using a mixed-method approach for Study 3. This chapter also reflects on the rationale behind these methodological choices and the challenges I encountered during the research journey.

Chapter 4: Study 1 Findings. This chapter presents the first study, which comprises three published papers. It begins with a recap on its research approach, then summarises the key findings and contributions to the thesis. This chapter then presents three publications, each demonstrating different dimensions of the impacts of AIED adoption from students' perspectives. It finally presents additional reflections that were not included in the publications. Chapters 5 and 6 follow the same structure outlined in this chapter.

Chapter 5: Study 2 Findings. This chapter presents the second study and a manuscript that has been submitted for publication and is currently under review. The paper demonstrates three critical ethical gaps in AIED adoption processes and the power dynamics that underpin these gaps.

Chapter 6: Study 3 Findings. This chapter presents the third study and a manuscript that has been submitted for publication and is currently under review. The paper demonstrates students' affective responses, beliefs on learning, and ethical priorities in the context of AI-enabled classroom sensing-intervention technologies.

Chapter 7: Discussion. This chapter synthesises the findings presented in Chapters 4-6, revisiting and responding to the main research questions and aims of the thesis. It weaves together the insight from the three studies into a cohesive manner, articulating the overarching arguments of the thesis. This chapter also presents the implications for the ethics of AIEd and the future of AIEd, provides suggestions for more inclusive, human-centred, and context-sensitive AIEd design and governance, and discusses methodological reflections.

Chapter 8: Conclusion. This chapter summarises the key contributions of the thesis, outlines its limitations and future directions where the study may be extended, and concludes with key takeaways for researchers and designers developing ethical AI in education.

The overall structure of this thesis is shown in Fig.1 and Fig.1 (a). (see next page).

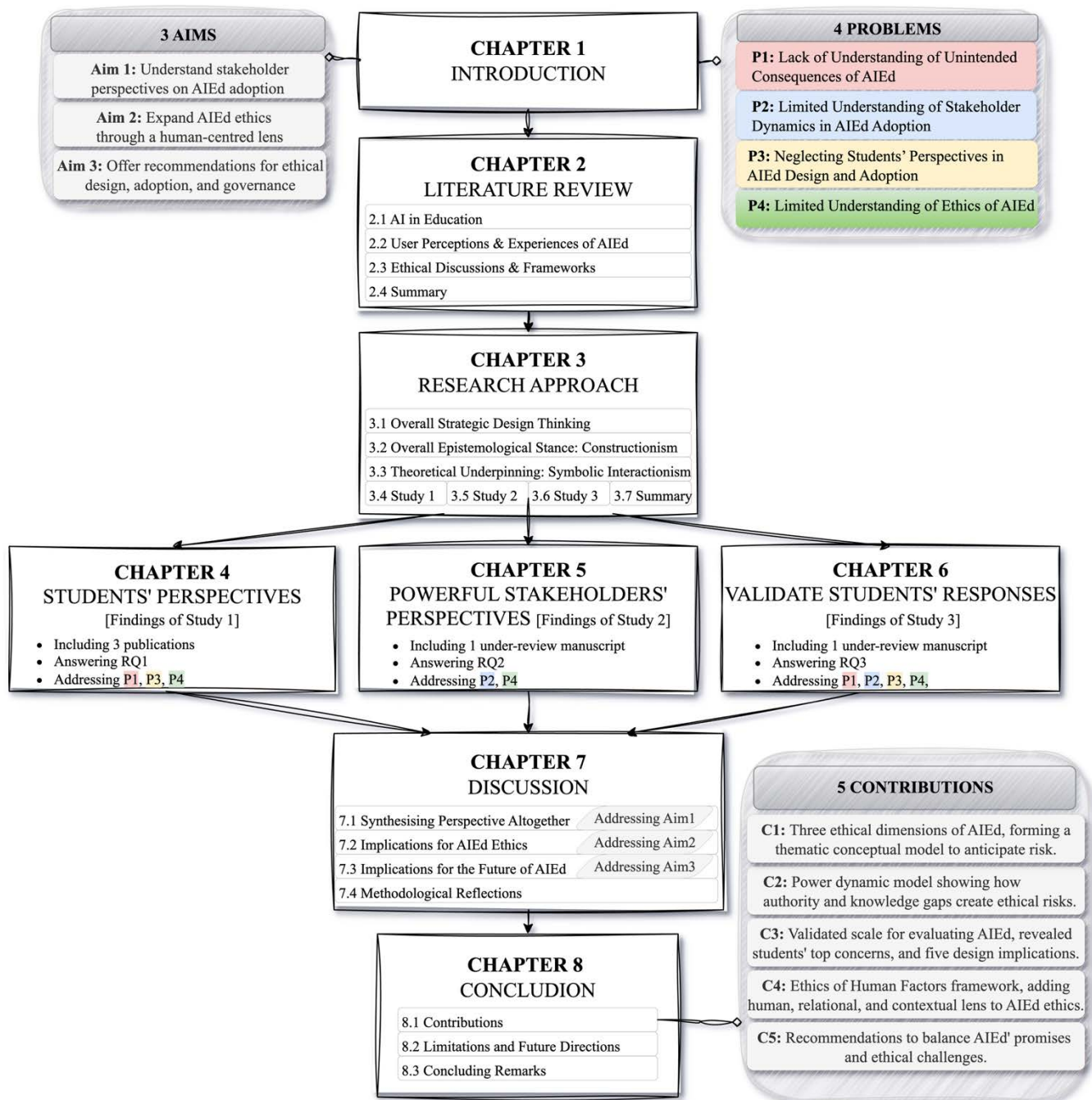


Fig. 1. Thesis Structure Overview

*Fig 1 and Fig 1(a) illustrate the overarching structure of the thesis. The research is driven by four identified *Problems*, which are translated into three key *Aims* that reflect the thesis's broader contribution. To achieve these aims, the overarching research question is broken down into three *Sub-Questions*, each addressed through one of the three empirical studies. Each study tackles the identified problems from a different angle. The Discussion chapter synthesises findings across the studies to respond to the overall aims, with each section aligned to a specific aim. The Conclusion chapter summarises key contributions. The relationship between problems, aims, sub-research questions (RQs), and chapters is detailed in the main text but summarised in Fig 1(a) to illustrate the thesis logic.

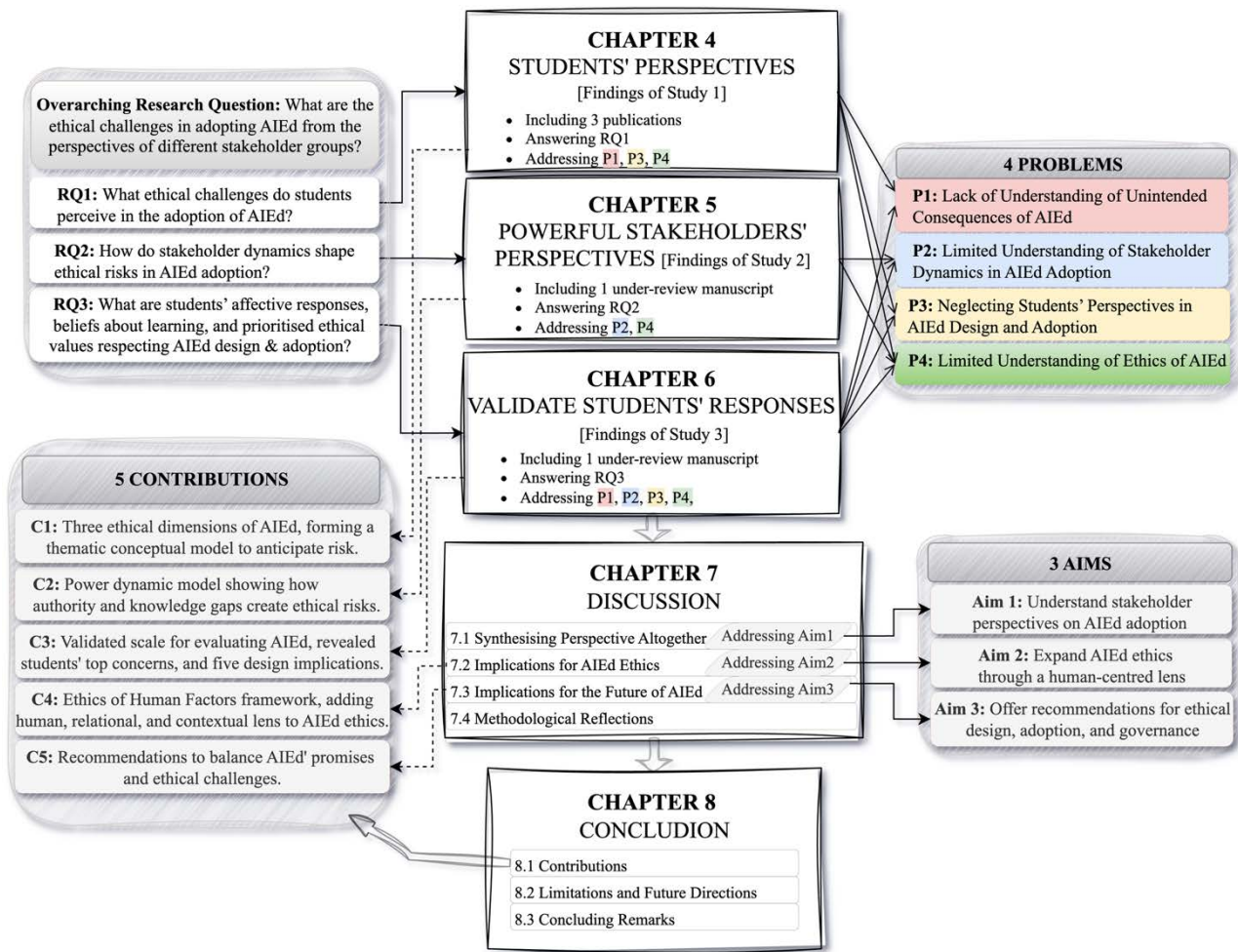


Fig. 1 (a). How problems, aims, research questions, contributions, and chapters relate to each other.

Chapter 2: Literature Review

This chapter reviews a range of interdisciplinary academic literature and institutional frameworks to examine the ethical challenges associated with the adoption of Artificial Intelligence in Education (AIED). It draws on research from AIED, AI ethics, and Human-Computer Interaction (HCI), as well as relevant government reports. The chapter is structured into three main sections: (1) the technological advancements and applications of AIED in recent years (Section 2.1), (2) user experiences and perceptions toward AIED (Section 2.2), and (3) ethical considerations and approaches surrounding AIED (Section 2.3).

Section 2.1 introduces the technological landscape of AIED, focusing on its increasing complexity and intrusiveness, setting the stage for the ethical focus of this thesis. Subsection 2.1.1 introduces key AI tools and techniques used in educational settings, while Subsection 2.1.2 narrows in on more sophisticated systems - AI sensing technologies in classrooms – which are central to this thesis. Then Subsection 2.1.3 reviews the current trend of AIED research. This section establishes the need for an empirical understanding of how stakeholders perceive and interact with these technologies as they become more deeply embedded in educational settings and increasingly process more sensitive user data.

Section 2.2 reviews existing research on users' perceptions and experiences with AIED, highlighting its narrow focus on system functionality and technical performance. It critiques the conventional technology evaluation frameworks and attitudinal scales that are commonly used in HCI and educational research, for failing to capture the deeper human aspects of learning. This section reveals a lack of understanding within existing AIED research and governance frameworks regarding users' lived experiences, along with insufficient empirical evidence on AIED adoption.

Section 2.3 explores ethical concerns surrounding AIED. Subsection 2.3.1 reviews existing ethics frameworks, highlighting their limited consideration of socio-technical implications and the absence of human-centred ethical principles. Subsection 2.3.2 expands on ethical concerns not covered by these frameworks and emphasises the lack of empirical evidence about AIED's broader consequences.

Together, this chapter identifies three key research gaps: (1) a lack of understanding of stakeholders' lived experiences and individual interactions with AIEd (Section 2.2), (2) a lack of empirical evidence on AIEd adoption and its broader consequences (Sections 2.2, 2.3.2), and (3) ethical frameworks failing to address human-centred concerns (Section 2.3). These gaps highlight the need for a more empirical and human-centred approach to assessing AIEd's ethical impact, forming the foundation of this thesis.

2.1 AI IN EDUCATION

Artificial Intelligence (AI) has been increasingly integrated into education over the past decade, promising to revolutionise teaching and learning. This thesis focuses on a specific category of AI in Education (AIEd) that is sophisticated and relatively more intrusive technologies used in classroom settings. These tools aim to empower teachers with data-driven insights for understanding students' learning status and enabling customised instruction while supporting students with timely assistance that is tailored to their needs.

This section first provides an overview of key AI techniques, tools and applications in the educational context (Section 2.1.1), then dives into the specific AIEd technologies central to this thesis – AI sensing technology (Section 2.1.2). It finally reviews the current trend of research in AIEd (Section 2.1.3).

2.1.1 Types of AI in Education

Three key AI techniques are broadly utilised in education: Machine Learning, Learning Analytics, and Educational Data Mining (Avella et al., 2016; Chen et al., 2020; Luckin et al., 2016):

1. Machine Learning (ML) is a subset of AI techniques. It discovers structured knowledge, identifies patterns, and builds predictive models from large datasets without being pre-programmed. ML in AIEd could be applied to tasks such as predicting students' academic performance and outcomes (Kučák et al., 2018).
2. Learning Analytics (LA) applies ML to education with a practice-oriented focus to enhance teaching and learning practices and outcomes (Lemay et al., 2021). LA involves collecting, measuring, analysing, and interpreting data from learners and their learning contexts (Romero & Ventura, 2013), and providing

insights into students' learning behaviours to inform effective teaching interventions (Arroway et al., 2015).

3. Educational Data Mining (EDM) develops methods to extract meanings from large-scale data generated by learners' actions in educational settings. It focuses on discovering new patterns and trends within educational datasets (Baker & Yacef, 2009). While EDM and LA overlap, EDM focuses more on methodological advancement, and often emphasises knowledge discovery using structured data; LA is more application-driven, aiming to directly improve educational practices (Lemay et al., 2021).

Those techniques, all of which rely on extensive student data, are often employed together in AIED artifacts. From the application perspective, Baker and Smith (2019) grouped AIED in three broad categories, noting that some artifacts combine features from multiple categories: 1) Learning-facing AIED, which supports students to receive and understand information and respond to their individual needs. For instance, the Intelligent Tutoring System (ITS) diagnoses students' strengths, weaknesses, and knowledge gaps to provide automated and customised feedback. 2) Teacher-facing AIED, which offers data-driven insights about individual students or class-wide progress, enabling teachers to effectively target their attention and automate administrative tasks like assessment and feedback. This positions teachers as "orchestrators of when and how to use AIED tools" (Luckin et al., 2016). 3) System-facing AIED, which assists with administrative and managerial decisions at the institutional level, such as organising timetables to predict school performance outcomes.

Current advancements in AIED technologies have primarily focused on enhancing teaching and learning processes (Southgate et al., 2019). Three key trending teaching-facing and learning-facing tools have been identified: pedagogical agents, intelligent tutoring systems (ITS), and smart classroom technologies (Southgate et al., 2019). These technologies (see Fig.2) are often integrated to foster collaborations and innovative interactions among teachers, students, and AI-enabled systems. These applications are at the heart of this thesis.

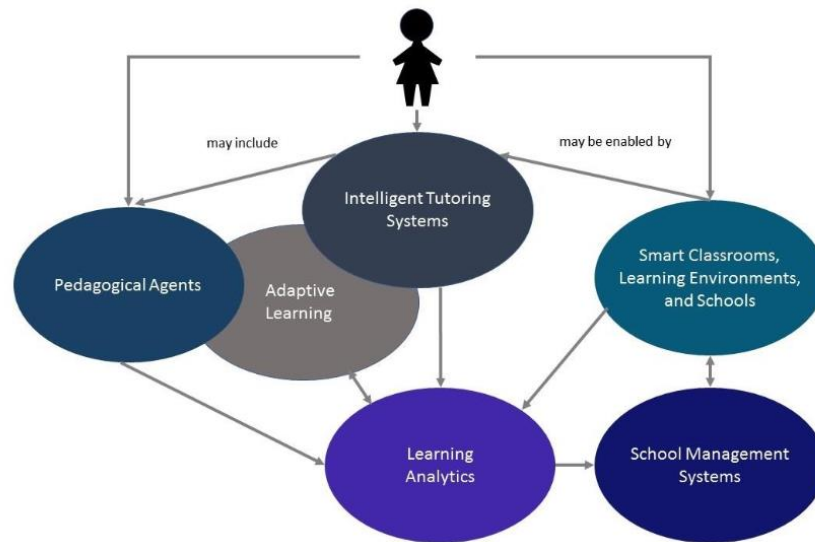


Fig. 2. Overview of typical AIEd applications and their relationships (Southgate et al., 2019).

Pedagogical agents are virtual characters - either text-based or visual - designed to act as digital tutors, mentors, or facilitators that are integrated into educational technologies. They facilitate learning by interacting with students through guidance, feedback, and navigation of learning content while enhancing students' engagement (Heidig & Clarebout, 2011). These agents are often embedded within ITS, providing personalised feedback and adaptive instruction to learners.

Intelligent Tutoring Systems (ITS) simulate one-on-one human tutoring (Luckin et al., 2016). From a technical view, ITS comprises several key components (Southgate et al., 2019): a domain or expert model that stores expert knowledge to assess the correctness of learner responses; a pedagogical model that contains pedagogical and tutorial strategies that guide the system in providing feedback and instructional steps; a learner model that stores learner characteristics and tracks the student's understanding and actions in the ITS system; and an interface that facilitates interaction between the learner and the system, employing written or natural language dialogue, with or without virtual pedagogical agents representing the ITS (Du Boulay, 2016).

The development of these technologies draws on interdisciplinary knowledge spanning cognitive science, psychology, education, and information systems to inform their design and functionality (Southgate et al., 2019). With these models, ITS systems collect data and make inferences about students' knowledge, emotions, or motivation, diagnosing gaps and offering personalised tasks, hints, or feedback to support learning (Koedinger et al., 2013). More sophisticated ITS, such as emotion-aware ITS,

dynamically adapt to students' emotional status (e.g., confidence, frustration, surprise or boredom) by using cameras and sensors with eye-tracking and emotion recognition software. This type of system is also known as a key tenant of smart classrooms or smart learning environments, which use technology to monitor and observe individuals' location, movement, and engagement within learning environments. We will dive into the details of this technology in the next subsection.

Smart classroom technologies integrate hardware (e.g., wearable biometric sensors, motion-sensing cameras, environmental sensors for tracking noise levels), software (e.g., ITS), and Internet of Things devices (e.g., AI-powered facial recognition cameras or eye-tracking techniques for monitoring attendance and student engagement) to create interactive, data-driven learning environments. These technologies often integrate with ITS and pedagogical agents, aiming to empower teachers by enabling real-time diagnosis of students' needs and delivering personalised learning resources and interventions (Baker, 2016). These AI-powered learning environments often capture significant amounts of data on learner performance and engagement, which, through learning analytics, aim to provide insights to guide resource allocation (e.g., teachers' attention), prompt reflection on learning and teaching practices, and improve student learning experiences (Siemens & Long, 2011).

These systems, which actively collaborate with teachers and students in AI-powered classrooms, are central to this thesis. While this type of AIED demonstrates significant potential, researchers have noted that existing ITS implementations in schools often rely on very simple pedagogy and student modelling, such as determining student comprehension based on solving three tasks in a row correctly (Southgate et al., 2019). More technically sophisticated applications, such as emotion-aware ITS, remain largely in pilot testing or feasibility stages (Baker, 2016). These systems, which involve close personal monitoring through technologies like eye-tracking and facial emotion recognition, represent a relatively new area of development within computer science, machine learning, and cognitive psychology. However, these advancements have sparked ethical debate regarding the collection and use of sensitive data, especially for vulnerable populations like students. Such unresolved ethical challenges represent a significant barrier to realising its full technological potential (Southgate et al., 2019).

This thesis focuses on these more sophisticated yet closely monitoring technologies, exploring stakeholders' perceptions and concerns. In the next section, we delve into the

key technologies enabling real-time classroom observation and detection of students' learning status, introducing a central concept – AI sensing technology.

2.1.2 AI Sensing Technologies in Education

AI sensing technologies in education are developed to monitor students' physical behaviours and emotional states to inform classroom dynamics and provide classroom observations. These insights are used for real-time analysis and performance assessments, which are then fed to human teachers and pedagogical agents to offer personalised feedback about student behaviours and suggestions for teaching strategies.

Within the realm of AI sensing, wearable sensors and environmental tools are developed to gather data on students' physiological responses, bodily expressions, and behaviours: (1) Wearable sensors, which involve devices such as wristbands or bracelets equipped with biometric sensors, detect physiological signals such as heart rate (e.g., to identify student stress (Yan et al., 2023)), measure electrodermal activity as an indicator for arousal states (Haataja et al., 2018), or sense skin conductance to monitor emotional arousal (Arroyo et al., 2009). (2) Environmental tools, such as AI-embedded cameras and ambient sensors, are employed to capture and assess students' physical behaviours, cognitive states, engagement levels, and interactions within educational environments (Vierhauser et al., 2024; Wang et al., 2023). Those tools leverage advanced computer vision and machine learning techniques, enabling facial and emotion recognition to monitor students' body movements, facial emotional expressions, and gaze behaviours. This thesis specifically focuses on this type of AI sensing technology.

Data can be obtained through facial emotional expressions such as confusion and frustration (Aslan et al., 2019; Ez-Zaouia et al., 2020); body language, including gestures (Kim et al., 2018; Lyu et al., 2023) and head orientation (e.g., nodding, shaking) (Ahuja et al., 2019; Peng et al., 2023); and eye movements, such as gaze direction, to determine attentional focus (Ahuja et al., 2021; Hutt et al., 2021; Linson et al., 2022).

These data feed into models of students' cognitive-affective and physiological states, which use indicators like body motions (Lyu et al., 2023) and arousal levels (Yan et al., 2023) to assess learning performance. Derived from quantified indicators (Zawacki-Richter et al., 2019), AIED tools interpret in-class behavioural performance (Alyuz et al., 2017) and provide insights into students' real-time emotional and attentional engagement (Aslan et al., 2019; Hutt et al., 2021; Ruan et al., 2023). These systems classify students

into categories such as “engaged,” “disengaged,” “struggling,” (Aslan et al., 2019; D'mello & Graesser, 2013; Nawaz et al., 2018), or even “at-risk” for poor performance or potential drop-out (Namoun & Alshanqiti, 2020). The inferred insights are then delivered to intelligent tutoring systems (Ruan et al., 2023) or to human teachers (Alzoubi et al., 2021; Aslan et al., 2019; Jiménez-Mijangos et al., 2023) for further functions or action.

AI sensing technologies can be deployed in education in two ways: 1) Class-wide sensing systems, which capture data from an entire classroom. For instance, a front-facing camera monitors all students in the classroom simultaneously, analysing both individual students and group dynamics. Existing technologies for class-wide sensing include Lumilo (Holstein et al., 2018), Edusense (Ahuja et al., 2019), TeachActive (Alzoubi et al., 2021), Classroom Digital Twin (Ahuja et al., 2021), and ClassID (Patidar et al., 2024). These systems may be seen as attempting to follow the “co-orchestration” model for teachers and AI tools (Holstein et al., 2019; Martinez-Maldonado et al., 2015; Yang et al., 2023), which aims to synergise human expertise and AI capabilities; 2) Individual-wide sensing systems, which capture data from specific students’ devices, such as laptop cameras (Lin et al., 2014) or wearable sensors like biometric bracelets (Arroyo et al., 2009).

2.1.3 Research in AIEd

Research efforts on designing and advancing AIEd, spanning decades, were intended to address educational challenges and enhance learning. The earliest AIEd research emphasised generating evidence to inform the design of systems that support individual learners. Efforts were dedicated to empirical studies aimed at understanding learners’ needs and how they could be effectively supported through engagement with AIEd systems (Kay, 2012). This learner-centred approach reflected a commitment to prioritising the learner’s well-being, placing their needs at the heart of AIEd research design (Self, 1998). Even a decade later, this focus on the learner persisted, with intentions to understand how expert teachers enhance learning in one-on-one contexts and to translate these insights into digital learning environments.

Over time, the scope of AIEd research expanded to develop precise computational models of educational, psychological, and social knowledge, aiming to make these models a cornerstone of AIEd’s contribution to education (Kay, 2012). The AIEd community began to recognise AIEd’s role within a larger learning ecosystem,

incorporating not just learners, but educators, families, and other external factors influencing the learner's experience. We can observe a shift in trends – from an initial focus on understanding learners' needs to an increasing emphasis on advancing computational models. While these approaches are not inherently in conflict, this shift raises questions about whether technological advancement continues to align with the goal of supporting learner well-being.

However, AIED scholars have long cautioned against an overreliance on data science and computational methods when seeking to innovate educational technologies. They argue that these approaches alone are insufficient for capturing the full depth of individuals' learning experiences and social contexts (Bond et al., 2019; Veletsianos et al., 2015). Learning processes, they assert, are inherently social and cannot be fully reduced to algorithms and computational models (Fynn, 2016; Holmes et al., 2021; Siemens, 2013). This social nature of learning requires qualitative rather than merely quantitative evidence to provide deeper insights into how AIED systems are designed and adopted. This is particularly important in educational settings, where technology use is deeply situated and shaped by emotional experiences and interpersonal relationships (Brown et al., 2007).

Despite these warnings, quantitative research continues to dominate the field of AIED (Fox, 2022). This persisting emphasis on quantitative methods has created a gap in deeper empirical evidence, particularly regarding the nuanced social and ethical dimensions of AIED design and adoption. We therefore need a more balanced approach that includes both quantitative and qualitative data to offer deeper insights into the social context of AIED adoption. This will ensure that advancements remain learner-centred and continue to prioritise the well-being and benefits of learners.

2.2 USER PERCEPTIONS & EXPERIENCES OF AIED

In the field of HCI, it has long been recognised that understanding users' perceptions and experiences of technology is critical to human-centred approaches that align with their values and needs (Aragon et al., 2016; Borning & Muller, 2012). In the context of adopting AI in education, learners as key users are the primary recipients of AIED's impact. However, there is a lack of empirical evidence for understanding students' perceptions and attitudes towards AIED. Current research regarding student attitudes tends to take an exploratory approach to general AI use in education rather than specific

aspects of AIED. User studies on specific AIED technology often focus narrowly on technology performance (Chu et al., 2022), overlooking broader factors. Although there are existing scales for evaluating user experience and attitude, these tools typically have a general focus and fail to adequately capture the unique experience of students within the educational context.

Studies showed that students generally have positive attitudes toward incorporating AI in education (Almaraz-López et al., 2023; Katsantonis & Katsantonis, 2024; Pinto dos Santos et al., 2019). For instance, students advocated for AI inclusion in educational training in fields like medicine (Pinto dos Santos et al., 2019), business, and education (Almaraz-López et al., 2023). They expected benefits from customised learning services such as identifying learning struggles (Gray et al., 2022), providing timely feedback (Schumacher & Ifenthaler, 2018), informing educators' pedagogical decisions (Falcão et al., 2020), and enhancing teaching quality (Gray et al., 2022). However, significant concerns from students also existed, including issues around data privacy (Jones et al., 2020; 2020), transparency and consent (Whitelock-Wainwright et al., 2021) (Whitelock - Wainwright et al., 2019), potential data misuse or misinterpretation (Schumacher & Ifenthaler, 2018) (Gray et al., 2022), and excessive monitoring of learning activities (Schumacher & Ifenthaler, 2018) (I will detail ethical concerns in the next section). While these studies highlighted a range of student attitudes, many were broad and did not focus specifically on any AIED technologies or delve very deeply into the nature of various concerns. In particular, many studies either focused on general AI in education or limited their scope to issues like data privacy. This left a gap in understanding the full range of students' perceptions and concerns in the AIED context.

Various scales specifically assessing student attitudes towards AI adoption in education have been developed. For instance, the General Attitudes towards Artificial Intelligence Scale (GAAIS) (Schepman & Rodway, 2023) evaluated students' attitudes towards AI applications in higher education. It includes two subscales: positive attitudes, which include perceptions of AI's utility and benefits and positive emotions (e.g., excitement); and negative attitudes, focused on unethical uses, errors, and negative emotions (e.g., discomfort). Using this scale, a recent study (Rodway & Schepman, 2023) found that while students were generally comfortable with AI, they expressed discomfort with its use in decision-making, such as grading and admissions. Another scale called Student Attitudes Toward Artificial Intelligence (SATAI) (Suh & Ahn, 2022) specifically

captures students' cognitive, affective, and behavioural responses to AI use in general. Recent findings from a SATAI-based study (Katsantonis & Katsantonis, 2024) indicated that students' emotional responses were the most significant factors shaping their overall attitudes toward AI. However, these scales typically maintained a general focus, leaving a gap in more context-specific investigations into student cognitive, affective, and behavioural responses within AIEd.

Although some empirical user studies have explored AIEd technologies, they typically focused on technology performance and usability (Chu et al., 2022), with limited attention to the broader implications and experiences. Most of those studies used quantitative methods to measure learning outcomes, effectiveness, and engagement as the impact of AIEd technologies on students. They often reported statistically significant improvements in learning (Fontaine et al., 2019), by comparing pre- and post-use scores (Latham et al., 2012; Lin et al., 2014) or task-completion times (Moundridou & Virvou, 2002) with the designed intelligent tutoring systems and affective tutoring systems. Those studies also adopted user attitude and technology acceptance scales, which often reported positive feedback regarding system usability, perceived usefulness, helpfulness, and increased confidence in learning (Latham et al., 2012; Lin et al., 2014; Moundridou & Virvou, 2002). However, the scales commonly used in those studies, such as UTAUT (Venkatesh et al., 2003), are very much technology-centred. Those scales barely dive deeper into the human aspects of learning with AIEd systems, such as the nuanced experience of students interacting with AI in the learning environment, including students' learning status, higher-order thinking skills, self-efficacy, and their specific beliefs and concerns (Chu et al., 2022). There is a lack of deeper empirical evidence of AIEd adoption, especially in understanding users' lived learning experiences and the nuances of their interaction with AIEd.

HCI has underscored the complex situated nature of technology adoption (Fitzpatrick & Ellingsen, 2013; Ismail et al., 2023) and the importance of understanding its impacts and user perceptions, not just on functionality but also on the broader social and emotional facets of people's lives (Kaye, 2006). Furthermore, it is well known that actual user interactions with technology can diverge from what designers anticipate (Preece et al., 2002; Suchman, 1987). Indeed, the value promoted by the system might even contradict and undermine design intentions (Flanagan et al., 2005). This is particularly concerning in the educational context, as any potential adverse effects of

unintended impacts are possibly significant and harmful. However, as I have stressed, the intricate dynamics between students and the evolving AIED environment, as well as the consequences for students, are underexplored. These concerns point to a deeper ethical dimension of AIED adoption - one that demands closer scrutiny. The next section explores how current ethical frameworks engage with these issues, and where critical gaps remain.

2.3 ETHICAL DISCUSSIONS AND FRAMEWORKS

Ethical discussions and frameworks addressing AI in education (AIED) often continue to mirror earlier trends in AI ethics, maintaining a primary focus on addressing data-centric and computational concerns. Gaining an understanding of this overarching AI ethics landscape is essential for positioning AIED ethics within the broader AI field, as well as for leveraging insights from this established body of work. Hence, this section first reviews the general ethical considerations of AI before delving into the specific ethical challenges of AIED. Additionally, this section highlights the limitations of current ethical approaches: a lack of human-centred ethical principles results from a lack of thorough examination of consequences. Finally, this section explores ongoing discussions about the broader implications of AIED technologies.

2.3.1 Current Approach to Ethics of AI and Ethics of AIED

The ethics of AI – an emerging interdisciplinary field spanning philosophy, computer science, law and social sciences - has garnered increasing attention from academia (e.g., (Floridi & Cowls, 2022; Jobin et al., 2019)), global institutions (e.g., (UNESCO, 2021; World-Economic-Forum, 2019)), and government departments (e.g., (Southgate et al., 2019)). In response, a wide range of frameworks and guidelines have been developed by policymakers, researchers, and industry consortia to guide ethical design, development and deployment of AI systems. These are typically grounded in a set of core principles, often drawn in part from bioethics.

Jobin et al. (2019) identified 84 sets of principles or guidelines, distilling five globally recognised principles: transparency, justice and fairness, non-maleficence, responsibility, and privacy. Similarly, Floridi and Cowls (2022) introduced the 5P framework, which consolidates six major ethical AI proposals, including the Asilomar AI Principles (2017), the Montreal Declaration for Responsible AI (2017), and the IEEE General Principles (2017), the ethical AI principles from the European Commission's European Group on Ethics in Science and New Technologies (2018), the UK House of

Lords AI code (2018), and the Tenets of the Partnership on AI (2018). The 5P framework includes beneficence, non-maleficence, autonomy, justice, and explicability, aiming for universal applicability across domains of AI use. Despite their broad scope, these frameworks often lack concrete and practical guidance for real-world AI applications (Holmes & Porayska-Pomsta, 2023) – for instance, how those principles might be actioned in specific contexts (Holmes et al., 2021), such as in education. UNESCO (2021) stresses the importance of contextualising AI ethics by considering the broader environment in which AI operates. Yet, in AI in Education (AIEd), ethical considerations remain peripheral (Holmes et al., 2021).

AI ethics has become increasingly interdisciplinary - moving beyond its historically dominant focus on technical, computational, and data-centric issues (Holmes et al., 2021) such as data privacy (Tsai, Whitelock-Wainwright, et al., 2020), algorithmic bias and fairness (Jiang & Pardos, 2021; Sha et al., 2022), and transparency (Robertson et al., 2021). However, in the context of AIEd, ethical discussions remain largely centred on AI techniques themselves (Holmes et al., 2021). Recent systematic reviews of AIEd literature confirmed this limited focus (Hakimi et al., 2021; Holmes et al., 2021; Murchan & Siddiq, 2021; Tzimas & Demetriadis, 2021), noting that little attention has been paid to the broader socio-technical impacts on students, teachers, and other stakeholders.

This narrow (though important) ethical focus extended to governance frameworks of AIEd and relevant AI-empowered educational techniques (e.g., Learning Analytics). A lot of concrete frameworks addressed data issues, but few offered structural mechanisms to address broader socio-technical and human-centred implications. For instance, DELICATE (Drachler & Greller, 2016) provided a step-by-step checklist for institutions implementing ethical Learning Analytics (LA), emphasising that failure to address privacy concerns risks student resistance or legal challenges. It suggested that institutions can use DELICATE to develop data governance policies to build student trust. However, students' concerns often extend beyond data privacy. There may be further, unintended negative consequences that directly harm students. Other framework, such as the EEAI framework (Southgate et al., 2019), IEAIEd's (The-Institute-for-Ethical-AI-in-Education, 2021), advocated for human-centred values in AIEd, yet their recommendations remained broad and lacked context-specific actionable strategies to mitigate unforeseen ethical risks. Even the SHEILA framework (Tsai et al., 2018), arguably the most comprehensive LA framework grounded in empirical research, focused

on securing stakeholder buy-in and identifying behavioural changes that LA systems should facilitate while overlooking how LA adoption may unintentionally reshape students' learning experiences. This gap was critiqued by SHEILA's architects, with Tsai et al. (Tsai & Gasevic, 2017) emphasising the lack of research on how technology affects the student experience and learning journey. Ignoring these possible unintended consequences and implications risks undermining the trust and acceptance of AIED, as compliance does not guarantee a positive user experience in practice.

Despite students being the primary users impacted by AIED, their role in governance structures and decision-making remained limited. Most governance models were not explicitly designed from a student-centred perspective, and few provided mechanisms for students to voice concerns or to help shape ethical guidelines. For example, Greller and Drachsler (2012)'s Learning Analytics framework introduced a stakeholder dimension, but it defined students merely as data subjects, while teachers and institutions - who use the data to inform educational decisions – were positioned as data clients and primary beneficiaries of LA. The Code of Practice for LA (2016) made a better attempt to adopt a participatory approach and include a student representative in the advisory group in developing the framework. The framework established a taxonomy of ethical issues in LA that accounts for potential adverse impacts on students. However, while this inclusion was a positive step, it still lacked empirical validation grounded in students' lived experiences, leaving a gap in understanding how AIED policies translate into real-world educational settings. Without stronger empirical evidence from students themselves, governance efforts regarding AIED risk being theoretically well-intended but practically misaligned with student needs and concerns.

This section has demonstrated that, while existing AI or AIED ethics frameworks address important technical issues, they fail to account for the real-world context of AIED interactions. Holmes et al. (2021) have created an AIED ethics framework (see Fig.3 below), calling for attention to the “unknown unknowns” in AIED ethics research. This thesis argues that a vital set of “unknown unknowns” is the human-centred ethical issues that arise for students and educators from their direct interactions with AIED systems. However, a significant gap remains in the development of human-centred ethical principles for AIED design and adoption. In the next section, I explore the broader ethical implications of AIED on its users.

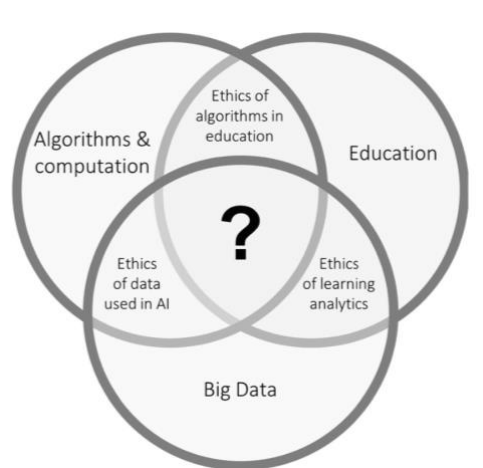


Fig. 3. AIED Ethics Frameworks (Holmes et al., 2021)

2.3.2 Broader Ethical and Sociotechnical Implications of AI in Education

The absence of human-centred ethical principles in AIED is closely tied to a lack of empirical evidence on the consequences of AIED adoption for key users (e.g., students and educators), their lived experience, and the broader educational ecosystem. Although these issues are frequently discussed, they remain scattered and underexplored.

Critics have raised concerns about the potential pedagogical and sociotechnical implications of AIED. For instance, some critics claimed that AI-enabled classroom monitoring and intervention, with a quantified analysis of students' learning and social interactions (Lu, Dillahunt, et al., 2021), may encourage compliance in student behaviour (Williamson, 2019), undermining learning autonomy (Thompson & Prinsloo, 2022), and fostering a surveillance-like atmosphere in educational settings (Cox et al., 2022). Others claimed that the data-driven approach might cause datafication of education (Manolev et al., 2019), which may also promote performativity, where educational practices shift to prioritise metrics over meaningful learning outcomes, often fostering a “league table mentality” (Cox et al., 2022). Furthermore, AIED algorithms may label students as ‘disengaged’ (Nawaz et al., 2018), perpetuating stereotypes (Lawson et al., 2016) and reinforcing biases related to race, gender, and culture (Lu, Marcu, et al., 2021). These biases can have subtle but significant social justice implications (Holmes & Porayska-Pomsta, 2023), yet they remain poorly addressed in existing literature. These critiques often lacked a robust empirical foundation, leaving many questions about the socio-technical consequences of AIED unanswered.

As Holmes and Porayska-Pomsta (2023) noted, most of the AIED literature has proceeded without serious engagement with its potential ethical consequences. As

introduced in Section 2.2, even though there are user studies from AIEd research, they typically focus on system functionality, rather than deeper user experience, such as students' learning journey. An understanding of specific ethical issues raised by AIEd is only beginning to emerge. However, there is no comprehensive framework, guidelines, or policies that have been developed to address ethical concerns (Holmes & Porayska-Pomsta, 2023). The ethics of AIEd is more nuanced than simply being 'good'; what constitutes an ethical approach remains uncertain (Holmes & Porayska-Pomsta, 2023).

A human-centred approach offers a promising solution to this gap, as it focuses on the perspectives and values of the individuals directly affected by AIEd, namely students and teachers. By aligning technology design with the values, lived experiences and well-being of its users, this approach ensures that AIEd systems support the needs of students and educators in meaningful ways.

Ultimately, a holistic, empirical understanding of user experience, particularly from the student's perspective, is crucial to comprehending the full scope of AIEd's impact. Such an understanding will be foundational in shaping ethical frameworks that guide AIEd development and adoption, ensuring that it serves both educational and ethical goals.

2.4 SUMMARY – THE NEED FOR HUMAN-CENTRED ETHICS

This chapter has reviewed AIEd advancements, particularly AI sensing technologies, which are becoming increasingly complex and reliant on sensitive user data. Existing research developing AIEd largely focuses on technological performance and system effectiveness while neglecting users' lived experiences and their nuanced interactions with AIEd (Gap 1). Furthermore, there is a lack of empirical evidence on how AIEd is adopted in real-world educational settings and its broader consequences (Gap 2). These two gaps contribute to a third fundamental issue: the lack of empirical understanding needed to inform human-centred ethical principles to guide AIEd design and governance. As a result, existing ethical frameworks remain primarily technical-oriented, failing to address the socio-technical and human-centred ethical concerns of AIEd (Gap 3).

To address these gaps, this thesis takes a human-centred approach to understanding the ethical challenges of AIEd adoption. While current ethical discussions and frameworks focus predominantly on data governance and computational concerns, this

study argues that the ethics of AIEd extend far beyond these technical-ethical aspects. The broader socio-technical and pedagogical implications of AIEd on users' interactions and learning experiences remain underexplored. Therefore, rather than applying predefined normative ethical frameworks to identify ethical challenges, this study adopts a bottom-up approach, deriving ethical insights directly from stakeholders' lived experiences. This thesis explores how different stakeholders perceive and respond to AIEd and its impacts. By providing an empirical understanding of AIEd's ethical consequences and challenges from the perspective of direct users, particularly students, this research aims to inform the human-centred ethical principles for AIEd design and governance.

This thesis seeks to answer the research question: *What are the ethical challenges in adopting AIEd, from the perspectives of different stakeholder groups?* This inquiry also shapes the methodological and epistemological stance of the study. Grounded in constructionism, this research acknowledges that meaning is constructed through individuals' perceptions and interactions with the world rather than being objectively predefined. This approach allows for a deeper exploration of how stakeholders interpret and navigate AIEd technologies, which will be further detailed in the next chapter on methodology.

Chapter 3: Research Approach

Following the identification of key gaps in the literature review, this chapter introduces the research design and methodology developed to address the research questions that emerged from these gaps, as outlined in Chapter 1:

Overarching Research Question: What are the ethical challenges in adopting artificial intelligence in education (AIEd) from the perspectives of different stakeholder groups?

RQ1: What ethical challenges do students perceive in the adoption of AIEd?

RQ2: How do stakeholder dynamics shape ethical risks in AIEd adoption?

RQ3: What are students' affective responses, beliefs about learning, and prioritised ethical values respecting AIEd design and implementation?

This chapter goes beyond merely outlining the research methodologies - the “how” of this research has been developed and unfolded. It also offers a personal reflection on the “why” behind my design choices and the “what” of the struggles and gains I encountered throughout the research journey.

As such, the structure of this chapter mirrors the progression of my reflective process, starting from Section 3.1, which presents the overall strategic design thinking of the research. It then follows Crotty (1998)'s four elements to “scaffold” the choices of research process design and justify the methodologies and methods: the epistemology; theoretical perspective; methodology; and methods. Section 3.2 discusses the epistemological stance – constructionism, which this research primarily leans on. Section 3.3 discusses the theoretical underpinnings – symbolic interactionism that informed the study's approach. Then, Section 3.4, Section 3.5, and Section 3.6 detail the methodology and specific data collection and analysis methods for study 1, study 2, and study 3, respectively. Finally, Section 3.7 summarises this chapter.

3.1 OVERALL STRATEGIC DESIGN THINKING

As introduced in Chapter 1, this PhD research is motivated by my aspiration to prioritise students' concerns in adoption decisions and ensure that future AIEd adoption considers their values and well-being. Reflecting on my discomfort at being monitored

and analysed while learning, I questioned why such adoption decisions are still being made despite potential students' dissatisfaction. Or why is technology being so enthusiastically developed and implemented without addressing students' concerns? These two questions led to my first and second studies – revealing students' concerns, and examining whether powerful stakeholders, like school decision-makers, developers, and parents, have effectively considered and addressed students' concerns. With this purpose, this thesis primarily leans on a constructionism approach, which I will detail in the next section.

Originally, my third study aimed to understand or resolve the potential value tensions between students and more powerful stakeholders, with the hope of achieving a model for more student-friendly AIEd design and adoption. However, the second study revealed the value tensions and power dynamics that I wished to understand earlier than expected. As a result of this boon, I started to reconsider my approach for my last study. I looked back at my previous two studies and found both are problem-oriented, focusing on exposing the ethical issues and gaps in current practices. I asked myself what would truly bring value to the final phase of this research. Value tensions among stakeholder groups seemed like a long-existing common problem in the field of HCI and ethics studies - too complex and far-reaching to solve in one study. Furthermore, AIEd introduces additional layers of complexity, such as issues surrounding explainability, transparency, and data ownership, which further complicate ethical decision-making and stakeholder alignment. These considerations led me to explore various possibilities: Should I create frameworks to empower less powerful stakeholders? Propose AI literacy initiatives for schools? Offer regulatory insights for restricting vendors, or ask developers to better anticipate risks? Should I focus on educating parents and students, and ensuring informed consent for the potential risks of AIEd? While these are valuable, I returned to the core reason that inspired this PhD: students' voices are often neglected.

What I ultimately wanted to contribute was a detailed and nuanced understanding of what students care about when we adopt AI in their learning - their concerns, preferences, and ethical priorities. I wanted those voices to be heard and recognised, and to inform future AIEd design. Rather than drawing conclusive solutions, I want my research to offer suggestive insights. Therefore, I decided to pivot my original aim of the third study - from resolving tensions to validating students' concerns and uncovering their ethical priorities. To do this, I chose a pragmatic approach with an experimental study.

Table 1 below provides an overview of the study approaches. In the following subsections, I introduce the overall epistemology of this thesis, as well as the theoretical lens and methodological approaches for each study. While these sections are presented in a sequence from epistemology to specific methods, this research was not initially driven by epistemological assumptions. As I stated earlier, this research grew from specific research aims and questions, which then informed the choice of methodology and data collection methods. The epistemology and theoretical perspectives were developed and articulated progressively as the research unfolded and progressed with my ongoing reflection.

Table 1. Summary of research approach and study components of this thesis

Epistemology	Constructionism		
Theoretical Perspective	Interpretivism - Symbolic Interactionism		Pragmatism
Studies	Study 1 Students' perspectives	Study 2 Powerful stakeholders' perspectives	Study 3 Students' responses and ethical priorities
Finding Chapter	Chapter 4	Chapter 5	Chapter 6
Methodology	Qualitative	Qualitative	Experimental Research
Methods Data Collection	Story Completion Method (n=71)	Semi-Structured Interviews (n=30)	Survey (n=132)
Methods Data Analysis	Reflexive Thematic Analysis; Grounded Theory Lite	Reflexive Thematic Analysis	Statistical Analysis; Inductive Thematic Analysis

3.2 OVERALL EPISTEMOLOGICAL STANCE: CONSTRUCTIONISM

Epistemology refers to the way of understanding and explaining how we know what we know and what it means to know. It deals with ‘the nature of knowledge’ (Crotty, 1998), its validity, scope, and general basis, as well as methods of acquiring knowledge. By introducing epistemology in this section, I explain how I, as a researcher, frame my research in the pursuit of knowledge, and how the reader can interpret the knowledge and evaluate its outcome.

My general way of understanding the world is with constructionism – truth or meaning emerges with human beings interacting with and interpreting the realities of the world. As Crotty (1998) noted, “There is no meaning without a mind. Meaning is not discovered but constructed.” This stance rejects the objectivist belief that there is an objective truth waiting to be discovered. Rather, in the constructionist view, meaningful

reality is developed by human practices within a social context. These contexts are not fixed – they are dynamic and situated within specific cultural, temporal, and experiential circumstances. As a result, knowledge or meanings might not be universal. They vary and evolve with changing interactions.

This dynamic and interpretive process underpins how I approach this research, especially within the field of Human-Computer Interaction (HCI). In HCI, users actively construct their understanding and experience of technology through their interactions with the technology, environments, and other users. Applied to AIED adoption, this constructionist perspective views reality and meaning as co-constructed. They emerge through the ongoing interactions between various stakeholder groups (students, parents, educators, and institutions) and with the AI technologies embedded within the educational ecosystems.

The educational context for AI adoption is uniquely complex. It involves diverse student populations shaped by intersecting factors such as age, gender, sexual identity, race, and culture. It is further affected by layered power structures involving multiple stakeholders, each with distinct values and priorities. Individuals' interactions with AIED systems are shaped by their backgrounds, leading to varied and potentially unpredictable specific circumstances and responses. Moreover, as the educational environment evolves by adopting ever-evolving AI technologies, so too do the interactions and meanings construct around them. While AIED technology may exist as a physical object in the world regardless of being observed by a conscious mind, the meaning associated with it – How AIED is applied, valued, and perceived - is shaped through human interaction and the sociocultural understanding in which it is used.

Ethics is a central focus of this thesis. Ethical understandings around AI are socially constructed and vary across time and cultural contexts (Wong, 2020). Ethical values and interpretations are shaped by individual lived experiences and broader societal norms. While much of the AI ethics discourse is governance-oriented - focusing on frameworks and strategies to manage risks (Corrêa et al., 2023) - this research contributes a bottom-up approach. It begins with individual perspectives, then identifies broader patterns, and ultimately contributes to a more collective ethical understanding. This form of inquiry echoes constructivism (Denzin, 2012).

Given these research purposes - to explore how stakeholders interpret and respond to AIED and its ethical concerns - the overall epistemological stance leans toward

constructionism². This approach is driven by a desire to understand the complex and multifaceted realities of technology adoption, rather than seeking to predict a singular truth. As Crotty (1998) noted, constructionist research does not claim to offer the only “true” way of understanding phenomena, as (post-)positivist studies might frame their research. Instead, the value of constructionist work lies in providing plausible, insightful, and contextually meaningful interpretations of the defined topic, grounded in the interactions and social dynamics it investigates.

From this perspective, the quality and value of my qualitative research could be evaluated through the depth of understanding it offers. Qualitative studies prioritise the identification and discussion of themes over numerical significance (as quantitative studies typically do) (Braun, Clarke, Hayfield, & Terry, 2019; Pyett, 2003). The rigour of constructionist work emphasises transparency in the analytical process and the production of “thick descriptions” supported by detailed examples of raw data (Seale & Silverman, 1997). Consistent with this approach, I aim for the research outcomes to be suggestive rather than conclusive, aligning with the principles of constructionism.

3.3 THEORETICAL UNDERPINNING – SYMBOLIC INTERACTIONISM

A theoretical perspective provides a way to understand the human world and social life within that world, and make sense of it by providing context of the process and grounding the logic and criteria within that world (Crotty, 1998). This section outlines how the constructionist epistemology furthers the symbolic interactionist theoretical foundation of this study.

As a specific theoretical perspective within constructionism, symbolic interactionism is a social theory that perceives the world as socially constructed through the meanings individuals assign to events and interactions (Crotty, 1998). It assumes that people create meaning for elements in their environment and respond to it through symbolic communication such as language, gestures, and actions. It emphasises micro-

² I adopted “constructionism” rather than “constructivism” – two closely related theories but differ in their focus, particularly in epistemology and research paradigms. Constructivism emphasises how individuals construct meaning from their personal experiences. Constructionism builds on constructivism by adding a social dimension, emphasising that knowledge is constructed not just individually but also socially, through interaction with others and external artifacts such as technology (Crotty, 1998). My research, which involves multiple stakeholder groups interacting and negotiating ethical concerns, considers the role of AIED (as artifacts) in mediating these interactions, shaping stakeholders’ perceptions of ethical challenges and influencing social dynamics. These characteristics align strongly with constructionism, which forms the theoretical basis for my study.

level social interactions, where individuals constantly interpret and re-interpret meaning in response to evolving social situations. In this way, reality is seen as constructed through dynamic interactions and negotiated meanings.

A central notion of symbolic interactionism is role-taking: placing oneself in others' positions to understand their perspectives. Methodologically, this requires researchers to adopt the standpoint of those being studied (Denzin, 1974) to see the social world from their perspective. This approach encourages researchers to understand how individuals construct their realities through their roles and interactions.

Since this thesis explores stakeholders' ethical understanding of AIED adoption and takes into account each individual's experience and social circumstances, constructionism and symbolic interactionism offer an effective lens for perceiving the social foundations of AIED. Especially, I examine how individual students, teachers, and AI co-create the education environment, focusing on the meaning and roles they attach to interaction patterns that can shape educational experiences and norms. With this approach, my personal experience and background also shape how the research is conducted. I am currently a student and educator (I worked as an academic tutor alongside my PhD). I am a former AI industrial professional (I completed my bachelor's and master's in information systems and worked for an IT consulting vendor and AI company as a project manager and product manager). With this identity, I apply symbolic interactionism's encouragement to empathise with various stakeholder perspectives. This theoretical grounding informs the intensive use of reflexive thematic analysis, which I will detail in section 3.5.

In the following sections, I will outline my choice of research method in the sequence of Study 1(section 3.4), Study 2 (section 3.5), and Study 3 (section 3.6). While I have detailed the research methods in each paper (see later chapters), this section focuses on my design thinking and reflections.

3.4 STUDY 1 – STORY COMPLETION RESEARCH

Study 1 was designed to answer sub-question 1: *What are the ethical challenges of implementing AIED from students' perspectives?* Study 1 adopted the story completion method to collect data to explore students' responses to AIED adoption. I initially used thematic analysis to analyse the data, then adopted grounded theory-lite.

3.4.1 Story Completion Method

Story completion is a method rooted in speculative and creative research approaches, enabling participants to project their beliefs, feelings, and assumptions through completing open-ended story prompts (Lawson et al., 2015). This method is particularly effective for uncovering socially constructed meanings and exploring complex and sensitive topics in an indirect and imaginative manner (Clarke et al., 2019). Given the ethical and personal nature of concerns around AI in education, story completion offers a valuable way to access students' deeper, and possibly unspoken, attitudes toward the technology. Furthermore, with AIED technologies still in their nascent state and few mature applications available, this approach facilitates proactive explorations for potential concerns before widespread adoption and any potential negative impacts emerge.

This approach aligns well with the constructionist epistemology underpinning the study, as it allows participants to actively construct and narrate their perspectives on AI within a hypothetical context. By engaging participants in speculative scenarios, this method creates rich data that reveals nuanced socially embedded concerns about AIED adoption – concerns that might otherwise go unexamined.

The speculative and indirect nature of story completion introduces certain methodological challenges. One key risk was to ensure the story prompts were realistic and sufficiently grounded to ensure participants' responses were relevant and meaningful to the research aims. To address this, I crafted three story stems anchored in existing AIED literature that reflect current advancements, to provide enough contextual detail to immerse participants in the scenario. The three scenarios, each progressively delving deeper into the implications of AI-driven tools in education, are provided in the publication appendix presented in Chapter 4.

Another challenge was fostering participants' imagination and engagement while balancing creativity with maintaining relevance to the research scope. To address this, I intentionally incorporated ambiguity in the scenarios to encourage imagination, allowing participants the space to project their own interpretations and concerns (Braun, Clarke, Hayfield, Frith, et al., 2019). For example, I intentionally left the AI's mechanisms and criteria for student evaluation vague to encourage participants to engage more deeply with the ethical and emotional dimensions of the scenarios. To further focus their responses, I offered hints in the task description to encourage affective, behavioural and cognitive

reactions (e.g., asking “If you were the main character in the scenario, how would you feel and behave?”), and suggesting a minimum time of 5 minutes per task. This immersive approach was validated by three pilot participants, who found the tasks engaging and produced rich, valid data, demonstrating the effectiveness of the method in eliciting relevant insights.

There is also a challenge that speculative prompts might inadvertently lead participants toward specific concerns or biases. To mitigate this, my research team and I carefully selected neutral language in the scenarios to avoid suggestive or leading phrasing. Additionally, the introductory materials emphasised that the goal was to understand participants’ genuine feelings and perspectives, rather than to critique the technology. These measures help ensure that the findings authentically reflect participants’ own values, beliefs, and concerns, rather than the researcher’s assumptions. I will further reflect on the story bias in the later section in Chapter 4.

3.4.2 Participant Recruitment – Prioritising Diversity of Opinions

Participants in this study were 71 university students and recent graduates. Initially, I recruited participants with strong storytelling skills, such as those studying film and screenwriting, to ensure rich narratives. While their responses provided valuable philosophical depth, I recognised the potential risk of bias if my dataset came from a large number of this participant group, due to their socially critical perspectives. To address this, I broadened the recruitment to include participants from diverse academic and personal backgrounds, ensuring a wider range of opinions. Reflecting on the essence of qualitative research and the goal of exploring ethical considerations, I realised that quantitative representativeness was less relevant. Thus, my primary focus shifted to gathering a broad, diverse spectrum of viewpoints rather than ensuring strict demographic proportions.

A related concern arose regarding participants with backgrounds in AI or related fields. I briefly considered excluding students from AI or IT majors to avoid potential bias, as I suspect their technical knowledge might lead them to overestimate or underestimate the ethical issues compared to non-technical participants (I incorporated this element of participants who are AI-literate to compare their responses to AIEd features in study 3). However, I ultimately decided not to filter participants based on their majors or AI experience, recognising that in real educational settings, students have diverse backgrounds and levels of AI literacy. This decision aligns with the research goal

of capturing a broad spectrum of concerns from all types of students, reflecting the diversity present in real-world university contexts.

3.4.3 From Thematic Analysis to Grounded Theory Coding

The findings were initially analysed from the thematic level by adopting Thematic Analysis (TA) (Braun & Clarke, 2019). I sorted the codes into initial themes and sketched their relationships on mind maps (Braun & Clarke, 2006). I found that many identified codes and themes interact, and TA could not support interpreting this intricate interplay.

To navigate this complexity, I transitioned to employing the techniques from Grounded Theory (GT), following Glaser and Strauss (2017), renowned for its capacity to elucidate relationships, particularly consequential ones, among concepts and codes (Walker & Myrick, 2006). Rather than adopting the full grounded theory, which aims at theorising or developing new frameworks grounded in data, I adopted the most frequently used form of grounded theory, known as grounded theory-lite. A grounded theory-lite involves using GT coding and relevant techniques to develop a set of categories, aiming at understanding the relationship between various categories to find hierarchy, shared, breaching/disagreed codes (Pidgeon & Henwood, 2013). The detailed coding and data analysis procedures have been introduced in the paper in the following sections, while this chapter primarily focuses on my reflection while conducting data analysis.

Challenging the Dogmatism of Grounded Theory Coding

During the process of adopting the Grounded Theory (GT) structure coding scheme to analyse my data, I reflected on the challenges of adhering strictly to purist GT frameworks. I encountered several practical issues that led me to question the rigidity of the methodology. For example, some codes did not fit neatly into a single category, while others were symbiotic, belonging to multiple categories simultaneously. In some cases, an open code could also serve as an independent axial code, or sub-codes were interrelated and tagged each other, even though they were located in separate higher-level codes. It is complex to assign certain data points to distinct categories due to their messy and interdependent relationships.

In practice, I found that not all data could be clearly divisible or categorised without losing its contextual richness. To address this, I occasionally deviated from the strict instructions of the GT coding structure, allowing for a more fluid approach that better captured the complexity of the data. This decision stemmed from the realisation that strict

methodological adherence can sometimes feel counterproductive when dealing with nuanced human data, causing the data to lose its meaning.

Ethically, I believe that pure methodological frameworks do not always align with the realities of research practice. Forcing participants' narratives into rigid categories risks being disrespectful to the depth and context of what they shared. It is important to honour the nature of participants' voices rather than impose artificial structures that may distort their meaning. By embracing a more flexible approach, I aimed to remain faithful to the integrity of the data and the experiences it represented, ensuring that my analysis was both ethically and contextually grounded.

Reflections on Applying Ethical Frameworks in Data Analysis

During the data analysis process, I initially attempted to align my findings with existing frameworks from AI ethics or AIED ethics as a way to provide theoretical grounding. However, this approach revealed significant challenges. As I emphasised in the previous chapters, current ethical frameworks predominantly address technical concerns and computational issues (Holmes et al., 2021) such as data privacy, algorithmic bias, transparency, and fairness. While these governance-oriented, legal-centred frameworks aim to prevent data-related or algorithm-related issues through regulation and principles, they rarely cover issues that grow from the nuanced interactions between users and the technology. In contrast, my focus highlights ethical concerns derived from users' lived experiences. The themes identified in my analysis, particularly the pedagogical impacts and sociotechnical implications such as students' emotional responses and potential behavioural change, extend beyond the scope of traditional ethical frameworks. I became concerned that forcing these findings into rigid structures could dilute the original objective of revealing students' concerns and diminish the focus of the discussion.

This realisation shaped how I approached the analysis. Instead of strictly adhering to pre-existing frameworks, I allowed the data to highlight ethical dimensions that governance-level concerns often overlook. By doing so, my research highlighted that governance-focused principles are necessary but insufficient. It also addressed a critical gap in AIED ethics, as highlighted earlier: the lack of validation of ethical impacts from the perspective of actual users, particularly students. This shift not only brought clarity to my analysis but also underscored the importance of grounding ethical considerations in real-world, context-specific concerns rather than abstract principles.

3.5 STUDY 2 – INTERVIEW STUDY

After revealing students' concerns about AIED adoption in depth, I sought to interrogate whether other key stakeholders – those with decision-making power - effectively considered or addressed these concerns. While study 1 focuses on higher education, study 2 shifted to the K-12 context, adding complexity to the decision-making dynamics by involving multiple stakeholders, including parents (who act as decision-makers for their children), school governors (who approve decisions on technology adoption in their schools), and industrial vendors (who develop and implement the technology). This shift was intentional, as K-12 students are under 18, making them more vulnerable. The study was designed to answer the sub-question 2 “*How do the complex stakeholder dynamics increase ethical risks?*”, by exploring the ethical considerations of responsible stakeholder groups (e.g., parents, school governors, developers) regarding AIED adoption. To gather perspectives, I conducted semi-structured interviews with 30 participants, 10 for each stakeholder group.

3.5.1 Semi-Structured Interview

Initially, I considered workshops or focus groups, as group discussions can reveal shared or conflicting attitudes among participants. However, the diversity within each stakeholder group posed challenges for these methods. For instance, parents have varying experiences based on their children's school types (e.g., rural vs. urban), while school representatives include individuals with vastly different responsibilities, such as IT admins or experts, teacher representatives, and principals. Similarly, developers' roles in designing, developing, or implementing AIED differ significantly. These variations made semi-structured interviews a more suitable choice, allowing for a deeper exploration of one's lived experiences, values, concerns, and reflections in individual contexts (Seidman, 2006).

In line with the constructionist commitment to understanding and interpreting data, the interview questions were mostly open-ended and flexible (Bearman, 2019; Creswell & Clark, 2017), allowing participants to construct meaning based on their lived experiences. As interviews are an active, co-constructed process (Terry & Hayfield, 2021) created through the conversational relationship between participants and the researcher (Kvale, 2009), the knowledge generated is collaboratively shaped through the interplay of questions and answers and “co-authored by interviewers and interviewees” (Kvale, 2009).

My background as an IT project professional, educator, and student, allowed me to build rapport with my participants, fostering a conversational tone that allowed for a rich exchange of perspectives. The interviews were designed to provide both descriptive insights into everyday experiences and epistemological reflections, which were discursively explored and justified through the interview conversation (Kvale, 2009).

To ensure ethical rigour, I carefully designed and refined interview questions to minimise bias. For instance, participants were first asked about general experiences with AIED, avoiding prompts that could steer them toward positive or negative responses. Follow-ups encouraged them to consider both sides if their answers were one-sided. When discussing the potential issues, I always prepared questions to encourage a balanced view, such as asking about opportunities, positive outcomes, or benefits. I also avoided leading questions. For example, instead of directly asking, “Who is responsible for potential ethical harm of technology adoption?” I rephrased it to, “Did you encounter any ethical challenges in your role?”

I piloted the interview process with one parent, one school governor, and one developer. I realised that participants without prior experience with AIED found it difficult to imagine specific scenarios using AIED. To facilitate more focused and consistent responses, I created AIED-specific artifact introductions aligned with the story stem scenarios from Study 1 (which I also used for Study 3). At the start of each interview, participants were asked to imagine their role in a given context: Developers - Imagine your client requests you to design and develop this AIED system; School governors - Imagine your school is planning to implement this AIED system; Parents - Imagine your child’s school is adopting this AIED system. This approach grounded three stakeholder groups’ responses in the same scenario more effectively.

The semi-structured interviews were conducted with 10 school governors, 10 parents, and 10 industry vendors, each lasting 45-70 minutes. I tried to create a non-judgmental and supportive conversational environment. At the beginning of the interview, I assured participants - particularly vendors and school governors - that their responses would remain anonymous and would not represent their organisations, encouraging them to share their views freely. During the interviews, while I adhered to the interview agenda to ensure all key questions were addressed, the discussions were guided by what I interpreted to be meaningful to the interviewees. This approach allowed conversations to flow naturally, often weaving in and out of different predefined topics and questions

(Byrne, 2022). I took notes during the interviews and documented my thoughts and reflections immediately afterwards. These notes significantly informed the interpretation and final analysis of my dataset.

3.5.2 Reflexive Thematic Analysis: From Frustration to Reflexivity

Study 2 analysis followed a Reflexive Thematic Analysis (RTA)³ approach. The data analysis process was iterative and reflexive, involving multiple rounds of refinement, self-reflection, and navigating uncertainty in method application. This section reflects on the evolution of the analytic process, the challenges encountered, and the insights gained.

After completing all the interviews, I began with an impression of shared and conflicting concerns among developers, school governors, and parents. I initially observed patterns that pointed to structured stakeholder interactions and power dynamics influencing AIED design and adoption decisions.

At first, I attempted to adopt Activity Theory (AT) (Engeström, 2000) as a framework to organise these findings to understand the systematic phenomena of AIED adoption. AT provides a lens to examine technology adoption as a dynamic, multiple-stakeholder process shaped by social, institutional, and material factors, and helps identify tensions that emerge when new technologies are introduced (Engeström, 2000). Using AT, I categorised codes and themes into “design and implementation” and “post-adoption” phases and mapped them against AT’s structural elements (e.g., Tools, Rules, Community, Division of Labour). While this framework offered a systematic way to organise the data, it felt overly mechanical. The resulting themes and storylines were broad, disconnected, and failed to capture the nuance of the data. Recognising these limitations, I shifted to a more interpretive approach, Reflexive Thematic Analysis (RTA) (Braun & Clarke, 2019).

I re-assembled the codes and reframed my analysis around key interpretive questions, as a starting point for constructing a logical argument: a) How do (ethical) considerations from each group influence the collection and implementation of design needs, and what is missing in their focus? b) What tensions and contradictions arise between stakeholder groups during AIED adoption? c) What are the power structures

³ Although Interpretative Phenomenological Analysis (IPA) was considered, this study adopted Reflexive Thematic Analysis because the aim was to map and theorise shared patterns of meaning across a diverse dataset rather than conduct idiographic, phenomenology-driven exploration of individual lived experience, for which IPA is primarily suited (Braun & Clarke, 2019).

among stakeholders, and what role do they play in decision-making? While this approach provided a more cohesive narrative, the findings still felt shallow. I felt the full potential of the dataset was untapped. I then refined the questions further to explore: a) What ethical issues are overlooked (covertly or overtly) by each group, and what risks do they pose? b) How do power dynamics and hierarchical relationships influence ethical AIED design and adoption? c) What gaps exist in stakeholders' expected versus actual roles and responsibilities? These refinements allowed me to identify richer themes, such as explicit value tensions and implicit gaps in knowledge and capacity. However, after multiple iterations, the analysis still lacked analytical depth. I found myself over-coding the data and struggling to generate meaningful insights.

A turning point came when I fully embraced the interpretive orientation of RTA. Rather than merely categorising data into descriptive themes, I realised that RTA required a deeper engagement with data - blending my observations and interpretations into the analysis. The themes need to move from "domain summaries" to a more integrated, narrative-driven approach (Braun, Clarke, Hayfield, & Terry, 2019). This shift enabled me to craft a storyline, connect patterns across stakeholder groups and articulate the deeper structural and ethical tensions underlying AIED adoption.

Crucially, I recognised my active role in shaping the analysis, consistent with the constructionist epistemology. I, as a researcher, with my own values, background, and interests, could reflexively influence the data interpretations and knowledge production (Braun & Clarke, 2019), rather than simply sitting objectively outside the data to uncover pre-existing truths (Terry & Hayfield, 2020). While recognising this role, I also aimed to faithfully take into account my participants' experiences and perspectives.

In this process, I considered *positionality* as a central analytics tool in RTA (Trainor & Bundon, 2021). Positionality helped me refocus my analysis by integrating my positional stances into data interpretation. I reflected on my dual perspectives - as a student, deeply concerned about the unintended negative consequences of AIED and the neglect of students' voices; and as an IT project professional, familiar with the practical constraints vendors face. While I strongly wished that students' concerns, which I revealed from Study 1, could be well covered and addressed, I also empathised with the vendor's challenges. This reflection clarified the core tension that interested me most in this research: addressing ethical issues without considering practical realities risks

remaining in a vacuum. The gap between ethical ideals and real-world project delivery became a critical lens for understanding the dataset and anchoring the analysis.

Another key aspect of my reflexive process involved revisiting *emotional and non-verbal cues* from the interviews. These moments, marked by surprise, frustration, or recognition of unexpected patterns, often signalled valuable findings. For example, the tone-shifting when vendors were asked who should bear responsibility for ethical consequences post-adoption, as when one vendor candidly admitted, “actually, I don’t give a shit, this is the common truth [because of a tight project deadline with financial pressure].” Vendors’ detachment from ethical considerations post-delivery, together with schools’ poor knowledge capacity about AI issues, revealed significant gaps in accountability. Similarly, parents expressed resignation to systemic challenges, likening AIED adoption to mandatory vaccination policies where individual agency is limited in the specific cultural context. These moments of surprise became analytical anchors for uncovering deeper underlying tensions.

Through six rounds of iterative refinements, the analysis transitioned from fragmented coding to a more cohesive analytical narrative, moving beyond surface-level observations to explore the underlying dynamics within the stakeholder ecosystem. Throughout this journey, I documented each stage of theme development with encountered frustrations and reflected critically on my analytical decisions. This documentation brought clarity to my findings and ultimately enabled a richer interpretation of the lived realities of the participants, particularly the stakeholder dynamics and the ethical gaps in AIED adoption decision-making.

In retrospect, this iterative journey exemplifies the essence of Reflexive Thematic Analysis, which emphasises the evolving nature of research questions and the importance of maintaining a dialogue between data patterns and analytical focus (Braun & Clarke, 2019). My approach aligns seamlessly with RTA’s principles, where initial research questions serve as an open “starting point” that might become more focused, or expand, or even shift in focus, as data collection and analysis progress. Honing and refining research questions are not indicators of poor-quality practice or design, but instead reflect a process through which deeper insight and arguments are generated (Byrne, 2022).

3.6 STUDY 3 – EXPERIMENTAL STUDY

Study 3 design intended to amplify students' voices by validating their concerns, preferences and ethical priorities regarding AIED adoption. Building on the themes from Studies 1 and 2, Study 3 aimed to test earlier conclusions, extend insights, and integrate findings from the previous studies to offer a structured examination of how students perceive AIED technologies in use. Study 3 addresses sub-question 3: *“What are students' affective responses, beliefs about learning, and prioritised ethical values respecting AIED design and implementation?”*

Study 1 and Study 2 highlighted a disconnect among stakeholders' considerations. Particularly, school representatives and parents expect to compromise students' privacy and emotional discomfort for perceived educational gains (e.g., improved scores), despite recognising students' potential resistance. Study 3 sought to investigate whether students agree with such trade-offs and which values, such as privacy or educational gains, they prioritise most.

To achieve these objectives, I adopted a pragmatic approach, using mixed methods in an online experimental study, integrating insights from Studies 1 and 2. One rationale for this experiment design stemmed from the realisation that not all AIED features elicit negative responses; some may be acceptable or even desirable to students. The experiment was designed to identify which features students favour and which they reject, providing clarity on their ethical priorities.

I carefully considered this transition from a purely qualitative approach to an experimental approach, with a methodological reflection: qualitative research, in my view, serves as a critical preliminary stage when little is known about a subject or phenomenon, uncovering insights from people's experiences, perceptions, and opinions. Once these insights take shape, quantitative methods can validate, refine and complement them, and strengthen the robustness of the research on a given topic. With Studies 1 and 2 being exploratory and revealing potential concerns and gaps, Study 3 aimed to validate these findings, offering deeper insights into the challenges of ethical AIED design and adoption.

The experimental study⁴ consisted of two parts. The first part focused on students' responses to different AIED features. I created video prototypes aligned with the prompts used in Studies 1 and 2, incorporating independent variables representing various AIED features. The independent variables were awareness of the sensing technology's presence, detection method (gaze-based attention detection vs. facial-based emotion detection), and teaching intervention (system-generated hints vs. human tutor assistance). While designing the survey instruments for dependent variables, I reviewed existing scales measuring user acceptance and attitudes. However, I found those scales were often shallow, emphasising task-oriented system functions rather than capturing users' deeper experiences, such as affective responses. To address this gap, I developed instruments that integrated themes from Study 1 that reflect students' specific concerns, existing theories or cognitive models for user attitude and technology acceptance, and user emotional reaction to technology models. At the end of this part, I included an open-ended question to invite participants to elaborate on their responses in relation to all of the prototypes viewed.

The second part of the experiment explored students' ethical priorities. Participants were presented with ethical terms related to the AIED features shown in the prototypes and asked to rate their level of concern for each term. They were then asked to rank these ethical terms in order of priority. This experiment was inspired by the Moral Machine Experiment (Awad et al., 2018), which asked participants to make moral decisions in hypothetical autonomous vehicle scenarios and examined how individuals prioritised values such as saving pedestrians versus passengers. Similarly, in Study 3, students were asked to select out of two offerings the ethical term that they value more. This approach provided insights into the ethical trade-offs students were willing to make and the values they considered most important.

Quantitative data were analysed using descriptive and comparative statistical methods. Open-ended responses were analysed qualitatively using inductive Thematic Analysis to gain deeper insights into students' perspectives.

⁴ In HCI research, studies that systematically expose participants to different conditions using mock-ups or technology probes thereby allowing controlled comparison of responses, are commonly described as "experiments" (Shneiderman, 2010). In education research traditions, however, such designs are often classified as a non-experimental elicitation or explanatory studies (Robson & McCartan, 2002; Johnson & Christensen, 2024). This thesis follows the HCI usage while acknowledging this disciplinary distinction.

3.7 SUMMARY OF RESEARCH APPROACH

This chapter outlined the research approach of this thesis, following Crotty (1998)'s four elements that scaffold the research design process. The epistemological stance underpinning this thesis is primarily constructionism, which views knowledge and meaning as constructed through human interactions with objects in the world and shaped by social and cultural context. This epistemological foundation was closely aligned with the theoretical perspective of symbolic interactionism, which emphasises the role of social interactions in the construction of meaning, reality and social behaviour. Symbolic interactionism further informs how individuals create and interpret meanings through their interactions and how these meanings influence their actions and relationships. Together, these frameworks guided the choice of methodologies and methods employed in this research.

Study 1 employed a speculative method - story completion - to explore students' responses and concerns regarding AIEd adoption. The data were analysed qualitatively using thematic analysis with partial application of grounded theory techniques to uncover patterns and themes. After understanding students' attitudes, Study 2 examined whether other powerful stakeholders involved in AIEd adoption, such as school decision-makers and developers, had addressed or considered student concerns. This study used semi-structured interviews for data collection and reflexive thematic analysis to analyse the data, focusing on understanding the dynamics of decision-making and stakeholder interactions. Then, Study 3 adopted a pragmatic mixed-method approach to investigate students' preferences for different AIEd features and their ethical priorities. Quantitative data were analysed using statistical methods, while qualitative data provided deeper insights through inductive thematic analysis.

Chapter 4: Students' Perspective: Ethical and Sociotechnical Consequences of AIED Adoption

This chapter presents Study 1 of this thesis - an exploratory investigation into the broad spectrum of students' responses regarding AIED adoption. Study 1 answers the RQ1: *“What ethical challenges do students perceive in the adoption of AIED?”* While ethical discussions on AIED largely focus on technical and governance issues, little is known about how students experience and respond to these systems in practice (as identified in Chapter 2). To bridge this gap, Study 1 adopted a student-centred speculative research method - story completion - and invited 71 higher education students to generate narratives about AIED adoption. Through qualitative analysis, the study reveals potential impacts in three key dimensions: individual, pedagogical, and classroom collaboration dynamics. The findings were disseminated through multiple publications, three of which are included in this chapter, each corresponding to one of the three dimensions.

The findings from Study 1 establish a thematic foundation for the subsequent studies by offering an in-depth empirical perspective on the ethical and sociotechnical implications of AIED adoption. Study 2 builds on these insights to examine whether key decision-makers acknowledge and address students' concerns. Study 1 findings also informed my approach to interviewing stakeholders in Study 2, allowing me to critically reflect on their responses. Study 1 findings also shaped the experimental design in Study 3, with scale elements derived from the themes identified in Study 1. As the thesis progressed, the significance of Study 1 became increasingly clear – its findings revealed concerns that are deeply important to students (validated in Chapter 6) but often dismissed or compromised by more powerful stakeholders (detailed in Chapter 5).

This chapter begins with a brief recap of the study's approach (Section 4.1), and its contribution to the thesis (Section 4.2), followed by an outline of publication information where the publications arising from Study 1 (Section 4.3), with detailed presentations of three published papers in Sections 4.4, 4.5, and 4.6. Finally, Section 4.7 introduces additional reflections and discussions of data analysis and findings that have not been covered in the publications.

4.1 APPROACH FOR STUDY 1

As detailed in Chapter 3, Section 3.4, Study 1 employed the story completion method for its speculative approach, enabling the proactive identification of potential concerns from students. Story stem scenarios were carefully crafted based on existing literature that reflects current AIED advancements. A total of 71 higher education students elaborated on these story stems, providing rich narratives for qualitative analysis. Thematic analysis, supported by techniques from grounded theory, was employed to identify relationships among themes and codes.

4.2 FINDINGS AND CONTRIBUTION TO THE THESIS

Study 1 contributes to the overarching research question of this thesis – “What are the ethical challenges in adopting artificial intelligence in education (AIED) from the perspectives of different stakeholder groups?” – by first examining the perspectives of students, the most vulnerable stakeholders directly affected by AIED. By focusing on students’ responses, this study uncovers a nuanced understanding of its potential impacts on individuals, the pedagogical environment, and the broader educational ecosystem. The key findings are summarised as follows:

Individual Impacts. Students reported potential negative emotional and psychological influences from AIED monitoring, such as heightened anxiety and stress. These were accompanied by behavioural changes, such as acting, pretending, or pandering to the AI monitoring mechanism. Such emotional responses not only affect students’ well-being but also influence how they engage with learning environments, which could further reshape the pedagogical landscape.

Pedagogical Impacts. Students highlighted shifts in learner autonomy, learning environments and pedagogical approaches, teacher-student interactions, pedagogical relationships and roles. These concerns raise critical questions about how AIED mediates and transforms traditional teaching and learning dynamics.

Classroom Dynamics Impacts. Students identified challenges with AI decontextualising educational contexts, where the complexity of pedagogical dynamics is oversimplified (e.g., the system reduces human affective status to quantifiable indicators). They also raised issues about AI-teacher collaboration, noting susceptibility to bias concerns and power disparities between teachers and

students. Additionally, unintended consequences of adopting AI monitoring and intervention can create a *Situated Changing Circle*: AI monitoring prompts inauthentic changes in student behaviour; these altered behaviours feed back into the system as collected student data, subsequently disrupting the AI system's efficacy.

Study 1 contributes a three-dimensional framework (individual, pedagogical, and classroom dynamics) to advance the understanding of unintended consequences of AIED adoption. These three dimensions together provide a comprehensive thematic picture and serve as the basis for a potential conceptual model that anticipates and critically evaluates the risks of educational AI. The identified themes, which have been presented separately across individual publications, are consolidated in Table 2 below:

Table 2. Dimensions and themes of potential consequences identified in Study 1.

Dimension	Themes	Where to review full code sets & illustrations
Individual Impacts	Affective Responses Behavioural Change Cognitive Beliefs	Fig. 1 in Section 4.4 (Han et al, 2021) OzCHI
Pedagogical Impacts	Learner Status (Autonomy) Learning Environment & Experience Educational Process & Approach Pedagogical Interaction Social and Educational Relationship Role Changing	Table 1 and Fig. 1 in Section 4.5 (Han et al, 2025) IJAIED
Classroom Dynamics	AI models Decontextualise Educational Context Human Bias and Power Disparity in AI-teacher Cooperation AI's Impact on Students' Behavioural Change Further Challenges AI Effectiveness	Fig. 1 in Section 4.6 (Han et al, 2025) CSCW

Study 1 directly addresses *Problem 3: Neglecting Students' Perspectives in AIED Design*, by foregrounding students' ethical concerns, expectations, and lived experiences. These findings also contribute to *Problem 1: Lack of Understanding of Unintended Consequences of AIED*, offering concrete empirical evidence on its potential ethical and sociotechnical implications. The findings across the three dimensions reveal that AIED's potential impacts are not isolated but interact in nuanced ways, highlighting the contextual complexity of real-world educational environments.

Building on this, Study 1 also expands the ethical discourse of AIEd, helping to address *Problem 4: Limited Understanding of Ethics in AIEd*. The findings challenge existing ethical frameworks in AIEd, which primarily address technology-centric concerns. It offers a human-centred lens demonstrating that ethical risks are deeply embedded in the structure of educational interactions. Instead of treating AIEd as a standalone tool, Study 1 highlights its reciprocal relationship with the broader educational ecosystem, emphasising that ethical considerations must be integrated into the social and pedagogical context.

Ultimately, by centring students' perspectives, Study 1 provides foundational knowledge that advances the thesis's overarching aim of promoting ethical, context-sensitive AI design and adoption in educational settings.

4.3 PUBLICATION INFORMATION

Study 1 has resulted in five publications⁵, reflecting its findings across various dimensions:

Individual Dimension. The findings were published at the 34th Australian Conference on Human-Computer Interaction (OzCHI 2021) and presented in Canberra, Australia, in November 2021. This paper and presentation won the best paper award.

Pedagogical Dimension. The findings were first published as a short paper at the 23rd International Conference on Artificial Intelligence in Education (AIEd Conference 2022), and presented in Tokyo, Japan, in June 2022. An extended full version of the paper was later published in the International Journal of Artificial Intelligence in Education (IJAIED) in 2025.

Classroom Dynamic Dimension. The findings were accepted and published at the 28th ACM SIGCHI Conference on Computer-Supported Cooperative Work & Social Computing (CSCW 2025), presented in Bergen, Norway, in October 2025.

⁵ While the findings across separate dimensions are presented in different papers, these dimensions are inherently interconnected. I emphasised this interplay, particularly the relationship between individual psychological and behavioural changes and their impact on the pedagogical environment. However, to maintain focus and cater to different audiences, each paper prioritised specific aspects of the findings.

An additional reflection paper expanding on Study 1's findings has been accepted as a book chapter in *Artificial Intelligence and Academic Integrity: Navigating Ethical Challenges of AI in Education*. This chapter explores the value of academic integrity in the era of monitoring AI in classrooms. While the book chapter is not included in the main body of this thesis, it is provided in Appendix A.

All five papers underwent anonymous peer review by at least three reviewers before acceptance. Below is a summary of the publication details and acceptance rates:

- OzCHI 2021: Acceptance rate ~50% (362 out of 729 submissions), with the Best Paper award (1 out of 729 submissions).
- AIED 2022: CORE A-ranked conference with an acceptance rate of ~21%.
- IJAIED Journal: A Q1 journal recognised as a top-tier venue for research on AI applications in education.
- CSCW 2025: CORE A-ranked conference with an acceptance rate of ~21%.

In accordance with the University of Melbourne's 2025 guidelines for thesis with publication, the author-accepted versions of these papers are included in the following sections. Each section corresponds to a specific dimension of the findings.

4.4 IMPACTS ON INDIVIDUALS: EMOTIONAL DISCOMFORT, BEHAVIOURAL CHANGE, AND COGNITIVE DISRUPTION

Full citation: Bingyi Han, George Buchanan, and Dana McKay. 2022. Learning in the Panopticon: Examining the Potential Impacts of AI Monitoring on Students. In Proceedings of the 34th Australian Conference on Human-Computer Interaction (OzCHI '22), November 29–December 02, 2022, Canberra, ACT, Australia. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3572921.3572937>

The published paper begins on the next page.

Learning in the Panopticon

Examining the Potential Impacts of AI Monitoring on Students

Bingyi Han *

School of Computing and Information
Systems, University of Melbourne,
Melbourne Vic Australia,
bingyih@student.unimelb.edu.au

George Buchanan

iSchool University of Melbourne
Melbourne Vic Australia,
george.buchanan@unimelb.edu.au

Dana Mckay

School of Computing Technologies,
RMIT University, Melbourne Vic
Australia,
dana.mckay@rmit.edu.au

ABSTRACT

In a panopticon, people are intrusively monitored across all areas of their lives. AI monitoring has been ever more widely adopted in education, with increasingly intrusive monitoring of students. These changes potentially create ethical harms, but current ethical discussions predominantly focus on legal and governance issues. The concerns of the majority of users—namely students—are neglected. Overlooking students' concerns further increases their vulnerability. We use a student-centred and speculative approach through the Story Completion Method (SCM) to explore how students would potentially respond to intrusive AI monitoring in a higher education setting. Our study included 71 participants who elaborated on the story stems we provided to them. Through a blending of thematic analysis coding and the techniques of developing grounded theory, we reveal that the common responses of students to extensive AI monitoring included impacts on personal psychology, changed behaviour, and cognition. There are likely major disruptions to personal autonomy, identity and educational relationships. If we are to avoid a future 'big brother' classroom, further investigations using HCI methods are critical to understanding how to protect students in AI-dominated learning.

CCS CONCEPTS

• Human-centered computing; • User studies; • Human computer interaction (HCI); Applied computing; • Education; • Computer assisted instruction;

KEYWORDS

AI Monitoring, Learning Analytics, Surveillance, Privacy, Story Completion Method, AI Ethics

ACM Reference Format:

Bingyi Han *, George Buchanan, and Dana Mckay. 2022. Learning in the Panopticon: Examining the Potential Impacts of AI Monitoring on Students. In *Proceedings of the 34th Australian Conference on Human-Computer Interaction (OzCHI '22)*, November 29–December 02, 2022, Canberra, ACT, Australia. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3572921.3572937>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

OzCHI '22, November 29–December 02, 2022, Canberra, ACT, Australia

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0024-8/22/11...\$15.00

<https://doi.org/10.1145/3572921.3572937>

1 INTRODUCTION

The educational landscape has been shaped by the increasing adoption of AI [1]. In the conventional education system, educators evaluate students' performance by observing and monitoring their in-class behaviours and activities [2], then reflect on learning suggestions and adjust teaching strategies [3]. The introduction of AI in Education (AIEd), especially the use of automated Learning Analytics, results in constant observation of students (by e.g. close real-time monitoring [4]) and the ability to collect learning artefacts and evaluate them at scale [5]. The extent of this monitoring would not be possible with human effort alone. The AIEd approach potentially facilitates more frequent formative assessment and predictions [6] or more insight into students' learning, for example, successful learning patterns [7]. Those benefits rely heavily on the extensive data that is pervasively collected and processed by AI during students' educational practices, which raises ethical concerns over the heightened surveillance and privacy issues [8, 9]. However, the ethical consequences from the lens of students are little understood [10].

A significant gap remains in the field: existing ethical discussions are mostly theoretical [11] and focus on governance [12]. While governance has been considered, the students affected by AIEd have been neither consulted nor considered: existing work is not student-centric [10]. There is also an absence of qualitative studies to understand the social-technical factors of AIEd's design [13] and empirical understanding of the ethical implications of AIEd's use [14]. An ethical discussion informed and led by the HCI community is missing. In pursuing the positive potential of emerging educational technology, the unintended (potentially negative) consequences are often overlooked [15]. HCI researchers are ideally placed to understand the potential negative effects of AIEd, particularly in vulnerable populations. HCI research can also advocate for a more fair and ethical approach. The HCI community can significantly contribute by providing a foundational understanding of the ethical problems by focusing on students' needs and concerns.

Within AIEd, AI monitoring technologies have been increasingly adopted in school sectors, from campus security to in-class behaviour management [8]. Techniques are becoming more detailed and intrusive, from facial recognition to check attendance [16] to affective recognition of emotions such as boredom, confusion, frustration, or delight, to in turn detecting students' (dis)engagement during learning [17]. AIEd has long been investigating facial expressions and affective analysis to detect signs of learning engagement, especially in the higher education [8]. Currently, there are no mature applications yet in the market but it remains a goal for both academia and industry [17, 18]. What is unclear is the degree to

which the possible ethical risks of this approach are understood. Typical 'perpetual beta' approaches, where the negative consequences of technology are considered only post-implementation are unethical and ineffective in sensitive settings such as education. Hence, rather than ask about current existing experience, we will take a further step by asking about the "might-become" impacts in the context of "what if" (when future monitoring AI comes to students) by using the Story Completion Method in a new study.

The Story Completion Method (SCM) leverages speculative design [19, 20] and research fiction [21, 22]. One of the significant threads of speculative design is to explore the ethical and moral dilemmas of emerging technology [19]. As a speculative method, the SCM has been specifically used as a critical tool to question the status and development of the existing or future technology [23]. In Story Completion Tasks (SCT), participants are invited to create narratives, which present their socially constructed ideas [24, 25]. Those fictions help speak out about the potential implications and social consequences of adopting a technology [26].

We collected 71 stories created by our participants with two story stem openings that describe different scenarios of using AI monitoring in the classroom, to see how students would likely react and respond to the AI during learning. Specifically, our focus is the role of AI monitoring in tertiary education. We then analysed those stories qualitatively. We revealed several unintended consequences of adopting AI monitoring in education, such as negative emotional/psychological impacts, behavioural changes in response to the presence of such systems, and influenced cognitions. Understanding those implications for individual students and their attitudes is key for the further design of safer educational environments. We summarise the contributions of this paper as: 1) A critical investigation of monitoring AI's impact on students from a user-centred HCI perspective; 2) Revealing the potential interactions among the impacts of AI monitoring, such as how emotional responses might lead to students' behavioural changes and disrupted cognition; 3) An initial study of the consequential ethics of the deployment of AI in Education.

2 RELATED WORK

In this section, we will review some existing monitoring educational technologies both developed for the research use and the market, followed by previous discussions around the impact of those technologies, to better understand the gap of lacking students' perspectives on AI monitoring's adoption and its implications.

2.1 Monitoring Technologies used in Education

Several monitoring educational technologies have been recently developed both in the market and for research. For instance, Holstein [27] developed the Spatial Classroom Log Explorer (SPACLE), a prototype tool that produces a moment-by-moment analysis of students' and teachers' behaviours within a particular learning context. Their study with SPACLE suggests that students' awareness of being monitored by teachers will promote learning; fewer "off-task" or "gaming" behaviours were shown. After SPACLE, Lumilo was developed by the same research group [28]. Lumilo is a mixed-reality tool provided for teachers, which monitors and evaluates

students' in-class performance by displaying the real-time indicators of students' metacognitive and behavioural learning process. It then offers customised suggestions of intervention for teachers. It demonstrated that this real-time analytics for teachers can support students' learning by integrating human and machine intelligence [28, 29]. Apart from these research-focused tools, there are several other AIED applications in the market that some schools have implemented. One such popular tool is the data-driven behavioural management system used in classroom settings called ClassDojo [30], another is the virtual school software called Class by Intel [31]. Both primary affordances include tracking and categorising students' in-class behaviours and reporting to teachers, or providing immediate feedback to students [30]. These systems aim to support personalization, reflection, recommendation and intervention [32] in learning and teaching.

Some AIED applications, specifically detect and quantify students' facial expressions and possible emotional status. These applications aim to increase teachers' awareness of students' emotional states to help teachers reflect on their teaching and promote positive learning outcomes for students. Emotion detection applications include the Intel Class [31], EMODA [33] and EMODASH [18]. Except for those emotional detection techniques which are mainly provided for teachers, it has also been raised that emotion analysis could be offered to students as a self-regulated learning tool. The findings of detection (such as students' behaviour, interaction, and affective status) can encourage self-reflection [34–36] and motivate positive changes [37, 38] during their learning process.

Those monitoring tools are often investigated for intended positive outcomes such as enhancing students' learning [28] and adjusting educators' teaching strategies [39] by providing personalization and intervention. However, unintended impacts caused by the AI monitoring do potentially exist. Negative concerns have been lightly discussed in previous research that report AIED developments. For instance, privacy issues have been mentioned with the monitoring tools which detect cheating behaviours in online tests [40]. However, privacy might not be the only ethical impact. Another recent CSCW study investigated the positive support for learning and teaching by tracking students' real-time understanding and sharing for both students and instructors [41]. In their future work, they note how this will impact interactions or discomfort that need further investigation. The following subsection will present existing ethical discussions about monitoring AI.

2.2 Impacts of Monitoring Technology on Education

Some studies aim to explore the impacts of AI monitoring tools across multidisciplinary fields involving not only HCI but also education and pedagogy. For instance, pedagogical studies raised concerns about the datafication of education [42, 43], which leads to changed relationships and role reconstruction within the educational environment [44]. Another example is concerns about quantification of students' behaviours and in-class performance, which encourages obedience in the classroom practices [9] and decreases learning autonomy [45]. Arguments have also been raised around other widespread ethical risks from the technological perspective such as data management and invaded privacy [15, 46],

algorithmic fairness and bias issues [11, 47, 48], and labelling and stereotyping learners [7, 49]. To this end, several frameworks aim at mitigating the ethical harm to students, such as Slade and Prinsloo's framework [50], Sclater's Code of Practice [51], and Southgate's The Education, Ethics and AI (EEAI) framework [52]. However, most of those discussions are theoretical [11] or from a governance perspective [12] with a top-down approach. They still lack actual users' perspectives and thus understanding of the socio-technical implications [14] of monitoring AIEs.

Several recent empirical studies from HCI such as Lu's [47], involve teachers in investigating the unintended effects resulting from the use of a classroom surveillance system, ClassDojo, which monitors and quantifies students' in-class performance and reports to teachers. Their interview study with teachers reveals the negative psychological impacts on children which come from the monitoring technologies' measuring, codifying, and simplifying the psychosocial factors existing in the educational environment. The same research group of Lu's [53] has also revealed that the monitoring AI has challenged teachers as they also feel surveilled by parents, students, and school administrators. Consequently, teachers had to be careful of using students' data, and negotiate care and control of those data. They have to balance the optimistic hopes that students can be better cared for (e.g., emotional issues), with ethical dilemmas such as intrusion. This series of unintended consequences might disrupt the pedagogical relationships in the conventional education landscape. However, this work is explicitly from the teachers' perspective, even though those studies aimed at revealing the impact on students. Understanding based on students' actual responses is still scant.

Some empirical research on AIEd involves students but the objectives are students' functional preferences for the AIEd tool design and implementation, rather than ethical impacts [54, 55].

There are some studies investigating ethical considerations from the students' perspective, but these are often focused on a specific known issue. One such issue is privacy, which is the main or even only targeted problem in some studies [15, 56]. Similarly, Tasi et al. [57] explore students' expectations of privacy, however, they mainly focus on the technical perspective of ethical issues such as transparency and information asymmetry. Hence, there remains a lack of students' perspectives in exploring the impact of monitoring AI.

We have summarised the literature on AIEd adoption. The spread of monitoring technologies is growing, and these technologies are increasingly intrusive. Some governance frameworks focusing on issues such as fairness have attempted to address the use of AIEd, but these have been highly theoretical. Some behavioural changes, disruptions to relationships, and negative consequences as a result of AIEd have already been detected in studies of teachers. Ultimately, though most of the impacts of the AIEd adoption, both positive and negative, are directly on students, and studies of students' experiences of monitoring technology have been highly focused on privacy, to the exclusion of other concerns. A lack of "evidence of impact" [58] especially from the perspective of learners increases the difficulty of understanding and preventing ethical harm, which is also seen as a critical barrier to buy-in for AIEd adoption [59]. To address this, discussions of AIEd must adopt consequential ethics, an understanding of the ethical issues that arise as the result of

the adoption of these technologies. However, technical artifacts of AI are often perceived as isolated in the ethical discussion while the consideration of the broader contexts in which AI systems are embedded is lacking [60]. We cannot use a framework of consequential ethics until the consequences surface. Hence, we conducted a speculative study to explore the possible consequential ethical risks of adopting monitoring AI that might come to students.

3 RESEARCH METHOD

The purpose of this study is to explore the possible impacts of the prospective AI monitoring applications on students in tertiary education. We focus on facial recognition and emotion detection techniques which are most ambitiously envisioned by both academia and industry [17, 18]. We hope to bring the future prospective AIEd scenarios to the current table of ethical discussions, and thus prevent ethical harm to students before it happens. Hence, we adopted the Story Completion Method (SCM) to collect data by inviting participants to co-create stories with the given Story Completion Tasks (SCT), where those stories are informed by their social and cultural knowledge or experience. This research approach was approved by the Human Research Ethics Committee at the University of Melbourne.

One criticism of this method is that SCM cannot inform us of actual psychological states and should be only considered as stories [24]. So, we will use SCM to investigate commonly held assumptions from the perspective of social constructivism and perceive our results as a "might be" instead of a "will be". By conducting this speculative study under dramatic futuristic scenarios of the classroom with AI monitoring, we will reveal the socially constructed visions of its impacts on students. Recent HCI studies have effectively used SCM for exploring emerging issues in sensitive topics [61] of technology design such as VR pornography [25] and sex robots [62]. In the school context, student groups comprise a vast population with a complex variety, such as sexual identity, race, and a range of cognitive abilities. We thus perceive SCM as an inclusive approach to address potential sensitivities, as the ethical concerns among cultural or identity groups vary.

The SCTs we crafted (see Table 1) constrained the participants to the context of our research interests [63]. Specifically, we focused on the invasive adoption of AI monitoring techniques in university classrooms. We left other details deliberately ambiguous such as the explanation of mentioned AIEd artifacts, functions, or concepts. This deliberate vagueness is a common tactic in SCM to motivate participants to cast their own interpretation and imagination [64]. Unlike other former HCI studies exploring future technology design (e.g. [25, 62]) which used story openings with only one or two sentences, our SCTs offer more immersive scenarios. We provided more descriptions to help participants understand the story context. A similar example of providing immersion through SCT can be seen in a recent HCI research [61] which extends the SCM with interactive storytelling gamification. At the end of the SCT, we prompted our participants by asking them what would happen next and how the character would think, feel, and behave.

We recruited 71 participants who completed both story completion tasks, along with a brief demographic survey that recorded

Table 1: Story Completion Tasks

Scenario 1: During this summer break, Charlie’s university has implemented an intelligent AI System with an education system which will assist teaching. It is the first lecture of this semester. Charlie walks into the classroom. He heard his classmates whispering: There will be an In-Class Surveillance by using the Classroom Behaviour Analysis System: - When students walk into the classroom, the system records their attendance by Facial Recognition Technology. It will then be generated as engagement data stored in Classroom Attendance System. - During the class, the system will capture students’ affective (emotional) states, such as when a student is either confused or confident, using an Emotion Recognition Technology. Then the analysis report will be created for teachers to understand students’ emotional readiness and cognitive needs for their learning. - Each student’s learning material and feedback will be generated via real-time analysis in class, to build a profile of the student’s progress in each class. Across a term, this data will generate a personal “Education Profile”. Now the class bell rings, Charlie feels . . .	Scenario 2: The lecturer said: “let’s have a group discussion”. Charlie then heard another whisper from his classmates: During the in-class groupwork practice, the system will also identify what each student is doing and saying, by using Facial Recognition and Voice Recognition. Eye Tracking Technology will be used to capture the amount of time that each student is focused on each learning resource. Then the analysis report will be created by the Eye Tracking for the teachers to understand the learning progress of the class and the efficacy of the chosen learning recourses. Now the classroom becomes noisy and full of conversations, Charlie . . .
--	--

gender, age, cultural background, and whether they had any teaching experience. Participants included current university students, and recent graduates. We recruited our participants from two cultural backgrounds: Chinese and Australian. We did this to understand whether there are cultural differences when perceiving ethical concepts or harm, because these two cultures demonstrably have varied cultural perceptions of, e.g., privacy [65]. We further split the Chinese group into students who only have only experience of studying domestically (within China), and students who have more than one year’s experience studying abroad. We did this to understand whether lived experiences in different cultures can influence people’s perceptions of ethical concepts. Finally, participants with teaching experiences were singled out in the lab, as we are curious whether having teaching experiences will differ their insights into learning with AI. In total, there are 25 participants from the Chinese cohort, 25 from the cohort of Chinese with international living experience, and 15 from the Australian cohort. 23 participants had teaching experience or are currently working as teachers. There are 6 incomplete stories that are not conducted in our analysis. We translated the Chinese stories to English before data analysis. Each participant is labelled with their cultural background (CHN for Chinese, IECHN for Chinese participants who studied abroad, and AUS for Australian participants), a number, and the suffix T if they have teaching experience, so CHN1-T is Chinese participant 1, who has teaching experience.

We firstly conducted the Thematic Analysis (TA) based on the participants-created stories, where we identified the lower-level codes and higher-level themes. In addition, we realised complex relationships exist between various codes and themes: many of them interplay or interact with each other, within and among different themes. We thus found that TA did not provide explicit support for this analysis. We therefore decided to adopt part of Grounded Theory Methodology (GT). Grounded Theory provides an exploratory method that enables the identification of relationships among concepts [66]. Building upon qualitative data sets [67], Grounded Theory offers a structure of three major coding stages: Open Coding, Axial Coding, and Selective Coding in the analysis

procedure, which helps with identifying the concepts, high-level phenomena, and interpretation. We translated the initial codes from TA to the open codes for GT. By following the GT analysis structure, we further identified the Causal Condition (e.g., a specific function - surveillance), the Phenomenon (e.g., the affected emotions of students), the Objective (e.g., to gain good performance), the Action and Interactive Strategy (e.g., behavioural change), and the Consequences (e.g., the affected learning experiences). Then we recognised the Process Relationship and Cause-Effect Relationship among those identified impacts due to the AI monitoring adoption. The analysis will be further demonstrated in the following sections.

4 FINDINGS: IMPACTS OF AI MONITORING IN EDUCATION

Through thematic analysis of the collected 71 stories, we identified the unintended social-technical and ethical impacts of AI monitoring on various elements in the educational system. Affected parts of the system include learners, their learning status (e.g., learning autonomy), pedagogical roles and relationships, learning environments and experiences. So far, there is not enough evidence to indicate significant cultural differences in participants’ attitudes toward AI monitoring adoption, so this is not further discussed in this paper. This paper will mainly present the impacts on individual students due to their direct interaction with AI monitoring, including the affected personal psychology, changes in behaviour and disrupted cognition. It was clear that those impacts entangled with each other, for instance, the impact on students’ emotions led to behavioural responses. So, by further conducting the analysis following the techniques of constructing Grounded Theory, we finally unfold the consequential relationships between those identified codes, as shown in Fig. 1 (See Section 4.4).

4.1 Influence of AI Monitoring on Personal Feelings and Emotions

As we did our first inductive coding, we saw there is a consistent focus by most participants on the predominantly negative emotional

Table 2: Codes Developed based on Participants' Responses

Core Category/Axial Code	Sub-Category/Open Code (Mentioned Frequency)	
IMPACTS ON INDIVIDUALS		
Behaviour Change	Cheat the surveillance (25)	Elude surveillance - privacy seeking (23)
	Pander to the surveillance mechanism (16)	Rebellion against the system (19)
	Forced performance (15)	Unauthentic behaviours (7)
	Decreased social interactions (7)	Have to be cautious of in-class behaviours (5)
Affected Personal Psychology	Personal Discomfort	Sense of uncertainty (rules, standards) (9)
	- Anxiety due to being monitored (21)	Unconfident (9)
	- Pressure due to being constantly evaluated (15)	
	- Need time to be away and off (21)	
Disturbed Cognition	- Fear of bad performance being recorded (6)	
	Internalised Disturbance:	Educational Disturbance:
	- On individual identity and autonomy (12)	On social interactions (7)
	- Invaded privacy (30)	Concentration on study (27)
IMPACTS ON PEDAGOGICAL ELEMENTS		
Affected Learning Environment & Experience	A scary learning environment (Students generally live with fear) (22)	
	Assuming strict rules and supervisions (18)	
	More efforts are needed for learning (6)	
	More engaged class (13)	
	Educational disruption (27)	
	Losing enjoyment and fun of learning experiences (6)	

effects of constant surveillance and evaluation. We will discuss these speculative emotional responses in detail in this section.

Personal Discomfort

The most frequently mentioned emotional response is the strong personal discomfort. For instance, being monitored and constantly evaluated make students feel “apprehensive” (AUS10-T), “stressed and under pressure” (AUS3-T), “nervous” (CHN22-T), “scared” (IECHN12), “fear” (AUS15-T), “irritable” (IECHN17), “overwhelmed” (CHN17), “anxious” (AUS14-T), “self-consciousness” (AUS10), “worried” (AUS13-T), “panic” (AUS2), “insecure for lack of privacy” (CHN2), “as he knows Big Brother is watching” (AUS3-T), or “cannot relax at all once stepping into the classroom with this surveillance” (IECHN11).

Lack Of Transparency Leads to the Sense of Uncertainty and a Fear of Showing Poor Performance

Under the surveillance, the student “fears his poor performance being recorded” (CHN2) and worries if “the system “interpreted or labelled his behaviour as negative” (IECHN8-T). It might be due to the system’s lack of transparency as we did not explicitly give the participants a detailed description of the monitoring and evaluation mechanism. Consequently, this opaque system provides the context which leads to the student’s sense of uncertainty. This sense of uncertainty includes the uncertain boundary and scope of monitoring, and the uncertain influence of evaluation. These uncertainties further construct the unclear standard of how students should behave in this learning environment with a monitoring AI. This feeling of uncertainty raises stronger personal discomfort such as worries, which further bring suspicion towards both surveillance and evaluation:

“The lecturers have told the students to just ignore it, but Charlie knows that he will not be able to do that. He feels uneasy knowing that his every move may be captured and calculated by the class behaviour analysis system. What exactly is it looking for? Should he be writing notes constantly? Should he be nodding along even when the teacher is not looking at him?” (AUS4-T)

“He feels really confused and taken aback that he had no idea about the system before beginning the class. He starts to feel angry that the university wasn’t transparent about how they were going to run the class. He worries that if he says the “wrong” thing it will go on his record and it will impact his grades. Charlie doesn’t know what to do - should he ask the lecturer more about the system? Should he lodge a complaint? He worries that his expression is being recorded and that his anger and anxiety will show through. He wonders what the consequence of that could be.” (AUS14-T)

Uncertainty and a New Surveillance Panopticon

It is worth noting how other consequences chain from the oft-mentioned sense of uncertainty described above. Once non-transparent surveillance with unquantified rules and judgement mechanisms is implemented, a power distance exists between the individual student and the system. As a result, the sense of uncertainty leads some individuals to subconsciously assume particularly stringent requirements alongside a high standard for behaviour. Participants described assuming that “the monitoring is ubiquitous” (CHN6) and “eyes staring at me all the time” (IECHN12) where

the surveillance catches every slack moment in class and “being distracted (or loaf on the study) is not allowed” (IECHN1). This psychological process was seen in many stories. Consequentially, students considered the classroom as an intense and strict learning environment with “a scary atmosphere” (CHN7) where individuals “generally live with fear, the fear of poor performance being recorded” (IECHN23). For this reason, the students believe more efforts are needed for learning under this system where he/she “must be constantly switched on” (AUS6-T), or “to engage more actively” (AUS12), or “to act as a perfect self” (IECHN2), to ensure his/her performances meet the imagined required evaluation standard. Unsurprisingly, this constant performance leads to personal discomfort. This discomfort is described in a large number of stories, for example, the participant described the environment as “too stressful to be a good student” (IECHN1); or said, “it is tiring as it seems not allowed to be away and off with the surveillance. . . even tigers sometimes doze” (CHN9-T).

The following story clearly shows a student’s assumption of strict surveillance where being distracted is not allowed:

“He feels uncomfortable that every word and deed of his own will be under surveillance. . . Thinking of this, Charlie suddenly realised that the scene of his wandering mind just now should be recorded by the surveillance. He then stopped thinking and continued to listen to the class.” (CHN3)

Two stories show the student has to always perform to avoid poor performance scores:

“I would feel very nervous because I feel like all my actions are being monitored and recorded and my every move affects my grades. So I would try to behave in an unnatural way/ in a way that I think will give me high marks from the perspective of the machine. . . Now with the additional voice recognition and eye tracking, I would feel even more nervous since not paying 100% attention at all times will be tracked, and therefore can impact my grades.” (AUS5)

“Charlie is worried. . . he can’t let his guard down and must be constantly “switched on”. If he is labelled as disinterested or if he has a bad day he may be called up and reprimanded. If he is performing poorly his teacher may assign more onto his workload and stress him out.” (AUS6-T)

Uncertainty Decreases Confidence During Learning

In addition to the consequence of the uncertainty demonstrated above, as this feeling of uncertain leads to the student assuming the strict rules with high standards, and assuming that the system compares all individuals. As a result, the student becomes less confident in his/her in-class performance, as described in many stories. Furthermore, this fear of poor performance being recorded seems to exacerbate confidence problems:

“As the system recorded his performance, he already felt nervous and worried about seeming dumb in front of his classmates. He is worried if the system shows that he is terrible at this subject and if his peers notice this and treat him differently than others. His mind

is racing. He can’t concentrate on learning as he is worried about so many things, such as if this will affect his grades, stop him from graduating, or affect his ability to get a job.” (AUS17-T)

“ . . . He is also a little reluctant to participate in answering questions since he fears that if his responses are not high quality, he may look a bit stupid.” (AUS7-T)

Alongside losing self-confidence, a few respondents expressed their worries that their historical recorded poor performance will make teachers lose confidence and raise biased opinions about them:

“Charlie was worried that the poor evaluation scores would stop his teachers from believing him as a good student and not giving him the attention that he used to have.” (IECHN7)

4.2 Behaviour Change and Performativity

We have just discussed one major theme of emotional and psychological responses. Interestingly, we have noticed in many stories that, most of those emotional impacts have later consequences on how individuals react to AI surveillance and evaluation. In this section, we will discuss the second major theme of potential behaviour changes of the students. We will demonstrate examples of the consequential relationships: each is labelled in *italics* in this section, and we use a narrative approach to show their interplay and similarities.

Most frequently, the student’s Fear of Poor Performance Being Recorded associated with their wish to gain good performance scores leads to the student’s various interactive strategies. Holding this objective, mentioned actions include Cheating/Pandering to the Surveillance Mechanism with Performativity after figuring out how to survive and adapt to it. E.g., the student would be compelled to act: “a smarter student is likely to become an actor and pretend with perfect performances” (IECHN2). More specifically, “everyone deliberately performed as a good student, pretending they understand the new delivered knowledge” (CHN21-T). Other mentioned strategies include “developing an Emotional Self-Check Software. . . to avoid being identified as a negative emotion” (CHN18), or “create the High-Score-Dictionary” which summarised the “high-frequency words” that help with gaining scores (CHN19). One story also mentioned, “to buy high score notes from seniors as a cheat sheet. . . to pass the strict evaluation” (AUS2).

There are also Privacy-Seeking Attempts to escape from monitoring, such as by speaking another language (IECHN13) or “undetectable accents and dialects” (IECHN8) to elude voice recognition, or by hiding real emotional or facial expressions for getting away from the emotion/facial recognition. Participants also described seeking private spaces for study or practice so the student can “be free to make mistakes” (AUS5) as “those mistakes won’t be tracked and recorded” (IECHN17-T). A more extreme story mentioned that the student would “drop out of this school”, as believing that “learning within this environment is harmful to mental health.” (CHN2)

The practice of Avoiding Mistakes commonly led to students Being Cautious of In-Class Expressions and Behaviours. Feeling this “self-conscious” (AUS17-T) is often accompanied by discomforts such as “feel more anxious” (AUS17-T). Many stories also expressed

a “forced performance” (CHN10-T), especially with more active and positive participation during class:

“Charlie forced himself to participate in the discussion although he had no idea about the topic. He participated only because he’s being watched, and he wants to behave himself.” (IECHN25)

“I have to be active in the group discussions but to make sure my expressions are correct.” (CHN8-T)

Knowing he was being monitored, “he started to pretend to engage more in the class, such as asking questions, taking notes, and wearing a positive emotion on his face. Although this may somehow motivate him to be actively involved in the class, a 2-hour lecture seems quite stressful and leads to some pressure he might not be able to handle.” (IECHN18-T)

In contrast, some stories consider this forced performance as a positive consequence of Enhancing the In-class Engagement of students:

“Even Kevin, the naughty boy who is usually the most difficult to engage in class, is actively participating in the discussion and expressing his views. . . Charlie felt that the classroom behaviour analysis system seemed to play a positive role in promoting learning in the classroom.” (CHN21-T)

“He should behave better than what he used to be... It is certainly a good thing that students can be made more engaged with the class and their learning materials.” (IECHN14-T)

More often, these behaviour changes especially forced performance and being cautious of self-behaviours, are aware as Unauthentic Performance, which is unnaturally different compared to how the student is supposed to behave without the monitoring AI. It is often accompanied by or leads to Personal Discomforts, such as feeling self-conscious and panicking:

“Charlie feels quite self-conscious about what his facial expressions may reveal. He tries to keep a ‘poker’ face all through the lesson. He is also a little reluctant to participate in answering questions since he fears that if his responses are not high quality, he may look a bit stupid. His friend is aware that Charlie is not being his usual self. . .” (AUS7-T)

He felt scared because of the AI monitored him invasively, “he used to be eloquent and outgoing, now he became taciturn, which was quite wrong compared to usual.” (IECHN12)

“It forces me to perform in a certain manner which I may not be comfortable with.” (AUS16)

4.3 Disrupted Concentration, Privacy, and Identity

We have seen the speculative emotional and behavioural responses to monitoring AI adoption in the classroom. Beyond that, there are other disturbances to the internal cognitive world of students, such as the threats to concentration on the study, privacy, social relationships, and self-identity. These concerns were not just within one person, but also compounded with the emotional and behavioural

changes. In this section, we will reveal those cognitive disruptions in detail.

Educational Disruption

Those earlier-mentioned forced pretending or acting behaviours, as well as being self-conscious, are frequently perceived as “an extra effort” (AUS12) responding to the student’s fear of bad performance being recorded. These extra efforts in turn disrupted the student’s learning. Many stories more directly pointed out that the existence of monitoring and constant evaluation itself “distracts the student’s concentration on studying” (IECHN4) and especially when the student is aware of the monitoring leading to personal discomfort:

“Eventually the emotion tracking will just encourage more people to put effort into faking emotions and distract from focusing on work.” (AUS6-T)

He knows that all his behaviours will be recorded and “he is feeling nervous. . . He struggles with the exercise because he is so distracted. . .” (AUS17-T)

Moreover, the sense of uncertainty distracts the student away as they have to spend time wondering about the unclear standard, and be cautious of their in-class behaviours:

He is stressed about the group discussion after knowing he will be watched and assessed constantly by the AI. . . “He is hardly thinking about the question the teacher asked, rather he is concerned about when he should enter the discussion. What if he accidentally cuts someone off? Will the machine interpret that as poor social behaviour? How can it interpret the natural flows and intricacies of conversation? What was the question again? Charlie’s head feels like spinning and he tries to bring his thoughts back to the class. . .” (AUS4-T)

As a result, quite a few respondents believe their learning environment and experience are dramatically influenced by the existence of the surveillance and evaluation system:

“He must behave carefully to meet the requirements of evaluation. Under this intense monitoring, group discussion loses its enjoyment. The learning experience is no longer relaxing and rewarding because every motion is captured for analysis. He thinks about the new system instead of the learning content. . . He feels no longer easy, creative to express his ideas as he used to be.” (IECHN19-T)

Internalised Disturbances

Alongside learning-related practical influences, many stories also expressed a direct disturbance to individuals’ sense of privacy, their social relations, and to minority groups’ social identities as a result of surveillance.

Nearly half the participants expressed concerns about violated privacy, along with varying levels of personal discomfort such as feeling “a bit worrisome” (AUS15-T), oppressed, anxious, and “unprecedentedly uneasy” (CHN9-T). Even more, many colourful similes/metaphors describe their psychological response to their invaded privacy. For example: “I feel myself being naked both physically and mentally” with nowhere to hide the mental state and body (CHN9-T); “feel like a captive animal” in front of the surveillance

for no privacy (CHN25); whilst the whole environment “sounds like a prison” (IECHN17-T) where students are “studying in a cage, not a university” (IECHN25); “everyone in this classroom is like a white mouse in a laboratory, being observed all the time with performed behavioural analysis” (CHN22-T). These privacy concerns predominantly come from the AI detection of in-class behaviours and emotions.

Concerns about other non-learning-related information about individuals being revealed have also been raised. Most of these worries are related to an individual’s social relations, especially the emotional intimacy that happens in schools:

“He is concerned that he spent some time looking at one of the new female students and thought he might want to get to know her better - did the AI system see that emotion?” (AUS13-T)

Consequently, many participants described their decreased willingness to have informal interactions with other students in the classroom and attempts to hide or avoid the social behaviours from the monitoring. This would in turn influence the social and pedagogical relationship among students:

Knowing all students are observed, “I wouldn’t feel comfortable with discussing with my classmates and sometimes joking around.” (IECHN17-T)

“He avoids eye contact with the friend who teased him before.” (AUS7-T)

“He doesn’t interact with his fellow students anymore, in case the system sees this as time-wasting or that he is lacking confidence. He misses that interaction - that’s where he learnt a lot.” (AUS13-T)

Interestingly, two stories reveal disturbance for individuals from a minority social identity group, such as being homosexual, an identity that can be either public or closeted. For closeted students, stress comes from the fact that the system might capture this secret, and the possibility of being outed becomes a burden to the individual student:

“Will the AI analysis conclude on his sexual orientation after recording the content he browsed on TikTok during the class?” (CHN15)

“I don’t like the feeling of being monitored. Maybe it is because of my fear that the teacher or other students could know my secret beyond my study through monitoring records. . . Then Charlie suddenly thought of his best friend John. Charlie has a special feeling to John. . . he is worried about his special feeling toward John will be monitored through his eyes and mood, which will all be captured and recorded. He didn’t want to be treated differently by others because of his little secret...” (IECHN7)

It is unethical that people cannot keep things they wish to hide from their groups such as classmates and schools. As respondent CHN16 expressed, “Everyone needs a sense of safety that comes from holding their secrets.” However, whether the AI monitor can walk away, or perform with silence and allow those things shrouded in secrecy is yet to be discussed.

4.4 Summary and Overview: Impacts of AI Monitoring and Their Interplay

We abstracted the most frequently mentioned consequences from the stories, including: the function of Surveillance and Evaluation directly impacts the Emotions (Section 4.1); many of those emotional responses further lead to Behavioural Change (Section 4.2), and Cognitive Disruption (Section 4.3). The Learning Environment and Experience and Educational Relationships as a pedagogical impact are also disturbed. There are other pedagogical impacts as further consequences, such as the Status of Learners, the Pedagogical Roles, which will not be discussed in this paper.

Participants’ stories have revealed various levels of psychological discomfort due to constant monitoring and evaluation. While on the other hand, the system’s lack of transparency leads to the student’s sense of uncertainty about the rules, supervision, and behaviour standards. As a result, individuals assume that the learning environment is harsh, with more or extra learning efforts needed for learning. The sense of uncertainty also decreases students’ confidence, which further causes behavioural change with frequent performativity. That is, the student panders to the surveillance by pretending or acting, to gain good performance scores. Also, the student fears showing poor/inappropriate performance as the academic scores will be affected. Consequently, they would force themselves to engage during the class, while being cautious of their in-class expressions and behaviours to avoid making mistakes. Generally, inauthentic behaviour or performance has been frequently expressed as a reaction to the monitoring, whilst these behavioural changes also cause personal discomfort. Those behavioural responses under the surveillance disrupt students’ concentration on the study and the initiative for learning. Moreover, the concern of invaded privacy includes a disturbance of individual identity and their social relation which all increases the vulnerability of students.

It is clear to be seen that those impacts do not exist standalone, they rather interact with each other which complicates matters further. In another word, one impact is often the result or cause of other impacts. Further, those interactions happen across the individual and pedagogical dimensions, as shown in Fig .1.

While most emotional and behavioural responses were described by participants as negative, there are still few positive attitudes. For example, several respondents see forced engagement as a way of enhancing learning during class. We have demonstrated both sides of responses from students’ perspectives where our focus in this paper is the classroom monitoring setting’s direct impact on individuals. In the next section, we will draw on broader discussions around the mentioned concerns.

5 DISCUSSION

As pointed out earlier by Prinsloo [68], surveillance has increased students’ vulnerability to privacy invasion and information collection, processing, and dissemination. These concerns have been reflected in our data: constant collection and evaluation of behavioural causes intrusion to privacy and individual identity, and disturbs learning. These concerns all potentially exacerbate the vulnerability of students, which consequently increases the difficulty of protecting students from ethical harm.

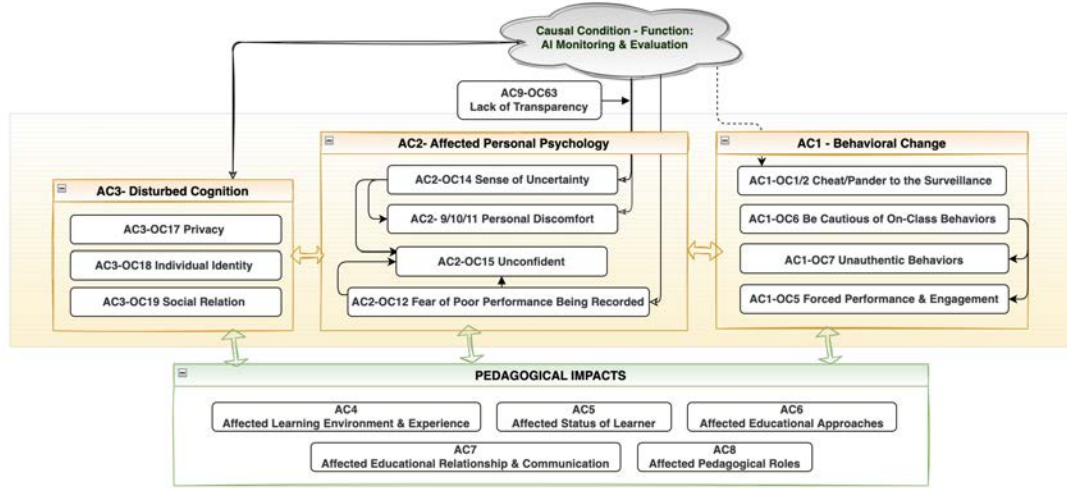


Figure 1: Impacts of AI Monitoring & Constant Evaluation on Individuals

In addition to the increased vulnerability, the speculated experience of students touches all three components of the Breckler's [69] ABC Model of Attitude, i.e., Affect, Behaviour and Cognition, reflecting students' attitude toward the adoption of AI in the classroom. We have seen the possible negative emotions as affective responses; the unauthentic behavioural changes and acts; and the cognitive beliefs of being invaded and disrupted. These concerns demonstrate the seriousness of AI monitoring, since it potentially affects the student's spirit (or emotion), body, and mind. In the following subsections, we will discuss in detail the implications of those three affected elements: emotional responses, behavioural change, and disrupted cognition. We will also discuss the utility of the Story Completion Method for exploring sensitive topics.

This paper is part of a wider study on how AIED adoption might impact students' learning. This paper focuses on AI monitoring before we explore further functions such as customised recommendations and intervention. The impacts mentioned are from the perspective of individual students. There are aggregated pedagogical impacts such as the benefits and issues that come from the evaluation of the whole class, and also the support of learning processes. Those topics will be covered in another paper.

5.1 The Emotional Impacts of AI Monitoring

From our findings, recurring personal discomforts come directly from the constant surveillance and evaluation, such as anxiety, fear, pressure, insecurity, and lack of confidence. Similar concerns have also been raised by other earlier research, which shows that constant surveillance used in education will impact students' physical and psychological well-being, e.g. the chilling effect of students' fear of surreptitious monitoring, or amplified performance-related stress [70]. We should be alert to these negative reactions, as these discomforts might change how people receive the adoption of Learning Analytics and how they live with it. Further, most psychological impacts can directly lead to students' behavioural changes.

Students' fear of been seen to perform badly is an important social factor related to self-esteem and fear of failure. The feelings of discouragement and insecurity found in our data contravene requirements identified in earlier work which includes student requests to learn in a safe environment where they could make mistakes without the feeling of fear [46]. Importantly, this fear of acquiring a record of bad performance results from students' sense of uncertainty, which itself is due to a lack of transparency about the system and its operation. In turn, this further intensifies discomforts and leads to further behavioural change. The lack of transparency and the related ethical concerns do not only exist in our stories; they have also been reported in existing real-world studies. For example, Slade and Prinsloo [71] have long raised concerns that a lack of transparent data collection process of educational institutions on students' information potentially leads to ethical harm. The lack of precise transparency in our description of the putative monitoring system is commonplace in users' understanding of such systems in the real world (e.g., [72]).

In a different context, an early empirical study investigated the long-term effects of ubiquitous surveillance in the home [73]. The results revealed that the surveillance may not be potent enough to bring significant stress and negative mental health effects, but it was shown to be a cause of personal discomforts such as annoyance, concern, and anxiety. The reason for these discomforts was tentatively attributed to the violation of privacy functions of the home, though most participants acclimatised to the system within months. In contrast to the home, students are not typically involved when making decisions about monitoring in classroom environments. Students are thus a vulnerable group within the learning environment. We have seen the speculated negative impacts of surveillance on learners' affective states in our context of higher education. We need careful consideration of the psychological implications, of such monitoring as those impacts are entangled with changes in students' behaviour.

5.2 Further Behavioural Change due to Affected Personal Psychology

There are theories from previous research describing people's behavioural change due to the existence of surveillance systems, such as panopticon theory, chilling effect, and performativity [30]. These have also been seen in our data. In this section, we proved that these effects also happen in different contexts, e.g., more intrusive AI techniques used in education. Furthermore, we revealed how these behavioural changes were raised from the affected personal psychology (emotions).

5.2.1 Surveillance, Panopticon, and Transparency. Foucault [74] has argued that the school functions as a social control mechanism. It governs students where the architecturally designed surveillance is deployed with a disciplinary and hierarchical power relation between students and the educational system [75]. In this way, the governmental mechanism of the education system reflects the panopticon. With the use of AI techniques in education, the information collection and data-based calculations of the surveillance of individuals will be more intrusive and overwhelming, which increases the vulnerability of students [68]. Criticisms from the pedagogical community have argued that the monitoring tools replicate and extend the Foucauldian Panoptic Surveillance on students' in-class behaviour (e.g., [9, 30]).

Our data reflects the panopticon in AIED that students' sense of uncertainty about the surveillance and behaviour standards leads to their assuming the monitoring is ubiquitous. This is due to the untransparent monitoring mechanism and evaluation operations where an asymmetric information relationship exists between students and the surveillance system. Students' fear of the monitoring and learning environment makes them believe they should put more effort into learning to meet the imagined standard, such as assuming they "must be constantly switched on", It brings further negative psychological impacts such as anxiety and exhaustion.

To mitigate the students' personal discomfort that comes from panoptic surveillance, it is necessary to first take care of students' sense of uncertainty by making the monitoring mechanism and procedures entirely transparent. It must be clear to students what information will be collected, the detailed rules of the evaluation and calculation, whether their performance will be ranked, and the consequences of the evaluation. It has been suggested that providing transparency and avoiding ambiguity in the data collection and utilization process is helpful to building trust [76] and improving students' comfort in sharing data and providing consents [77]. In addition to data use issues, our findings suggest that making monitoring operations more transparent might decrease students' discomfort, mitigating the panopticon effect when being monitored in an educational environment.

5.2.2 Constant Evaluation and Performativity: Manolev argued that a learning environment where the students' behavioural performance is constantly evaluated creates a culture of performativity as a behavioural control mechanism [30]. Students will constantly be subject to, and participate in calculating and organising their in-class performance, in response to evaluations and benchmarks [78]. This has been seen in our data where students tend to pander to the evaluation mechanism by performing themselves. Consequently,

the learning experience under constant evaluation reshapes students, especially their learning initiatives and autonomy.

This culture of performativity leads to various behavioural changes in class. We will not dive deeply into these changes in this paper. However, when students deliberately present themselves in an artificial way, such as forcing themselves to participate and act perfectly, or cheating and pandering to the surveillance and evaluation, this gives rise to adverse behavioural impacts [51] and creates personal discomforts such as the sense of unnaturalness ("not being the usual self"), exhaustion, and anxiety. An ethical argument has been raised that students should be free of the feeling of embarrassment caused by less active participation [79]. However, as seen in our findings, the undesirable behavioural changes and emotional discomforts interact with or even exacerbate each other.

5.2.3 Chilling Effect. Previous studies argued that monitoring systems have chilling effects on students [43, 80]. That is, surveillance promotes individuals' self-censorship, conformity, and inhibition, which renders them more cautious and careful of their engagement in candid discussion [81–83]. This is consistent with our data including complaints that the student must be cautious of in-class expressions and behaviours, with feelings of discouragement, lack of confidence and increased panic. These behavioural changes and psychological impacts also affect learners' status, such as self-determination and learning autonomy.

Our findings might address the gap in Holstein's experiment [28], which suggested that real-time teacher analytics enhances students' learning. Holstein's study [28] has shown that, when students are monitored by real-time analytics, they are more willing to receive hints, spend less time on unproductive persistence, make fewer mistakes on specific questions, and have more idle time in their learning process. The context of Holstein's study is with school children while our context is in tertiary education. His results echo our findings that students are cautious of their learning practice and avoid making mistakes, as they fear poor performance being recorded by the system. Our results can be seen as a supplement with psychological factors to his findings that students' awareness of being monitored reduces the frequency of maladaptive learning behaviours. However, on the basis of our data we remain concerned about whether all these influences on students are positive, even though Holstein's group perceived the improved learning outcomes as an advance.

We have seen how psychological impacts intertwined with behavioural changes, an interaction to which we must remain alert. These interactions refocus our attention on the rationale for designing and deploying AIED systems. Participatory design techniques that include learners are necessary to understand negative consequences. We might also need to engage in specific focus on different contexts or lenses toward more vulnerable groups, such as younger students, or students such as LBGTQI students or students of colour for whom extra monitoring will create increased anxiety [43].

5.3 Disturbed Privacy and Individual Identity

The invasion of privacy by AI monitoring is much discussed in the literature (e.g., [84]). When considering privacy in the educational context, it is necessary to consider the existing relationships, e.g.,

between student-teacher, student-school, or even younger student-parents. These relationships bring value tensions among different groups which add complexities when addressing privacy concerns, even in other contexts. For instance, early research has shown clashing preferences between parents and teens regarding what data should be included in audit logs on Home-Entryway Surveillance [85]. Teens coming and going will be monitored by their parents and their privacy needs to be negotiated with the parents. However, even though both parents and children agree that children's privacy should be respected, they perceive privacy concepts differently [86]. Those differences which lead to value tensions need to be considered.

Further, the unclear balance between the monitoring's essential caregiving and privacy invasion for child-parent relationship [87] is also reflected in student-school or student-teacher relationships. In addition, there are inherent power differentials between students and teachers, or between students and schools backed by authorities that exist in the educational context. These power differentials have been indicated as one of the features of intimate threats towards privacy invasion, which makes it complicated to ameliorate [88].

We are aware of the privacy-preserving computing technology developed for facial recognition or eye-tracking, which might mitigate privacy concerns by anonymising both data collection and use [89]. However, these techniques may not solve be able to solve the privacy issues seen in the context of AIED. The ethical use of AIED tools regarding privacy is in tension with the tool's utility and delivered benefits, especially when providing customised recommendations for individual learning. Those intended outcomes might not be realised without a linkage between individuals and their personal data [90]. Hence, this might be a new field for privacy preserving HCI and should consider students' experiences when trying to realise the intended benefits.

Except for the privacy invasion by the monitoring that has been popularly discussed [56, 91], students feel squeezed in the learning environment as surveillance causes educational disruption. It is similar to another context of surveillance at home that, the monitoring devices, e.g. cameras were perceived as life-disrupting [73].

More sensitively, the ethical concerns among cultural or identity groups vary where unique vulnerability needs to be considered (e.g. for students with disabilities [92]). Our data raises these issues with a few sexuality-related responses. We have mentioned the concern from a minor social identity group that the surveillance might be a burden to disturbing their individual identity and revealing their privacy. A recent article has shown that some school surveillance systems flag LGBTQ keywords from students, which may unintentionally out students' sexuality [93]. This surveillance might be more harmful to certain marginalised groups than others, regarding well-being and safety (abuse issues) for individual students and their families.

School is supposed to be a safe place for all students. We should have inclusive consideration to make AIED more thoughtful and friendly to all individuals, regardless of race or sexuality. As pointed out earlier how the surveillance horrified LGBTQ students, individuals from different cultural backgrounds or with different identities might differ in the way they perceive the AIED environment [58, 94]. These differences might also exist in other dimensions of Affective, Behavioural, and Cognitive responses of the ABC model of attitude,

while these need careful consideration to prevent ethical harms from coming to more vulnerable groups.

5.4 Contribution to the Story Completion Method

Like other recent HCI studies (e.g., [25, 95]) using Story Completion Method (SCM), our findings showed that SCM could probe sensitive topics by constructing fictional responses from nonpersonal storytelling with indirectness [63]. Participants can be more comfortable discussing sensitive topics using this third-person lens than asking them to tell "your" perspective. For instance, our respondents spoke out about sexuality-related concerns smoothly during made up narratives, which are informed by and reflect their experience and socio-cultural values [24].

In contrast to that earlier work in HCI, though, we have not exclusively used thematic analysis to code our stories. Instead, we have followed the techniques for developing Grounded Theory. This different analytical method allowed us to tease out the relationships among codes and themes in our complex dataset. By extending the analytical techniques used, we have demonstrated the potential of SCM to further reveal the complexity of the social-technical implications of merging technology's adoption.

5.5 Future Work

Our immediate future plan is to carry on the exploration of AIED's ethical impacts on students including analysing our data with a pedagogical theme. This includes the behavioural change in detail, the affected learner status, learning environment and relationships, learning approaches and pedagogical roles.

The scenarios of this study focus on the context of higher education where the student groups narrowed to university students. Future studies might involve other age groups such as under-18 teenagers and younger children. We also did not consider stigmatised or marginalised groups within the student community. One potential avenue for future research may be considering the consequences or perceived threats to LGBTQIA+ students from deploying AI monitoring.

Extending the work into the future by repeating this work in the context of a real-world deployment, to validate the perceived likely threats as ones that emerge in actual real-world contexts is still key. As we noted above, some of the reported concerns of students have already been discerned in actual deployments. However, we have found new likely consequences that future developers should be aware of. However, this raises the concern that exposing students to stressful and threatening technologies may be unethical. How to forge a path towards understanding real-world consequences that minimise the risk of real-world harm is itself a vital concern to address as work in AIED progresses.

6 CONCLUSION

The traditional learning landscape involving teacher and student is increasingly interrupted by a third party: AI. We now need to build a model of how learning works for students including this third "person", and to understand whether and how to safely integrate AI into education. We need to understand the voices and concerns

of all the stakeholders, including teachers, school, governance, and especially learners and their family.

It is crucial to understand the possible impact of AIED on students, to further ensure the ethics provide a workable ground and a useful baseline for all the components of the system [1]. This understanding might include their acceptance or preferences towards technology, how they would negotiate with themselves to balance the benefit and harm that technology brings, how their tolerance would change when facing different cases. We must also consider whether there are any cross-cultural differences, and what the cultural reasons are for these differences. If we can identify the constraints that bring this community together and create a new sustainable model, we will provide a foundation of AI in education and improve the confidence of educators and students to use any future systems.

REFERENCES

- [1] Luckin, R., *et al.* Intelligence unleashed: An argument for AI in education. 2016.
- [2] Romero, C. and S. Ventura, Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2013. 3(1): p. 12-27.
- [3] Margolis, A.A., Zone of Proximal Development, Scaffolding and Teaching Practice. Cultural-Historical Psychology, 2020. 16(3): p. 15-26.
- [4] Marcu, G. and A.N. Spiller. Collaborative Aspects of Collecting and Reflecting on Behavioral Data. in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020.
- [5] Siemens, G., Learning analytics: The emergence of a discipline. American Behavioral Scientist, 2013. 57(10): p. 1380-1400.
- [6] Bulger, M., Personalized learning: The conversations we're not having. Data and Society, 2016. 22(1): p. 1-29.
- [7] Siemens, G. and P. Long, Penetrating the fog: Analytics in learning and education. EDUCAUSE review, 2011. 46(5): p. 30.
- [8] Andrejevic, M. and N. Selwyn, Facial recognition technology in schools: Critical questions and concerns. Learning, Media and Technology, 2020. 45(2): p. 115-128.
- [9] Williamson, B., Datafication of education: a critical approach to emerging analytics technologies and practices. Rethinking pedagogy for a digital age, 2019: p. 212-226.
- [10] Mangaroska, K. and M. Giannakos, Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. IEEE Transactions on Learning Technologies, 2018. 12(4): p. 516-534.
- [11] Arnold, K.E. and N. Slater. Student perceptions of their privacy in learning analytics applications. in Proceedings of the seventh international learning analytics & knowledge conference. 2017.
- [12] Schiff, D., *et al.* What's next for ai ethics, policy, and governance? a global overview. in Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. 2020.
- [13] Wise, A., Y. Zhao, and S. Hausknecht, Learning analytics for online discussions: Embedded and extracted approaches. Journal of Learning Analytics, 2014. 1(2): p. 48-71.
- [14] Krach, S.K., M.P. McCreery, and H. Rimel, Examining teachers' behavioral management charts: A comparison of Class Dojo and paper-pencil methods. Contemporary School Psychology, 2017. 21(3): p. 267-275.
- [15] Kumar, P.C., *et al.* Privacy and security considerations for digital technology use in elementary schools. in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019.
- [16] Selwyn, N., Digital lessons? Public opinions on the use of digital technologies in Australian schools. Melbourne, Monash University. Retrieved May, 2019. 10(2020): p. 2019-01.
- [17] Dewan, M., M. Murshed, and F. Lin, Engagement detection in online learning: a review. Smart Learning Environments, 2019. 6(1): p. 1-20.
- [18] Ez-Zaouia, M., A. Tabard, and E. Lavoué, EMOASH: A dashboard supporting retrospective awareness of emotions in online learning. International Journal of Human-Computer Studies, 2020. 139: p. 102411.
- [19] Dunne, A. and F. Raby, Speculative everything: design, fiction, and social dreaming. 2013: MIT press.
- [20] Wong, R.Y. and V. Khovanskaya, Speculative design in HCI: from corporate imaginations to critical orientations, in New Directions in Third Wave Human-Computer Interaction: Volume 2-Methodologies. 2018, Springer. p. 175-202.
- [21] Blythe, M., *et al.* Imaginary design workbooks: Constructive criticism and practical provocation. in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018.
- [22] Troiano, G.M., J. Tiab, and Y.-K. Lim. SCI-FI: shape-changing interfaces, future interactions. in Proceedings of the 9th Nordic Conference on Human-Computer Interaction. 2016.
- [23] Lawson, S., *et al.* Problematising upstream technology through speculative design: the case of quantified cats and dogs. in Proceedings of the 33rd annual ACM conference on human factors in computing systems. 2015.
- [24] Kitzinger, C. and D. Powell, Engendering infidelity: Essentialist and social constructionist readings of a story completion task. Feminism & Psychology, 1995. 5(3): p. 345-372.
- [25] Wood, M., G. Wood, and M. Balaam. "They're Just Tixel Pits, Man" Disputing the 'Reality' of Virtual Reality Pornography through the Story Completion Method. in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017.
- [26] Lindley, J., P. Coulton, and M. Sturdee, Implications for Adoption, in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017, Association for Computing Machinery. p. 265-277.
- [27] Holstein, K., B.M. McLaren, and V. Aleven. SPACLE: Investigating learning across virtual and physical spaces using spatial replays. in Proceedings of the Seventh International Learning Analytics & Knowledge Conference. 2017.
- [28] Holstein, K., B.M. McLaren, and V. Aleven. Student Learning Benefits of a Mixed-Reality Teacher Awareness Tool in AI-Enhanced Classrooms. 2018. Cham: Springer International Publishing.
- [29] Holstein, K., B.M. McLaren, and V. Aleven, Informing the design of teacher awareness tools through causal alignment analysis. 2018, International Society of the Learning Sciences, Inc.[ISLS].
- [30] Manolev, J., A. Sullivan, and R. Slee, The datafication of discipline: ClassDojo, surveillance and a performative classroom culture. Learning, Media and Technology, 2019. 44(1): p. 36-51.
- [31] Kaye, K., Intel calls its AI that detects student emotions a teaching tool. Others call it 'morally reprehensible'. 2022: Protocol.
- [32] Khalil, M. and M. Ebner. Learning analytics: principles and constraints. in EdMedia+ Innovate Learning. 2015. Association for the Advancement of Computing in Education (AACE).
- [33] Ez-Zaouia, M. and E. Lavoué. EMODA: A tutor oriented multimodal and contextual emotional dashboard. in Proceedings of the seventh international learning analytics & knowledge conference. 2017.
- [34] Lim, L., *et al.* Exploring students' sensemaking of learning analytics dashboards: Does frame of reference make a difference? in Proceedings of the 9th international conference on learning analytics & knowledge. 2019.
- [35] Verbert, K., *et al.* Learning analytics dashboard applications. American Behavioral Scientist, 2013. 57(10): p. 1500-1509.
- [36] Noroozi, O., *et al.* Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: A systematic literature review. Education and Information Technologies, 2020. 25(6): p. 5499-5547.
- [37] Park, Y. and I.-H. Jo, Development of the learning analytics dashboard to support students' learning performance. Journal of Universal Computer Science, 2015. 21(1): p. 110.
- [38] Jivet, I., *et al.* From students with love: An empirical study on learner goals, self-regulated learning and sense-making of learning analytics in higher education. The Internet and Higher Education, 2020. 47: p. 100758.
- [39] Holstein, K., B.M. McLaren, and V. Aleven, Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms. 2019, Springer International Publishing. p. 157-171.
- [40] Li, X., *et al.* Massive Open Online Proctor: Protecting the Credibility of MOOCs certificates, in Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. 2015, Association for Computing Machinery: Vancouver, BC, Canada. p. 1129-1137.
- [41] Sung, G., T. Feng, and B. Schneider, Learners Learn More and Instructors Track Better with Real-time Gaze Sharing. Proc. ACM Hum.-Comput. Interact., 2021. 5(CSCW1): p. Article 134.
- [42] Selwyn, N., Data entry: Towards the critical study of digital data and education. Learning, Media and Technology, 2015. 40(1): p. 64-82.
- [43] Kwet, M. and P. Prinsloo, The 'smart' classroom: a new frontier in the age of the smart university. Teaching in Higher Education, 2020. 25(4): p. 510-526.
- [44] Jarke, J. and A. Breiter, Editorial: the datafication of education. Learning, Media and Technology, 2019. 44(1): p. 1-6.
- [45] Lundie, D., Authority, autonomy and automation: The irreducibility of pedagogy to information transactions. Studies in philosophy and education, 2016. 35(3): p. 279-291.
- [46] Drachler, H., *et al.* Ethical and privacy issues in the application of learning analytics. in Proceedings of the fifth international conference on learning analytics and knowledge. 2015.
- [47] Lu, A.J., *et al.* Coding Bias in the Use of Behavior Management Technologies: Uncovering Socio-technical Consequences of Data-driven Surveillance in Classrooms, in Designing Interactive Systems Conference 2021. 2021, Association for Computing Machinery: Virtual Event, USA. p. 508-522.
- [48] Durán, J.M. and K.R. Jongsma, Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. Journal of Medical Ethics, 2021. 47(5): p. 329-335.

- [49] Lawson, C., *et al.*, Identification of 'at risk' students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology Research and Development*, 2016. **64**(5): p. 957-968.
- [50] Prinsloo, P. and S. Slade, Student vulnerability, agency, and learning analytics: An exploration. *Journal of Learning Analytics*, 2016. **3**(1): p. 159-182.
- [51] Sclater, N., Developing a Code of Practice for Learning Analytics. *Journal of Learning Analytics*, 2016. **3**(1): p. 16-42.
- [52] Southgate, E., *et al.*, Artificial intelligence and emerging technologies in schools. 2019.
- [53] Lu, A.J., *et al.*, Data Work in Education: Enacting and Negotiating Care and Control in Teachers' Use of Data-Driven Classroom Surveillance Technology. *Proceedings of the ACM on Human-Computer Interaction*, 2021. **5**(CSCW2): p. 1-26.
- [54] Whitelock-Wainwright, A., *et al.*, The Student Expectations of Learning Analytics Questionnaire. *Journal of Computer Assisted Learning*, 2019. **35**(5): p. 633-666.
- [55] Whitelock-Wainwright, A., *et al.*, An exploratory latent class analysis of student expectations towards learning analytics services. *The Internet and Higher Education*, 2021: p. 100818.
- [56] Jones, K.M., *et al.*, "We're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology*, 2020. **71**(9): p. 1044-1059.
- [57] Tsai, Y.-S., A. Whitelock-Wainwright, and D. Gašević, The privacy paradox and its implications for learning analytics, in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. 2020, Association for Computing Machinery: Frankfurt, Germany. p. 230-239.
- [58] Macfadyen, L.P., *Institutional Implementation of Learning Analytics: Current State, Challenges, and Guiding Frameworks*. by Charles Lang, Alyssa Friend Wise, Agathe Merceron, Dragan Gašević, and George Siemens. 2nd ed. Vancouver, Canada: SOLAR, 2022.
- [59] Tsai, Y.-S., *et al.*, Learning analytics in European higher education—Trends and barriers. *Computers & Education*, 2020. **155**: p. 103933.
- [60] Hagendorff, T., The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 2020. **30**(1): p. 99-120.
- [61] Troiano, G.M., *et al.*, "Are We Now Post-COVID?": Exploring Post-COVID Futures Through a Gamified Story Completion Method. in *Designing Interactive Systems Conference 2021*. 2021.
- [62] Troiano, G.M., M. Wood, and C. Harteveld, "And This, Kids, Is How I Met Your Mother": Consumerist, Mundane, and Uncanny Futures with Sex Robots. in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020.
- [63] Clarke, V., *et al.*, Editorial introduction to the special issue: using story completion methods in qualitative research. 2019, Taylor & Francis. p. 1-20.
- [64] Braun, V., *et al.*, Qualitative story completion: Possibilities and potential pitfalls. *Qualitative Research in Psychology*, 2019. **16**(1): p. 136-155.
- [65] Ma, Y., Relational privacy: Where the East and the West could meet. *Proceedings of the Association for Information Science and Technology*, 2019. **56**(1): p. 196-205.
- [66] Walker, D. and F. Myrick, Grounded theory: An exploration of process and procedure. *Qualitative health research*, 2006. **16**(4): p. 547-559.
- [67] Adams, A., P. Lunt, and P. Cairns, A qualitative approach to HCI research. 2008.
- [68] Prinsloo, P. and S. Slade, Student Vulnerability, Agency and Learning Analytics: An Exploration. *Journal of Learning Analytics*, 2016. **3**(1): p. 159-182.
- [69] Breckler, S.J., Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of personality and social psychology*, 1984. **47**(6): p. 1191.
- [70] Reidenberg, J.R. and F. Schaub, Achieving big data privacy in education. *Theory and Research in Education*, 2018. **16**(3): p. 263-279.
- [71] Slade, S. and P. Prinsloo, Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 2013. **57**(10): p. 1510-1529.
- [72] Roberts, L.D., *et al.*, Student attitudes toward learning analytics in higher education: "The fitbit version of the learning world". *Frontiers in psychology*, 2016. **7**: p. 1959.
- [73] Oulasvirta, A., *et al.*, Long-term effects of ubiquitous surveillance in the home, in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. 2012, Association for Computing Machinery: Pittsburgh, Pennsylvania. p. 41-50.
- [74] Foucault, M., *Security, territory, population: lectures at the Collège de France, 1977-78*. 2007: Springer.
- [75] Foucault, M., *Discipline and Punish: the Birth of the Prison*. New York: Vintage, 1979.
- [76] Corrin, L., P.G. de Barba, and A. Brooker, Understanding the learner perspective to inform institutional learning analytics strategy and practice, in *Handbook of Digital Higher Education*. 2022, Edward Elgar Publishing. p. 248-259.
- [77] Li, W., *et al.*, Disparities in students' propensity to consent to learning analytics. *International Journal of Artificial Intelligence in Education*, 2021: p. 1-45.
- [78] Ball, S.J., The teacher's soul and the terrors of performativity. *Journal of education policy*, 2003. **18**(2): p. 215-228.
- [79] Chen, B. and H. Zhu, Towards Value-Sensitive Learning Analytics Design, in *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, Association for Computing Machinery: Tempe, AZ, USA. p. 343-352.
- [80] Kwet, M., Operation Phakisa education: Why a secret? Mass surveillance, inequality, and race in South Africa's emerging national e-Education system. *Mass Surveillance, Inequality, and Race in South Africa's Emerging National E-Education System* (December 4, 2017), 2017.
- [81] Solove, D.J., A taxonomy of privacy. *U. Pa. L. Rev.*, 2005. **154**: p. 477.
- [82] Penney, J., Internet surveillance, regulation, and chilling effects online: A comparative case study. *Regulation, and Chilling Effects Online: A Comparative Case Study* (May 27, 2017), 2017. **6**(2).
- [83] Sklansky, D.A., Too much information: How not to think about privacy and the Fourth Amendment. *Calif. L. Rev.*, 2014. **102**: p. 1069.
- [84] Alkhatib, S., *et al.* Privacy and the Internet of Things (IoT) Monitoring Solutions for Older Adults: A. in *Connecting the System to Enhance the Practitioner and Consumer Experience in Healthcare: Selected Papers from the 26th Australian National Health Informatics Conference (HIC 2018)*. 2018, IOS Press.
- [85] Ur, B., J. Jung, and S. Schechter, Intruders versus intrusiveness: teens' and parents' perspectives on home-entryway surveillance. in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2014.
- [86] Cranor, L.F., *et al.* {Parents'} and {Teens'} Perspectives on Privacy In a {Technology-Filled} World. in *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. 2014.
- [87] Ghosh, A.K., *et al.* A matter of control or safety? Examining parental use of technical monitoring apps on teens' mobile devices. in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018.
- [88] Levy, K. and B. Schneier, Privacy threats in intimate relationships. *Journal of Cybersecurity*, 2020. **6**(1): p. tyaa006.
- [89] Erkin, Z., *et al.*, Privacy-Preserving Face Recognition. 2009, Springer Berlin Heidelberg. p. 235-253.
- [90] Rubel, A. and K.M. Jones, Student privacy in learning analytics: An information ethics perspective. *The information society*, 2016. **32**(2): p. 143-159.
- [91] Stahl, B.C. and D. Wright, Ethics and privacy in AI and big data: Implementing responsible research and innovation. *IEEE Security & Privacy*, 2018. **16**(3): p. 26-33.
- [92] Baek, C. and S.J. Aguilar, Past, present, and future directions of learning analytics research for students with disabilities. *Journal of Research on Technology in Education*, 2022: p. 1-16.
- [93] Carballo, A. Remote Learning Accidentally Introduced a New Danger for LGBTQ Students. 2022 [cited 2022 Jun 14, 2022]; Available from: <https://slate.com/technology/2022/02/remote-learning-danger-lgbtq-students.html>.
- [94] Dawson, S., *et al.* Rethinking learning analytics adoption through complexity leadership theory. in *Proceedings of the 8th international conference on learning analytics and knowledge*. 2018.
- [95] Kelly, R.M., *et al.* "It's About Missing Much More Than the People": How Students use Digital Technologies to Alleviate Homesickness. in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021.

4.5 IMPACTS ON PEDAGOGICAL LANDSCAPE: LEARNING AUTONOMY, EDUCATIONAL PROCESSES, AND ROLES

Full citation: Bingyi Han, Sadia Nawaz, George Buchanan, and Dana McKay. 2025. Students' Perceptions: Exploring the Interplay of Ethical and Pedagogical Impacts for Adopting AI in Higher Education. *International Journal of Artificial Intelligence in Education*. Jan 2025:1-26. <https://doi.org/10.1007/s40593-024-00456-4>

The published paper begins on the next page.



Students' Perceptions: Exploring the Interplay of Ethical and Pedagogical Impacts for Adopting AI in Higher Education

Bingyi Han¹ · Sadia Nawaz² · George Buchanan³ · Dana McKay³

Accepted: 22 December 2024

© International Artificial Intelligence in Education Society 2024

Abstract

Artificial intelligence is increasingly implemented in higher education, offering customized interventions and timely feedback to enhance learning experiences. While these tools have the potential to improve educational outcomes, they also introduce ethical risks and sociotechnical implications such as reduced learner autonomy. Current ethical discussions often focus on computational issues and overlook the nuanced impacts from students' perspectives, which may increase students' vulnerability. Taking a student-centered approach, we apply the Story Completion Method to investigate students' concerns about adopting analytics-based AI tools in education. Seventy-one participants responded to the story prompts, which we analyzed qualitatively to uncover perceptions about how these tools may reshape pedagogical aspects such as learner autonomy, learning environments and approaches, interactions and relationships, and pedagogical roles. Our findings reveal that these potential impacts not only occur in isolation but also interact with one another. This study makes two primary contributions: first, it marks a novel application of the speculative design method to explore students' perceptions of AIED tools. Second, it provides a qualitative analysis of key themes derived from student responses, offering design implications for AIED systems that are sensitive to student concerns and ethical considerations. These insights offer a foundation for future research and contribute to a more student-centered approach to the ethical development of AIED.

Keywords AI in education · Learning analytics · Ethics · AIED · Pedagogy · Learning autonomy

Extended author information available on the last page of the article

Published online: 09 January 2025

Springer

Introduction

The application of artificial intelligence in education (AIED)¹ is rapidly gaining popularity, with analytics-based tools and adaptive technologies to address diverse educational needs. These technologies have been applied to promote cooperative and collaborative learning (Holstein et al., 2019), address educational challenges such as institutional retention and students' experiences (Colvin et al., 2016), assess students' performance (González-Calatayud et al., 2021), identify 'at-risk' students (Queiroga et al., 2022), visualize learning patterns and provide automated feedback (Vehmas et al., 2022). The primary goal of these developments has been to enhance the effectiveness and efficiency of learning processes, including promoting students' affective experience (Baker et al., 2010; Nawaz et al., 2018, 2020) and self-regulation (Gašević et al., 2017) to ultimately drive better learning outcomes (Leitner et al., 2017).

Despite these intended benefits, ethical concerns about AIED's impact on students have emerged alongside these developments and intended positive outcomes (Holmes et al., 2021; Mouta et al., 2023; Swist et al., 2024). For example, improperly designed or misused AI systems can inadvertently reinforce existing biases, potentially disadvantaging minority groups (Baker & Hawn, 2022), or they may compromise learner autonomy (Sclater, 2016) as students may feel increasingly surveilled (Han et al., 2022). While these risks are significant, current ethical frameworks for AIED remain undeveloped, and both the ethical and pedagogical implications of AIED are not yet fully understood (Baker & Hawn, 2022; Holmes et al., 2018a, b; Schiff, 2022).

AI integration in education introduces challenges that extend beyond technical concerns (Siemens et al., 2013), touching on cultural and social dimensions that require nuanced ethical considerations due to the complexities of educational contexts. Emerging technologies in education have the potential to reshape learning environments in dynamic ways. As Tsai et al. (2020) highlighted, the absence of strong pedagogical grounding could exacerbate student vulnerability. Therefore, understanding the potential pedagogical influences of AIED from a student perspective is essential.

This paper explores students' responses to AIED and examines its potential pedagogical implications. Our research addresses two primary questions: RQ1: "*What are the socio-technical consequences that happen to students' learning experience due to AIED adoption?*" (addressed in Sect. 4). RQ2: "*How may these consequences impact one another?*" (summarized in Sect. 5). We adopted a speculative research method called Story Completion Method to collect data. Grounded in qualitative research and informed by learning theories, our study reveals and discusses likely impacts on student engagement and pedagogy, including the Learner Status, Learning Experience and Environment, Learning Process and Approach, Educational Interaction and Relationship, and Pedagogical Roles. We also investigate the complex relationships that may exist among these identified impacts. While impacts of other dimen-

¹ Our study design sees the future implementation of AIED techniques, where AIED would entail AI, LA (Learning Analytics), and EDM (Educational Data mining techniques) tools. In the current context, we use AIED to illustrate our findings.

sions including individual (Han et al., 2022), and technical efficacy [Publication in progress] are presented in other publications, this paper aims to present primarily the pedagogical impacts and their interaction with individual behavior changes. Our findings highlight a wide range of potential impacts, both positive and negative. We advocate for the positive aspects and propose that negative responses can serve as hypothetical worst-case scenarios, guiding future designs to anticipate and address student concerns more effectively, thereby enhancing acceptance.

Based on our findings, we offer several key design implications for future AIED systems to address identified student concerns (more details are in Sect. 5). Firstly, supporting student autonomy and providing options for self-direction can help maintain a sense of control, allowing students to choose how their data is collected and used. Secondly, transparency in data collection and feedback processes is also crucial, with user-friendly interfaces that clarify the impact of AI on learning. Thirdly, AIED systems should offer customizable features for feedback mechanisms to cater to diverse learning goals and respect cultural differences, creating a more inclusive experience. Finally, encouraging active learning within AIED systems can prevent over-reliance on AI, fostering a balanced and engaging educational environment.

Our study makes two primary contributions: (1) it offers the first demonstration of how the speculative design technique, Story Completion Method (SCM), can be utilized in AIED to explore how students may perceive potential tools, and (2) it presents a qualitative analysis of key themes identified from student responses, providing insights with direct implications for the design of AIED systems that are sensitive to student concerns and ethical considerations.

Related Work

Among the most widely promoted benefits of AIED is personalized learning, where content is adapted to meet individual needs through tailored recommendations (Chen et al., 2022). While AI's potential to enhance learning is acknowledged, empirical research on its broader impact on learners remains limited (Ryan et al., 2021). Key areas that warrant exploration include students' perspectives on aspects of AIED that may be contentious - such as being scrutinized during the data collection process, data and algorithmic bias (Baker & Hawn, 2022), standardization or the one-size-fits-all design (Teasley, 2017; Winne, 2017), the tendency to rely too heavily on AI for hints or answers (e.g., with the "gaming the system" strategy (Baker et al., 2008) – these are critical concerns that can influence students engagement and trust.

The focus of most AIED studies has been on quantifying student interactions with learning materials to improve AI systems and data-focused system indicators (Ahmad et al., 2022; Altun et al., 2022). However, this approach might overlook the nuanced qualitative aspects of AI adoption that affect uptake. AIED researchers have noted the lack of qualitative studies (Gašević et al., 2015) addressing these broader socio-technical dimensions, which are often limited to evaluating stakeholder expectations, functional preferences and intended outcomes (Whitelock-Wainwright et al., 2021).

Ethical concerns in AIED research tend to be narrowly framed, emphasizing computational ethical issues such as privacy, fairness, and algorithmic bias (Hakimi et

al., 2021; Tzimas & Demetriadis, 2021). However, AIED's pedagogical impacts—such as on learner autonomy and well-being—are rarely addressed. A limited ethical focus is also evident in institutional governance frameworks (e.g. Prinsloo & Slade, 2016b); Sclater (2016); Southgate et al. (2019); Clark et al. (2020), where only a fraction mention ethical considerations (Khalil et al., 2022). Compliance with ethical and legal standards alone does not guarantee that unintended consequences will be avoided, nor that student trust and acceptance will be maintained (Howell et al., 2018; Klein et al., 2019). The education community advocates for addressing broader pedagogical concerns, but there is a lack of empirical, student-centered evidence to inform these discussions (Tzimas & Demetriadis, 2021).

The challenge, then, comes as most students have little or no direct experience with AIED, making it difficult for them to anticipate its impacts, and yet waiting for consequences to surface post-adoption risks unintended harm. To address this dilemma, we adopted a speculative research method – the Story Completion Method – which we now describe in detail.

Research Method

The objective of this research is to delve into the broad spectrum of students' responses regarding the potential pedagogical impacts of implementing AI in higher education, and to address the existing gap in AIED research that has yet to fully explore students' viewpoints. This research was approved by the Human Research Ethics Committee from the University of Melbourne (Ethics ID Number: 22256).

Story Completion Method and Data Collection

While AIED community shows its ambition to pursue AIED technology advancements, the widespread adoption remains nascent, with relatively few learners having fully experienced its functions and impacts. The prevalent 'perpetual beta' approach, which waits to address the unintended or adverse consequences of technology until after deployment, is particularly unethical and ineffective in sensitive environments such as education. The inherent complexity and potential invasiveness of AIED necessitate a forward-looking strategy to pre-emptively identify, confront and mitigate potential issues before they materialize. To this end, we have opted to extend our inquiry beyond students' current existing experience, exploring the "might-become" impacts of AIED through speculative "what if" scenarios, facilitated by the Story Completion Method (SCM). This stance enables us to proactively consider the ethical implications of future prospective AIED scenarios in the present. It is akin to the typical ethics evaluation approach such as conducting ethical pre-mortems to forestall systemic ethical failures of a technology or a project (Vallor, 2018). Our goal is to prevent ethical harm to students before it occurs and offer insights that can guide the adoption of AIED designs that are not only innovative but more ethically responsible.

Hence, we collected our data through this established approach—the Story Completion Method (SCM)—which leverages speculative design (Dunne & Raby, 2013)

and research fiction (Blythe et al., 2018). SCM is particularly effective in probing the ethical and moral implications of emerging or future technologies (Dunne & Raby, 2013). SCM has often been used for sensitive topics (Troiano et al., 2020). Education, with its diverse population encompassing various identities, races, and backgrounds, is no different.

In SCM, participants were invited to create narratives based on given story openings as Story Completion Tasks (SCT) that elicit their responses to specific contexts and the values that matter to them (Kitzinger & Powell, 1995). The SCM method elicits participants' responses to specific contexts. We developed three futuristic scenarios of university classrooms based on existing or proposed technologies in the AIED research literature. We used iterative pilot studies to ensure that responses were substantive. Scenarios 1 and 2 introduced how AI monitoring tools (such as facial recognition and emotion recognition) might collect students' information, including their engagement patterns and affective states. Scenario 3 described how customized learning services could utilize the collected data, including real-time evaluation and analysis, personalized recommendation, and automated intervention. In line with the established SCM approaches, careful attention was given to the design of our story prompts, ensuring that each word was thoughtfully chosen to elicit meaningful responses (Clarke et al., 2019). When describing technology artifacts and AIED functions, there was a deliberate ambiguity in our scenarios to support students' imagination and enable a wide range of responses for uncovering potential issues. In addition to providing the story stems we also asked the participants what happens next and how the characters would behave².

We targeted participants recently or currently experiencing higher education (undergraduate, postgraduate, PhDs, and recent graduates). To explore potential cultural influences on the perception of ethical concepts in AIED, we recruited participants from two different demographics (China and Australia), via social media platforms. However, we did not ultimately observe substantive differences in the identified themes across the two cultural groups, suggesting that these themes might be broadly applicable across cultural contexts. Our recruitment did not filter participants based on their majors or AI literacy, since we examine students' broad perspectives on AI in education, considering the diverse backgrounds of students in real-world learning situations, regardless of their AIED experience or AI literacy.

Data collection was conducted anonymously through the Qualtrics online survey platform, where 71 participants were invited to complete story-writing tasks. The survey began with a plain language statement and consent form, explaining our research aims and ethical approval. Participants were then presented with three Story Completion Task (SCT) scenarios and asked to complete a story based on each. Each participant's story consisted of three sections, corresponding to the three different AIED scenarios provided. In total, we collected 71 story submissions. However, six participants did not complete all three scenarios, so their stories were excluded from the final analysis. This left us with 65 fully completed stories for analysis. We also collected demographic information from participants including gender, age, and

² The full scenarios are presented at: <https://sites.google.com/view/storysystems>.

teaching experience. Of the participants, 54% were male, 46% were female; none identified as non-binary.

Data Analysis

The focus of this research is on capturing a wide array of student perceptions to inform future ethical considerations and design improvements. Hence, we prioritize qualitative breadth over quantitative precision. Our methodology follows the qualitative epistemologies with interpretivism. We endeavor to interpret and convey the diversity of the perspectival realities from our participants concerning this topic of AIEd adoption (Bauer et al., 2000).

Our analysis commenced with a reflexive thematic analysis to establish initial codes and overarching themes (Braun & Clarke, 2019). The stories were coded at the sentence level, with open codes being organized inductively. These codes were either semantic, which captured the explicit ideas expressed by participants, or latent, which reflected the researcher's interpretations of the underlying meanings present within the data (Terry & Hayfield, 2020). Then, we sorted the codes into initial themes and sketched their relationships on mind maps (Braun & Clarke, 2006).

However, we found complex interactions exist among many identified codes and themes, and Thematic Analysis (TA) could not support interpreting this intricate interplay. Techniques from Grounded Theory (GT) are renowned for their capacity to elucidate relationships, particularly consequential ones, among concepts and codes (Walker & Myrick, 2006). To navigate this complexity, we transitioned from TA to adopting GT techniques, following the approach introduced by Glaser and Strauss (2017).

Rather than adopting the full grounded theory which aims at theorizing or developing new frameworks grounded in data, we adopted the most frequently used form of grounded theory, known as Grounded Theory-Lite. A grounded theory-lite involves using GT coding and relevant techniques to develop a set of categories, aiming at understanding the relationship between various categories (Pidgeon & Henwood, 2013).

GT analysis included three stages: Open Coding, Axial Coding, and Selective Coding. This structured approach facilitated identifying concepts, high-level phenomena, and interpretation. In the open coding process, we translated the initial TA codes into open codes. Then in the axial coding process, we related codes (categories and properties) to each other. We organized codes within a frame that encompasses several elements: *Phenomenon*, the central concept unifying the analysis (e.g., the affected learner status); *Causal conditions*, events leading to the occurrence of the phenomenon (e.g., the specific function - customized recommendation and intervention); *Action strategies*, responses to the phenomenon, goal-oriented activities undertaken by agents (e.g., forced performativity under monitoring from the individual impact); *Consequences*, outcomes of these strategies, both intended and unintended (e.g., the affected learning experiences). During axial coding, we thoroughly examined and compared the categories derived from open coding, integrating relevant concepts and summarizing major categories based on their interrelations. Since our primary goal was to explore the interplays among identified codes rather than building a fully

comprehensive theory, we employed the “lite” version of Grounded Theory (Pidgeon & Henwood, 2013). Specifically, we skipped the final stage of the GT - selective coding, which unifies the entire data set for theory development - retaining only the key elements from open and axial coding.

We present identified pedagogical impacts in Table 1. We illustrate the interactions among pedagogical impacts and the individual impacts in Fig. 1 at Sect. 6. Additionally, we include extracts from participant narratives in Appendix Table 2 to exemplify our coding process. We provide example quotes for each code in Appendix Table 3.

The first author led and conducted the data collection and analysis while discussing with the research team regularly. The analytics procedure was reported transparently to the team to ensure research rigor (Popay et al., 1998; Seale & Silverman, 1997; Yardley, 2000).

Table 1 Codes developed based on participants’ responses – pedagogical dimension. We display the core category and sub-category that we identified from the data analysis. The categories include both positive, negative and neutral themes. Specifically, we display the positive themes by underlining them with green colour, negative themes with red colour, neutral themes with black colour. Aside from each sub-category, we display how many distinct respondents mentioned it

Core Category	Sub-Category (Distinct Respondents Mentioning It)
IMPACTS ON PEDAGOGICAL ELEMENTS	
Learner Status	Decreased Learner Autonomy (8) Decreased Learning Independence (8) Forced Learning Initiatives (10) Passive Learning Pathway (10) Restricted Self-Exploration Freedom of How (16) and What (13) to Learn Manipulation (11) Overly Relying on AIEd (9)
Learning Environment & Experience	Intense Learning Environment (40) - Students generally live with fear (22) - Assume strict rules and supervision (18) - Assume more effort is needed (for imagined standard) (6) Disrupting Concentration (27) <u>More Engaged Class (13)</u> Result-Oriented (Scores) (13) Losing Learning/Exploration Enjoyment (6) Over-Digitalized Learning (6)
Educational Process & Approach	<u>Improve Self-Reflection Effectiveness (49)</u> <u>Provides Analytical Insights to Inform Education (21)</u> <u>Equity (Equal Attention) for Individuals (6)</u> <u>Friendly for Introverts with Active Help (5)</u>
Pedagogical Interaction, Relationship & Role	Decreased Interaction and Communication: among Students (12); between Students & Teachers (3) Affected Social & Educational Relationship: among students (9); Students & the teaching body (12) Role Changing: students (3), teachers (5), other teaching agents (e.g., AIEd) (17)

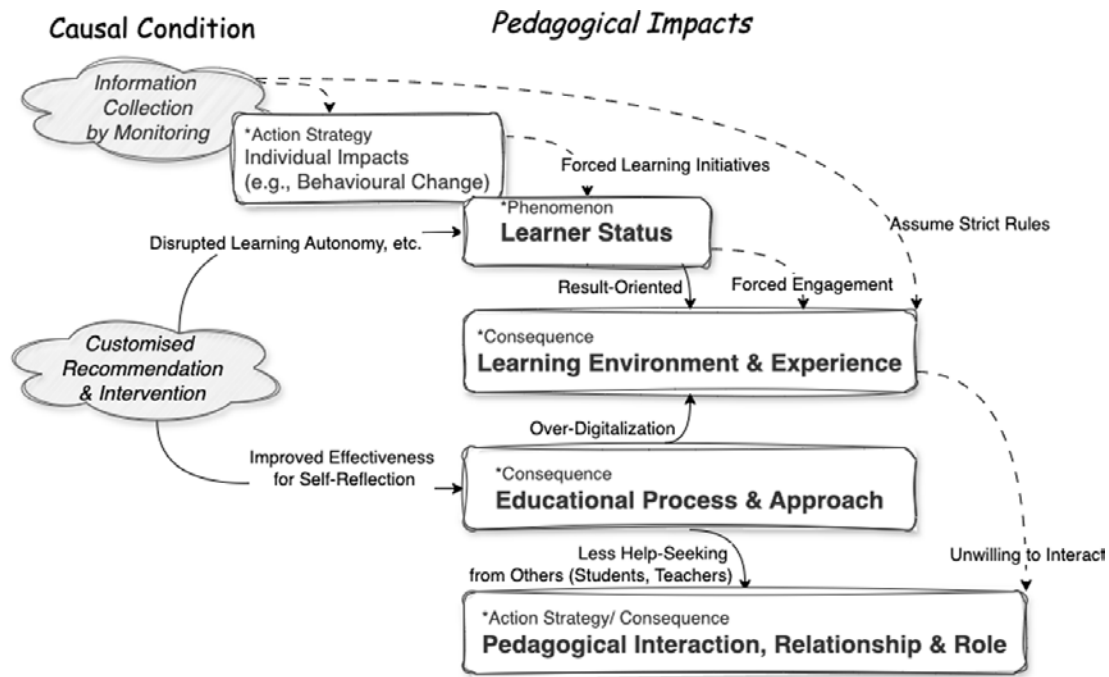


Fig. 1 *Impacts of AIED Adoption* - This figure summarizes our findings of the interplay among identified potential impacts. The relationship elements according to Grounded Theory techniques include Causal conditions, Phenomenon, Action strategies, and Consequences. The two causal conditions are Information Collection by Monitoring, and Customized Recommendation & Intervention. We labeled each identified main theme with the relationship element, such as *phenomenon. The likely impacts due to AI monitoring are shown in dashed lines (---); the potential impacts due to AIED customization are shown in solid lines (—). We adopted Botschen et al. (1999)'s criterion, considering the relationship elements mentioned by over one-third of respondents as significant. We excluded elements mentioned by fewer than one-third of the respondents

Results and Discussion

Analysis of the 3 story scenarios from the SCT tasks highlights the potential impact of AI adoption under four dimensions: Individual, Pedagogical, Technological, and System-design. These impacts seemed to interact with each other. In a prior study (Han et al., 2022), we demonstrated the impact on individuals due to their direct interaction with AI monitoring. We also assessed the challenges to technological efficiency and system-design acceptance, which will be presented in a future paper. The current paper describes primarily the pedagogical implications, using four core categories, see Table 1. In the following context, we demonstrate how individual impacts such as behavioral change can lead to pedagogical impacts and how they interact. In the text, we refer to Table 1 core categories in **bold** and the open codes in *italics*.

Learner Status

The first pedagogical implication identified through the SCT scenarios was the most frequently anticipated changes by participants in **learner status**. Participants perceived that AI monitoring and its associated functions such as feedback provision and adaptivity could pose a threat to their *Autonomy* and reduce their *Learning Independence*. As participant P24 mentioned: "Everyone is nervous knowing they are

watched”, and P1 mentioned: “It’s uncomfortable to be watched. Now I have to force myself to participate and express myself even though I indeed do not have any ideas in the group discussion...” This seems to suggest that fear of showing poor performance under monitoring can lead students to unauthentic behavioral changes, e.g., the engagement with a *Forced Learning Initiative*.

Similarly, some participants reported that automated scaffolding could leave them with *Passive Learning Pathways*, inhibiting them from carrying their own choice or *Self-Exploration Freedom for What to Learn and How to Learn*. e.g., P40 described that “students are forced to be incorporated into this system” and that the automated system often suggests learning pathways based on the assumption of “one established good standard” (P35). Participants perceived that such learning design could cause forced compliance and disruption to their original choice. For instance:

“He feels this early warning system is established to force everyone to learn. Learning itself is a personal matter: individuals should be able to choose how and when to learn, and even whether to learn (for a person who wants to drop out, he has the right to decide to do so).” (P35).

“He has a very personal learning style which he thinks suits him. He believes the university should be a place of freedom and self-realization. However, this system disrupts his usual learning habit and makes him less able to be what he has always comfortably been.” (P39).

Quite a few stories described more practical conflicts where a student whose curiosity or passion locates somewhere else. They perceived that customization could restrict their exploration of learning interests. They also perceived that the system interventions were *Manipulative* due to a *Forced or Passive Learning Pathway*, when “following all system-suggested guidance steps to study” (P45). The following quote exemplifies this perceived passive learning pathway:

“He enjoys studying because he has been able to pursue topics that are interesting to him - but now he will become a slave to the system and have to learn what it says... all their behaviours and knowledge, will be shaped by AI not themselves.” (P63).

By comparison, a few participants considered forced engagement as a positive push. For instance:

“It’s uncomfortable to be watched. Now I have to force myself to participate and express even though I indeed do not have any ideas in the group discussion... I found this lecture is actually more interesting than I thought.” (P1).

Other positive attitudes were expressed with trust and willingness to receive system-allocated learning resources as the participants believed in its analytical power. For instance, some participants believed that the system-generated resource allocation

may prevent a student from stress or anxiety by inhibiting the learning content that the system believes a student is not ready for.

Discussion Based on SCTs, our findings uncover a student perspective about disruptions to autonomy and when these effects may emerge. Several participants perceived that AIED depicted in the SCTs may pose a threat to their autonomy where they would not be at liberty to make choices of their own or to engage in self-exploratory behaviors. This mirrors existing concerns that interventions may limit students' self-determination or reduce learning independence (Whitelock-Wainwright et al., 2020). These concerns are also shared by other researchers indicating that AIED may undermine learner autonomy (Schumacher & Ifenthaler, 2018). It further extends an ethical issue "echo chambers" (Sclater, 2016), which relates to AI recommendations, that these personalized services can entrench personal traits.

Participants also perceived that AIED personalization can be manipulative, which resonates with another ethical issue "nudging". Nudge theory suggests using subtle methods to influence people's behaviors (Southgate et al., 2019), e.g., rather than telling people what to do, indirect methods of reinforcement can be used where certain options can be made more accessible. Existing evidence on nudging suggests that it can be perceived as a manipulation by students (Borenstein & Arkin, 2016), which supports our participants' concerns.

Lastly, in digital environments, students expect to be self-reliant and control their own learning (Roberts et al., 2016). A study on student vulnerability found that decisional interference such as recommended settings can often affect students' decision-making (Prinsloo & Slade, 2016a). A possible recommendation to address the autonomy and choice-making behaviors of learners is to provide options and encourage the learners for decision-making, even when their choice might disagree with the AIED recommendations.

Learning Environment and Experience

Our second pedagogical implication is the learning environment and experience, which is often expressed after describing the influenced learner status. Many participants believed that individuals are not encouraged to explore but must follow the instructions on what and how to learn. This undermines student enthusiasm for learning, as (P63) expressed: "A little bit more of his intellectual curiosity and spark dies as he starts to think along these lines (to learn what system told him to learn instead of topics of his interests)." Further, the learning experience can be joyful when the "intention of learning is out of the interests and desire for knowledge exploration" (P12). However, when the system removes independence, learning will "lose its *enjoyment*" (P38). This sense of diminished learning enthusiasm and influenced learner experience can be illustrated in the story from P44:

"The system has planned every step of learning for him. He no longer feels excited about exploring the world with uncertainty. He knows this study experience will be SMART, but no longer a fun journey...He feels he must behave

to meet the requirements of the system. The learning experience is no longer relaxing and rewarding as he knows his every move will be captured for analysis purposes...” (P44).

A few stories criticized the intervention that provides instant help for students’ learning encourages *Result-Oriented Education*: “learning under this system is... just to complete tasks and goals set by the system” (P26), “The learning process should not pursue efficiency (speed of learning for high scores). Knowledge obtaining should be polished and precipitated by time and thinking” (P34). Many participants also perceived their learning with these data-driven tools as *Over-Digitalizing Learning*: “This system makes his experience less personal and more transactional...” (P54), where students feel like “a piece of data” (P64) or “become a number...as it completely treats people as a goal rather than as a person.” (P45).

The majority of respondents expressed that the monitoring could dramatically influence their learning experience. The presence of AI monitoring can make the students *Assume Stricter Rules and Supervision*, and an expectation of compliance where students *Assume More Effort is Needed*. The learning environment becomes intimidating and intense rather than encouraging, where students frequently express that they will *Generally Live with Fear*. The sense of being watched *Disrupts their Concentration*. For example, one participant (P55) described this anxiety and the behavioral shifts, a common theme among participants:

“This is a bit scary... I would feel very nervous because I feel like all my actions are being monitored and recorded and my every move affects my grades. So I would try to behave in an unnatural way/ in a way that I think will give me high marks from the perspective of the machine..Now with the additional voice recognition and eye tracking, I would feel even more nervous since not paying 100% attention at all times will be tracked, and therefore can impact my grades.” (P55).

Nevertheless, for a few participants, monitoring positively *Enhances Class Engagement* due to forced performativity. One respondent illustrated this view:

“Even Kevin, the naughty boy who is usually the most difficult to engage in class, is actively participating in the discussion and expressing his views..Charlie felt that the classroom behaviour analysis system seemed to play a positive role in promoting learning in the classroom.” (P21).

Discussion Respondents’ concerns about AI-enhanced outcome-oriented learning, echo existing arguments such as risks of infantilizing learning in higher education (Sclater, 2016) and spoon-fed learning (Tzimas & Demetriadis, 2021). Most stories that mentioned the sense of disruptions also expressed a preference for learning on their own. These students perceived their learning as hindered by automatic interventions. They believed that active learning opportunities are not provided if they cannot be the creators of knowledge and felt that they are treated as passive learners who can only take information passively. This is against the pedagogical concepts of con-

structivism (Piaget, 1977) which is one of the fundamental theories for educational technologies. Hence, it is vital that AIEd should encourage students to engage in exploratory behaviors where they take charge of their own learning. Meanwhile, education with emerging educational technology should cultivate individuality (Clayton & Halliday, 2017). The future role of teachers should be to assist students in developing self-regulatory behaviors where students set goals for themselves and adjust learning processes accordingly (Van Der Graaf et al., 2021).

Educational Processes and Approaches

Our third pedagogical implication of AIEd is **educational processes and approaches**, often expressed with positive attitudes. A lot of participants embraced the AIEd's personalization depicted in the SCT, foreseeing *Improved Self-Reflection Effectiveness* by targeting their weaknesses. Individuals can *receive undivided Equal Attention* from AIEd, whereas teachers' time and energy per student are limited. Some participants also believe customized intervention is more *Friendly for Introverts* as the system actively offers help. For instance:

“As an introverted person, he may be pleased with this system. He is always too shy to take the initiative to ask questions. The system can identify his struggles and hint to teachers to provide further help... Compared with communicating with people, he'd love to provide his data to receive the customized service.” (P31).

“Charlie has been struggling with gaining good grades even though he studied hard... The university teachers are knowledgeable but... fail to deliver their knowledge effectively as there are so many students in each lecture. Or they don't care whether students understand... Charlie is excited about this system as it takes care of every student by providing targeted help. The system will... remind the teacher to elaborate on difficult concepts if it detects many confused faces.” (P36).

Quite a few respondents expressed an unwillingness for the AIEd depicted SCT to detect their learning weaknesses. The primary concern was its exposure to others (including teachers, other students, and the system), as weakness is seen as part of individual privacy. On the contrary, a few participants favor the weakness analysis for seeing its shortcoming-targeted service for enhancing self-learning reflection. For instance, participant P60 described his increased learning confidence when studying with the recommended study plan targeting solving his weakness, after he agreed on the captured weakness and perceived the identification was accurate and effective:

“Charlie looks over at his suggested study plan. He thinks to himself that the AI has captured his skills and weaknesses quite accurately and begins to study with greater confidence.” (P60).

Most students perceive AIED can assist both teachers and students by *Providing Analytical Insights to Inform Education*.

Discussion Similar to our participants' expectation of AIED nudging teachers to offer assistance, Holstein's experiment in K-12 classrooms (Holstein et al., 2018) illustrated the potential for AIED to direct teacher attention to students with lower prior ability. His human-AI hybrid real-time analysis of student learning converged learning outcomes across students of varying learning abilities. His findings might indicate the possibility of solving the complaints from our data that, struggling students wish to receive more assistance which might be overlooked by teachers themselves with limited attention.

Our study demonstrates students' conflicting perspectives on acceptance due to their different expectations of AI customization services. These negative attitudes demonstrate clear barriers to AIED uptake, which the AIED community needs to embrace in its frameworks. This divergence is also seen in a recent investigation into students' attitudes towards AIED (Whitelock-Wainwright et al., 2021): low expectations of AIED are linked to perceptions of its stifling independence, while high expectations are linked with hopes of AIED to potentially assist particular needs. The limitations of adopting AI-based tools have not been well-understood [47], for example, regular feedback provided to large student cohorts with varying demographics can be perceived as unhelpful especially by students with stronger self-regulatory skills (Winne, 2017). This then raises the question of how to balance the benefits and harms of AIED according to students' varying expectations and needs.

Affected Interaction, Pedagogical Relationship, and Roles

During the actual learning practices, participants frequently anticipate the disruption of **interaction, relationships, and roles**.

In the context of greater monitoring, quite a few students may find informal interactions such as small talk with other students uncomfortable. E.g., "His classmates were acting unnaturally, without the usual humming of greetings and chattering discussions." (P24). Students' interactions can become inauthentic, such as "losing the gentle way of communication" (P2). Students can also become self-cautious of their expressions, e.g., "No one dares to really express their thoughts anymore" (P52).

These *Decreased (Social) Interactions and Communication among Students* include both learning and non-learning-related behaviors, which often result from students' fear of inappropriate performance being recorded. Those changes are often shown with forced performativity under monitoring, where students feel the need to pander to expectations of tone and expression. The following stories provide clear examples:

"The group discussion begins. He noticed that the communication in the group lost the cheerful tone. Everyone looked solemn and expressed their opinions concisely and seriously... he flipped through the materials and pretended to read rapidly during the discussion, hoping to convince the system that he had

mastered the knowledge. But it seems not a good attempt that he made himself in a muddle...when the teacher announced the end of the discussion, everyone felt relieved...” (P7).

“...they all have an initial tacit understanding of how they all should behave... and be cautious of expressions to avoid low evaluation scores. As a result, everyone seems to be discussing confidently, but in fact they were only complimenting each other’s ideas and being very cautious about their words. Charlie was very disappointed at first...he joined the discussion with his classmates and started their pretended fierce discussion.” (P52).

Students’ Pedagogical Relationships are thereby affected compared with when students are used to learning from the conversations with each other:

“He doesn’t interact with his fellow students anymore, in case the system sees this as time-wasting or that he is lacking confidence. He really misses that interaction - where he learned from his peers a lot” (P63).

Since the intervention can provide automatic learning assistance, struggling students might decrease their help-seeking interactions with peers, a negative consequence given that these interactions have been shown to benefit both stronger and weaker students (Boud & Cohen, 2014). The dominance of technology can further influence *Students’ Pedagogical Roles*: peers could once educate each other, but now “students are more like competitors against each other” (P26). A broader opinion of the affected student’s role was that students are more “being managed but not to be treated” (P15).

Similarly, the direct *Interaction between Students and Teachers* might decrease due to the intervention once students trust AIED effectiveness and rely on AIED-provided help. This fragment shows a student game the AIED system to obtain attention and receive teacher assistance:

“...he knows that the system might help the teacher assist him so he might... play up his weaknesses to make the AI think that he needs further help... Because he is aware the system tracks his progress, he feels less...need to ask the teacher directly for help...the academic profile will inform the teacher either way.” (P58).

Nevertheless, in this way, AIED might create an indirect communication channel between students and teachers, thereby becoming a medium that delivers students’ wishes for more attention.

Participants frequently expressed their concerns that AIED will make *Teacher-Student Relationships* impersonal. For example:

“Since the system has tools over some of the teachers’ administrative roles such as attendance check and providing support, there is no longer close contact with teachers. Will teachers still bother to know him?” (P63).

Discussion Automatic learning assistance may weaken students' pedagogical roles such as peer learners. Peer learning is powerful, moving learners "beyond independent to interdependent or mutual learning" (Boud & Cohen, 2014). Once students perceive AIED assistance as efficient for problem-solving, established interactions among students would be reduced. This might covertly subvert their pedagogical relationships of peer-learning. AI-assisted learning environments should create spaces for facilitating human relationships (in the classroom) by enhancing human interactions.

How Impacts Interact and Shape the Perceived Learning Landscape: Implications for AIED Design

The preceding sections have outlined the potential impacts of AIED adoption on students. These identified impacts not only exist in isolation but may also interact with one another. In this section, we synthesize key findings to illustrate the interplay of these effects within AI-enhanced educational environments. We also discuss how these interactions can influence the overall learning landscape. We aim to offer insights into how AIED may shape learning environments on a larger scale and inform better and more effective educational technology design. Finally, we present design implications that focus on student-centered approaches, addressing key concerns identified from our findings to support more thoughtful and ethical AIED development.

Interactions Among Impacts

According to the Grounded Theory techniques and coding structure, causal conditions are presented in *Italics*, and the identified main themes are presented in **Bold**. The identified phenomena, consequences, with their cause-effect relationships between identified impacts are summarized in Fig. 1.

When AIED collects student data by monitoring: students may alter their behavior to align with perceived expectations, leading to unauthentic engagement in learning activities that affects Learner Status (Sect. 4.1). Monitoring may also shift students' perceptions of the Learning Environment such as assuming strict rules (Sect. 4.2), which can further discourage social Interactions with other students (Sect. 4.3). Nevertheless, this monitoring can positively enhance engagement within the learning environment and offer insights into progress.

When AIED delivers customization: it may disrupt Learner Autonomy (Sect. 4.1), as students might feel compelled to follow recommended pathways, thus reducing exploratory learning, and influencing their Learning Experience (Sect. 4.3). Nevertheless, the customized analysis and AIED recommendations were perceived to enhance self-reflection for their learning, thereby increasing the effectiveness of the Educational Process and Approach. As students might trust AIED's analytical power and rely on its assistance, it can influence educational Interactions and Relationships between students and teachers and among students, and Pedagogical Roles (Sect. 4.4).

Our findings suggest that AIEd adoption dynamically reshapes educational interactions and beliefs in the learning environment. These changes align with the social anthropological view that technology adoption within a social context continually transforms human practices and perceptions (Suchman, 1987). The consequences are unforeseeable but must be accommodated.

Implications for Future Educational Landscape with AIEd

Our study reveals a diverse range of student perspectives on AIEd, often differing or even diametrically opposed, depending on individual learning preferences, needs and expectations. This diversity underscores the need for future educational landscapes shaped by AIEd to accommodate and respect individual differences through flexible, customizable options. Specifically, a significant concern among students relates to how AIEd might infringe on their learning autonomy and self-determination. The desire for control over their educational journey is paramount (Roberts et al., 2016). To foster trust and reduce potential resistance to AIEd-facilitated personalized learning (Holmes et al., 2018a, b), it is crucial to prioritize features that preserve and promote student autonomy.

Transparency and consent are also vital in addressing students' apprehensions about AI monitoring and data collection. Our findings indicate that students often fear constant monitoring might expose their shortcomings, potentially leading to increased anxiety and stress. Therefore, establishing a trust-based environment is essential. This can be achieved by clearly communicating what data will be collected (Corrin et al., 2022), how evaluations are conducted (e.g., AI mechanism, evaluation criteria), and the implications of these assessments, as well as clarify that making mistakes and testing ideas is a valued part of the learning process.

The AIEd community should also foster an inclusive, student-centered design process (Strohmayr et al., 2019). By involving students in the development and refinement of AI tools, educators and designers can ensure that these technologies align with a broad spectrum of student needs and perspectives. This collaborative approach not only helps in addressing specific student concerns but also empowers students by giving them a role in shaping the technologies they use. Ultimately, a commitment to continuous dialogue among developers, educators, and learners will be vital to ensure that AI tools are both ethical and effective in educational settings.

Design Implications for AIEd Development

Building on these broader implications, our study offers specific design implications to guide student-centered AI design in educational settings:

- **Support for Autonomy and Choice:** AIEd systems should incorporate features that allow for student choice and self-direction. For instance, AIEd design should enable students to control or limit the data collected on their activities and offer opt-in choices for different levels of AI interaction. For example, allowing students to select which data they share for AIEd and how it is used to strengthen their sense of agency.

- **Transparency in AI Monitoring and Feedback:** Students expressed concerns about feeling monitored. AIED systems should be transparent about the data they collect and how it is used, which could be achieved through clear, user-friendly dashboards that show students what data is being collected and offer explanations about how it impacts their learning.
- **Inclusive Feedback Mechanisms:** Since feedback can be perceived differently by students based on their learning goals, AIED systems should provide customizable feedback settings (Tsai et al., 2021). This could include options for students to adjust the frequency, tone, and type of feedback, thereby fostering a more supportive and personalized learning experience.
- **Ethically Grounded Design for Diverse Learning Environments:** AIED should consider the socio-technical implications that can vary widely across different educational settings. Designing systems that respect cultural and individual differences (Winstone et al., 2016) can help AIED adapt to diverse learning environments, thereby reducing the risk of unintended harm or bias. This approach may also enhance the effectiveness of AIED adoption; for instance, AI-assisted feedback might be perceived positively in some cultures but seen as disruptive to learning autonomy in others.
- **Encouraging Active Learning:** To counter the risk of students becoming overly reliant on AI hints and support, AIED systems should include mechanisms that encourage active learning. This could involve gradually reducing AI assistance as students build skills, or offering prompts that encourage students to reflect on their learning processes before receiving AI-generated hints.

Limitations and Future Work

This study employed the story completion method to explore students' perceptions of AI-enhanced educational environments. Story prompts might have unanticipated impacts on participants' responses. For instance, terms such as "monitoring" might lead to certain narratives. To mitigate this, we explicitly clarified in the participant introductory materials that our research objective is not to scrutinize AIED technology negatively. Instead, our goal is to uncover any potential concerns and perspectives that could inform more ethically responsible and improved AIED designs. While existing SCM literature does not explicitly address the influence of word choice within story stems on participants' responses, it emphasizes the importance of careful stem design to ensure clarity and avoid narrowing responses (Clarke et al., 2019). Although SCM is a well-established method in psychology and medical research (Pong et al., 2024), most studies do not pre-test story stems for bias. The key limitations discussed typically focus on the interpretation of findings rather than potential bias introduced by the prompts themselves. A commonly discussed concern with SCM is social desirability bias, which can be mitigated by the third-person perspective storytelling approach. Nonetheless, we acknowledge the potential story bias represent a gap in the cur-

rent SCM literature and should be a consideration for future research using this method.

In future work, an A/B testing approach could be employed to explore how different word choices may influence students' perceptions and responses to the scenarios. This approach could also provide insights into how more thoughtful phrasing and framing of AIEd concepts might encourage better acceptance and more positive responses toward its adoption.

Given that our study specifically focuses on students' perspectives in the context of higher education, we encourage future studies to broaden the scope to include perspectives of other stakeholders such as institutions and leadership, and to explore other contexts like K-12 education. This will facilitate a more comprehensive understanding of the attitudes and concerns toward AIEd adoption.

Future research should also delve deeper to examine how students' academic disciplines and varying levels of AI literacy influence their perceptions of AI use in educational settings. These nuanced insights can further inform the ethical design and implementation of AI in education, ensuring it aligns with the diverse needs and perspectives of the student body.

Conclusion

Our study sheds light on the ethical and pedagogical implications of AIEd from a student-centric perspective, an area that has been insufficiently explored in prior research. Students are the primary data producers, the main subject of analytics, and the beneficiaries of analytics-driven insights. They must be placed at the heart of AI adoption and their concerns should be addressed. Our research reveals students' perceptions of the potential impact of AIEd on learner status, learning environment and approaches, interactions and relationships, and pedagogical roles. We highlight that these aspects not only exist independently, but also interact and influence one another, weaving a complex web of ethical and sociotechnical considerations. As AI continues to shape the future of education, it is crucial that we prioritize the ethical considerations and perspectives of those most directly impacted- the students themselves. This approach ensures that the evolution of AIEd aligns with the students' and teachers' needs and fundamental pedagogical values, thereby fostering an equitable, effective and ethically responsible educational future.

Appendix

Table 2 *Coding example* - this table provides examples of how we conduct coding and identify the interplay elements from participants' stories. The example below is from an actual participant's story. The sentence identified as a code is underlined; the keyword identified as a casual condition is underlined with curly lines

Example Story from Participant 31	Codes	Interplay Elements
As an <u>introverted person</u>, he may be pleased with this system. He is always too shy to take the initiative to ask questions. The system can identify his struggles and hint to teachers to provide further help. Compared with communicating with people, he'd love to provide his data to receive the <u>customized service</u>. With its analysis, teachers can then make appropriate adjustments on their teaching, it sounds helpful. But Charlie doesn't like the customized study plan formulated from AI, he prefers to learn in his own way rather than being restricted by AI's automatic ones. He just prefers to trust his own methods. However, Charlie also has some concerns. He worries that he may not always be able to maintain full concentration in class. If his mind wanders, it could potentially affect his grades or evaluations as AI could detect that. The class vibes are a bit scary under real-time monitoring and evaluation. He feels that he needs to put more effort on focusing and engaging on the class. But Charlie is smart, he might pretend to read by staring at the slides while actually he isn't paying attention and thinking about the girl he had crush with. Charlie also tried to have fewer casual talks with her peers as they are all trying to "behave" under the AI monitor, they don't want to be identified with too many records of non-learning behaviour. Without those interactions, he feels that he and his classmates become pure competitors rather than peers, as they are just numbers in the ranking system.	[Friendly for Introverts with Active Help] [Provides Analytical Insights to Inform Education] [Restricted Self Exploration Freedom of How to Learn] [Assume Strict Rules and Supervision] [Assume more effort is needed (for imagined standard)] [Intense Learning Environment] [Individual Dimension - Behavioural Change] [Decreased Interaction and Communication: among Students] [Pedagogical Roles]	Casual Condition: Customization Casual Condition: Monitoring Phenomenon Phenomenon/ Action Strategies Action Strategies Consequences

Table 3 *Codes with example quotes* - this table provides sample quotes for each identified code that is displayed in Table 1

Codes	Quotes
Learner Status	
Decreased Learner Autonomy	I feel like my pursuit of achievements doesn't come from my own will, but to according to what the system suggest, like how long should I stay focus on the lecture. And I am forced to be incorporated, otherwise I will be punished for scores. My autonomy was surrender. (P26)
Decreased Learning Independence	What if future students over rely on AI's automatic interventions since it is so handy, students might just don't take time to think to solve the problems but waiting for AI to hint the next step. "Only pursue results by using AI but not to take time with thinking by ourselves" (P30)
Forced Learning Initiatives	Under monitoring, students are just forced to learn and participate, because they are frightened if the AI catches their poor performance being recorded. I feel that I am not learning for myself but to completing tasks to satisfy the evaluation system. (P8)
Passive Learning Pathway	"Charlie has always believed that the most fundamental reason why tools are tools is that they can be controlled by humans, rather than letting tools determine human behavior. The development of this system is based on a high-sounding purpose, but in essence it is to control people rather than to help people. The final result is to let people walk according to the established "good" way, which greatly obliterates people's will. (P45)

Table 3 *Codes with example quotes* - this table provides sample quotes for each identified code that is displayed in Table 1

Codes	Quotes
Restricted Self-Exploration Freedom of How and What to Learn	What to Learn: “One of the missions of education at school is to provide the various possibility of education, to endow it with the greatest fairness, without presuppositions and restrictions. It is very likely that some students will miss their own possibility under this AI system. For example, perhaps student A’s academic weakness is programming, and his strengths areas are at art, history and painting. However, A actually likes learning programming and logic. A would rather get middle or even lower scores in subjects that he is not good at. Even so, it is more rewarding for A to feel fulfilled than to achieve the upstream level in the field of learning art or history. The learning system can customize individual learning plans and educational routes for each student, and may also automatically generate customized learning materials and elective courses recommended for students; but they can never manipulate a person’s true heart and inner driving interest. And this internal drive will change. So don’t set limits to it. Under this AI system, it is possible to miss the self-achievement of this student A, or miss the free growth and choice of another student B or student C.” (P34) How to Learn: “Charlie has a very personal learning style for which he thinks is best most suitable for him. And Charlie also believes that the university is supposed to be a place of freedom, self-exploration and self-realisation. However, something like a real-time classroom performance evaluation system might disrupt his usual learning habit and makes him less able to ‘be what he has always been’.” (P39) “Instead of stimulating students’ curiosity and desire for exploration of knowledge, the university use monitoring and quantification to characterize students and criticize students for not focusing enough on class. It is ridiculous! We have different ways of thinking, and different speed of problem solving. The intention of taking this course should be our interest.” (P12)
Manipulation	“AI’s intervention with analysis might cause me to deviate from my original needs and lead me to accept its guidance. I hope that I am the one using the tools, rather than being instrumentalized and manipulated by the technology.” (P28)
Overly Relying on AIED	“Chris (The lecturer) knows that there are students who can seem to have low engagement, then pull on a massive study blitz just before the exam, and get high scores in the exam. But these students make him uneasy, as he doesn’t believe that they have deep learning in the subject. He tries to remember back over the semester about this student. Before the system was in place he used to take note of students who didn’t seem to be engaging, and try to get a feel for what they were thinking. But he has become so used to relying on the system to find these students and offer them a guided path that he no longer has that close contact with his students. He wonders if professors are really needed any more.” (P63)
Learning Experience & Environment	
Losing Learning/Exploration Enjoyment	“Under strict system monitoring, group discussions will lose its enjoyment. Everyone is pretending with good performance and tried so hard. He used to be chilling in group discussions and feel free to express Himself, test out ideas and talk freely with his friends and peers. But He bites his lip, worried about how surveillance in the classroom might ruin this aspect for him. He couldn’t chill out anymore.” (P67)
Result-Oriented (Scores)	“He always feels that learning under this system is not to acquire knowledge, but to complete tasks and goals. He believes that learning should be free to explore, like where Charlie’s passion locate, but not only pursues high scores.” (P49)
Over-Digitalized Learning	“To Charlie, this feels like yet another thing that is being implemented by the university to make his experience less personal and more transactional. Will the teacher even bother getting to know him if the machine is taking the attendance and marking the class participation. He feels like he has become a number rather than an individual.” (P54)

Table 3 *Codes with example quotes* - this table provides sample quotes for each identified code that is displayed in Table 1

Codes	Quotes
Disrupting Concentration	“Attempts to speak up in order to be recognised by the voice recognition AI. He might either skim through parts he would have otherwise taken longer to read or waste time on things he’d otherwise skim through to show that he is taking in all the material equally. Again, Charlie’s attention is dedicated more to how the AI will perceive him rather than the actual discussions in class. Due to this, he isn’t actively listening to his classmates and is instead trying to think of interesting and thoughtful things to add to the conversation. He feels uncomfortable with silences and has the need to fill in gaps with conversation to show he is engaged. He begins to wonder what the profiles of his fellow classmates look like.” (P58) “This monitoring thing is so disturbing! I can’t even focus on the materials for discussion with the idea of ‘someone is watching me’ haunting in my mind, it is watching every single detail!... I wish I did not learn anything about this whole setup. Without knowing that I’m being watched, I may focus more on my task right now...” (P29)
Intense Learning Environment	“I feel like AI is watching me all the time, I feel anxious to learn in the classroom.” (P9)
- Students generally live with fear	“The system knows everything, which makes Charlie feel scared.” (P48)
- Assume strict rules and supervision	“There is no way to loaf on the study as AI is monitoring and recording everything. Even tigers sometimes doze. Anyhow the sluggards on the class will be seen and caught with an ironclad evidence. How can I avoid laziness for every second of my living as long as I am still breathing” (P26)
- Assume more effort is needed (for imagined standard)	“I feel overwhelmed and worries about if I would get an early warning since I find the topic on algebra quite hard... I wouldn’t feel free to make mistakes as every mistake will be recorded. I would try my best to pretend to show a lot of confidence during the moments that I am being tracked. This will also encourage me to practice algebra at home where I know that I am not being tracked and I free to think and make mistakes” (P55) “With this system, students must make much more effort than before to deal with the courses, and that happens to the teachers too. Life will not be easy anymore for both of them... Charlie knows he had to perform so well to get a good score. He is worried. He feels he can’t let his guard down and must be constantly “switched on”. (P6)
More Engaged Class	“Charlie forced himself to participate in the conversation although he didn’t have any idea around the topic. He participated just because he’s being monitored, and he wants to behave to meet the standard. Charlie and his classmates shared some random fact about the topic and started to roll their eyes at each other as the whole thing starts to become absurd... The class discussion is seemingly more energetic than before.” (P50)
Educational Processes & Approaches	
Improve Self-Reflection Effectiveness	“It (the personalized analysis and recommendation) helps me studying with focus – my strength and weakness. It can guide me to optimize my learning methods and improve my shortcomings.” (CH11)
Equity (Equal Attention) for Individuals	“There are too many students in the lecture, he always felt that lecturers are more like presenters but not educators. He takes every lecture seriously but he found the knowledge of university level is too difficult and divergent. He believes the system can ensure every student to be taken care of. Since the system can accurately capture the emotions of each student, if the student does not understand a certain concept, the system will surely remind the teacher to elaborate again. This is a more scientific method to transfer knowledge, students will really benefit a lot” (P30)
Friendly for Introverts with Active Help	“He is a shy student. He is often not confident, to answer and ask questions in public. He believes the system is thoughtful, can detect if he is struggled and help. He doesn’t have to feel embarrassed to ask simple questions to others.” (P1)

Table 3 Codes with example quotes - this table provides sample quotes for each identified code that is displayed in Table 1

Codes	Quotes
Provides Analytical Insights to Inform Education	“Teachers were excited about the system at first - it offered really interesting insights into students’ learning and how to engage with students and to improve their experience.” (P64)
Educational Interaction, Pedagogical Relationship & Roles	
Decreased Interaction and Communication: among Students; between Students & Teachers	Among Students: “He doesn’t dare show his real emotions but tries to look intelligent and interested all the time. But he’s so worried about how he will seem to the system that he isn’t learning as much as he used to. He doesn’t interact with his fellow students anymore, in case the system sees this as time wasting or that he is lacking confidence. He really misses that interaction - that’s where he learnt a lot.” (P63) Between Student and Teacher: “He doesn’t feel the need to send his lecturer the email. Instead, Charlie proceeds with the suggested study plan. Depending on how affective it is, this will either reinforce his confidence in the AI study plan and therefore lead him to asking less questions from the lecturer.” (P58)
Affected Social & Educational Relationship: among students; Students & the teaching body	Among Students: “He noticed the monitoring camera tried to avoid long greetings with his mates...” (P11) Between Student and Teacher: “He is aware of the AI system tracking his progress, he feels less of a need to ask the teacher directly for help because he know the academic profile will inform the teacher either way.” (P58)
Role Changing: students, teachers, other teaching agents (e.g., AIEd)	Students: “I feel we are more like competitors with each other since we are just a number being ranked in AI systems, rather than collaborators as we barely share and talk on class.” (P12) Teachers: “The current teachers are too easy to work, almost no need to spend time to observe and examine the students’ learning conditions, everything is done by the new education system! Conversely, such a system has higher requirements for teachers, who must have more professional teaching skills to become educational strategists and designers” (P13) Other Teaching Agents: “Teachers sometimes have favoured students for evaluation. AI can play as an un-biased 3rd party providing justice when evaluating.” (P58)

Funding This study involves no funding.

Declarations

Competing interests This study involves conflicts of interest.

References

- Ahmad, A., Schneider, J., Griffiths, D., Biedermann, D., Schiffner, D., Greller, W., & Drachsler, H. (2022). Connecting the dots—A literature review on learning analytics indicators from a learning design perspective. *Journal of Computer Assisted Learning*.
- Altun, D., Krauss, C., Streicher, A., Mueller, C., Atorf, D., Rerhaye, L., & Kunde, D. (2022). Lessons Learned from Creating, Implementing and Evaluating Assisted E-Learning Incorporating Adaptivity, Recommendations and Learning Analytics. *International Conference on Human-Computer Interaction*.
- Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092.

- Baker, R. S., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in gaming the system behavior in interactive learning environments. *Journal of Interactive Learning Research*, 19(2), 185–224.
- Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4), 223–241.
- Bauer, M. W., Gaskell, G., & Allum, N. C. (2000). Quality, quantity and knowledge interests: Avoiding confusions. *Qualitative researching with text, image and sound*, 3–18.
- Blythe, M., Encinas, E., Kaye, J., Avery, M. L., McCabe, R., & Andersen, K. (2018). *Imaginary Design Workbooks: Constructive Criticism and Practical Provocation* Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada. <https://doi.org/10.1145/3173574.3173807>
- Borenstein, J., & Arkin, R. (2016). Robotic nudges: The ethics of engineering a more socially just human being. *Science and Engineering Ethics*, 22(1), 31–46.
- Botschen, G., Thelen, E. M., & Pieters, R. (1999). Using means-end structures for benefit segmentation: An application to services. *European Journal of Marketing*, 33(1/2), 38–58.
- Boud, D., & Cohen, R. (2014). *Peer learning in higher education: Learning from and with each other*. Routledge.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676x.2019.1628806>
- Chen, X., Zou, D., Xie, H., Chen, G., Lin, J., & Cheng, G. (2022). Exploring contributors, collaborations, and research topics in educational technology: A joint analysis of mainstream conferences. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-022-11209-y>
- Clark, J. A., Liu, Y., & Isaias, P. (2020). Critical success factors for implementing learning analytics in higher education: A mixed-method inquiry. *Australasian Journal of Educational Technology*, 36(6), 89–106.
- Clarke, V., Braun, V., Frith, H., & Moller, N. (2019). *Editorial introduction to the special issue: Using story completion methods in qualitative research* (Vol. 16, pp. 1–20). Taylor & Francis.
- Clayton, M., & Halliday, D. (2017). Big data and the liberal conception of education. *Theory and Research in Education*, 15(3), 290–305. <https://doi.org/10.1177/1477878517734450>
- Colvin, C., Rogers, T., Wade, A., Dawson, S., Gasevic, D., Shum, B., Nelson, S., Alexander, K., Lockyer, S., L., & Kennedy, G. (2016). *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement*. Australian Government - Office for Learning and Teaching.
- Corrin, L., de Barba, P. G., & Brooker, A. (2022). Understanding the learner perspective to inform institutional learning analytics strategy and practice. *Handbook of Digital Higher Education* (pp. 248–259). Edward Elgar Publishing.
- Dunne, A., & Raby, F. (2013). *Speculative everything: Design, fiction, and social dreaming*. MIT Press.
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. <https://doi.org/10.1007/s11528-014-0822-x>
- Gašević, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics*, 4(2), 113–128. <https://doi.org/10.18608/jla.2017.42.10>
- Glaser, B., & Strauss, A. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.
- González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial Intelligence for Student Assessment: A systematic review. *Applied Sciences*, 11(12), 5467. <https://doi.org/10.3390/app11125467>
- Hakimi, L., Eynon, R., & Murphy, V. A. (2021). The Ethics of Using Digital Trace Data in Education: A thematic review of the Research Landscape. *Review of Educational Research*, 00346543211020116.
- Han, B., Buchanan, G., & McKay, D. (2022). Learning in the Panopticon: Examining the Potential Impacts of AI Monitoring on Students. Proceedings of the 34th Australian Conference on Human-Computer Interaction.
- Holmes, W., Anastopoulou, S., Schaumburg, H., & Mavrikis, M. (2018a). Technology-enhanced personalised learning: Untangling the evidence.

- Holmes, W., Bektik, D., Woolf, B., & Luckin, R. (2018b). June 27–30). Ethics in AIED: Who cares? Artificial Intelligence in Education (AIED), London, UK.
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., & Bittencourt, I. I. (2021). Ethics of AI in education: Towards a community-wide Framework. *International Journal of Artificial Intelligence in Education*, 1–23.
- Holstein, K., McLaren, B. M., & Aleven, V. (2018). Student Learning Benefits of a Mixed-Reality Teacher Awareness Tool in AI-Enhanced Classrooms. In C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. du Boulay, *Artificial Intelligence in Education Cham*.
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. *Journal of Learning Analytics*, 6(2).
- Howell, J. A., Roberts, L. D., Seaman, K., & Gibson, D. C. (2018). Are we on our way to becoming a Helicopter University? Academics' views on learning analytics. *Technology Knowledge and Learning*, 23, 1–20. <https://doi.org/10.1007/s10758-017-9329-9>
- Khalil, M., Prinsloo, P., & Slade, S. (2022). *A Comparison of Learning Analytics Frameworks: a Systematic Review* LAK22: 12th International Learning Analytics and Knowledge Conference, Online, USA. <https://doi.org/10.1145/3506860.3506878>
- Kitzinger, C., & Powell, D. (1995). Engendering infidelity: Essentialist and social constructionist readings of a story completion task. *Feminism & Psychology*, 5(3), 345–372.
- Klein, C., Lester, J., Rangwala, H., & Johri, A. (2019). Learning analytics tools in higher education: Adoption at the intersection of institutional commitment and individual action. *The Review of Higher Education*, 42(2), 565–593. <https://doi.org/10.1353/rhe.2019.0007>
- Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education—a literature review. *Learning analytics: Fundaments, applications, and trends*, 1–23.
- Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sánchez, E. M. (2023). Uncovering blind spots in Education Ethics: Insights from a systematic literature review on Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-023-00384-9>
- Nawaz, S., Kennedy, G., Bailey, J., Mead, C., & Horodyskyj, L. (2018). Struggle town? Developing profiles of student confusion in simulation-based learning environments. 35th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education, ASCILITE.
- Nawaz, S., Kennedy, G., Bailey, J., & Mead, C. (2020). Moments of confusion in simulation-based learning environments. *Journal of Learning Analytics*, 7(3), 118–137.
- Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. (Trans A. Rosin). Viking.
- Pidgeon, N., & Henwood, K. (2013). Using grounded theory in psychological research. *Doing qualitative analysis in psychology* (pp. 245–273). Psychology.
- Pong, C. Y., Roberts, N. J., & Lum, E. (2024). The what, why, and how? Of story completion in health services research: A scoping review. *BMC Medical Research Methodology*, 24(1), 159. <https://doi.org/10.1186/s12874-024-02274-7>
- Popay, J., Rogers, A., & Williams, G. (1998). Rationale and standards for the Systematic Review of Qualitative Literature in Health Services Research. *Qualitative Health Research*, 8(3), 341–351. <https://doi.org/10.1177/104973239800800305>
- Prinsloo, P., & Slade, S. (2016a). Student Vulnerability, Agency and Learning analytics: An exploration. *Journal of Learning Analytics*, 3(1), 159–182. <https://doi.org/10.18608/jla.2016.31.10>
- Prinsloo, P., & Slade, S. (2016b). Student vulnerability, agency, and learning analytics: An exploration. *Journal of Learning Analytics*, 3(1), 159–182.
- Queiroga, E. M., Machado, B., Paragarino, M. F., Primo, V. R., T. T., & Cechinel, C. (2022). Early Prediction of At-Risk students in secondary education: A Countrywide K-12 Learning Analytics Initiative in Uruguay. *Information*, 13(9), 401.
- Roberts, L. D., Howell, J. A., Seaman, K., & Gibson, D. C. (2016). Student attitudes toward learning analytics in higher education: The fitbit version of the learning world. *Frontiers in psychology*, 7, 1959.
- Ryan, T., Henderson, M., Ryan, K., & Kennedy, G. (2021). Designing learner-centred text-based feedback: A rapid review and qualitative synthesis. *Assessment & Evaluation in Higher Education*, 46(6), 894–912.
- Schiff, D. (2022). Education for AI, not AI for education: The role of Education and Ethics in National AI policy strategies. *International Journal of Artificial Intelligence in Education*, 32, 527–563. <https://doi.org/10.1007/s40593-021-00270-2>

- Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, 78, 397–407.
- Sclater, N. (2016). Developing a code of practice for learning analytics. *Journal of Learning Analytics*, 3(1), 16–42. <https://doi.org/10.18608/jla.2016.31.3>
- Seale, C., & Silverman, D. (1997). Ensuring rigour in qualitative research. *The European Journal of Public Health*, 7(4), 379–384.
- Siemens, G., Dawson, S., & Lynch, G. (2013). Improving the quality and productivity of the higher education sector. *Policy and strategy for Systems-Level Deployment of Learning Analytics*. Canberra, Australia: Society for Learning Analytics Research for the Australian Office for Learning and Teaching, 31.
- Southgate, E., Blackmore, K., Pieschl, S., Grimes, S., McGuire, J., & Smithers, K. (2019). Artificial intelligence and emerging technologies in schools.
- Strohmayr, A., MacArthur, C., Spors, V., Muller, M., Vigil-Hayes, M., & Alabdulqader, E. (2019). CHIInclusion: Working toward a more inclusive HCI Community Extended abstracts of the 2019 CHI Conference on human factors in Computing systems. Glasgow Scotland Uk. <https://doi.org/10.1145/3290607.3299012>
- Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press.
- Swist, T., Shum, B., S., & Gulson, K. N. (2024). Co-producing AIED Ethics under Lockdown: An empirical study of deliberative democracy in action. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-023-00380-z>
- Teasley, S. D. (2017). Student facing dashboards: One size fits all? *Technology, Knowledge and Learning*, 22(3), 377–384. <https://doi.org/10.1007/s10758-017-9314-3>
- Terry, G., & Hayfield, N. (2020). Reflexive thematic analysis. *Handbook of Qualitative Research in Education* Edward Elgar Publishing Limited, 430–441.
- Troiano, G. M., Wood, M., & Hartevelt, C. (2020). And This, Kids, Is How I Met Your Mother: Consumerist, Mundane, and Uncanny Futures with Sex Robots. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.
- Tsai, Y. S., Rates, D., Moreno-Marcos, P. M., Muñoz-Merino, P. J., Jivet, I., Scheffel, M., Drachsler, H., Kloos, C. D., & Gašević, D. (2020). Learning analytics in European higher education—trends and barriers. *Computers & Education*, 155, 103933.
- Tsai, Y. S., Mello, R. F., Jovanović, J., & Gašević, D. (2021). *Student appreciation of data-driven feedback: A pilot study on OnTask LAK21*: 11th International Learning Analytics and Knowledge Conference, Irvine, CA, USA. <https://doi.org/10.1145/3448139.3448212>
- Tzimas, D., & Demetriadis, S. (2021). Ethical issues in learning analytics: A review of the field. *Educational Technology Research and Development*, 69(2), 1101–1133.
- Vallor, S. (2018). *An Ethical Toolkit for Engineering/Design Practice*. <https://www.scu.edu/ethics-in-technology-practice/ethical-toolkit/>
- Van Der Graaf, J., Lim, L., Fan, Y., Kilgour, J., Moore, J., Bannert, M., Gašević, D., & Molenaar, I. (2021). Do instrumentation tools capture self-regulated learning? International learning analytics and knowledge (LAK) conference.
- Vehmas, J., Khan, A. I., Kaliteevskii, V., & Chechurin, L. (2022). Learning analytics overview: Academic Approach and Machine Learning possibilities. *Digital Teaching and Learning in Higher Education* (pp. 123–143). Springer.
- Walker, D., & Myrick, F. (2006). Grounded theory: An exploration of process and procedure. *Qualitative Health Research*, 16(4), 547–559.
- Whitlock-Wainwright, A., Gašević, D., Tsai, Y. S., Drachsler, H., Scheffel, M., Muñoz-Merino, P. J., Tammets, K., & Delgado Kloos, C. (2020). Assessing the validity of a learning analytics expectation instrument: A multinational study. *Journal of Computer Assisted Learning*, 36(2), 209–240.
- Whitlock-Wainwright, A., Tsai, Y. S., Drachsler, H., Scheffel, M., & Gašević, D. (2021). An exploratory latent class analysis of student expectations towards learning analytics services. *The Internet and Higher Education*, 100818.
- Winne, P. H. (2017). Leveraging big data to help each learner and accelerate learning science. *Teachers College Record*, 119(3), 1–24.
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2016). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52(1), 17–37. <https://doi.org/10.1080/00461520.2016.1207538>

Yardley, L. (2000). Dilemmas in qualitative health research. *Psychology & Health*, 15(2), 215–228. <https://doi.org/10.1080/08870440008400302>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Bingyi Han¹  · Sadia Nawaz²  · George Buchanan³  · Dana McKay³ 

✉ Bingyi Han
bingyi.han@unimelb.edu.au

¹ School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

² Department of Human Centred Computing, Faculty of IT, Melbourne Metrics, Faculty of Education, Monash University, University of Melbourne, Melbourne, Australia

³ School of Computing Technologies, RMIT University, Melbourne, Australia

4.6 IMPACTS ON CLASSROOM DYNAMICS: INTERACTIONS AMONG STUDENTS, TEACHERS, AND SYSTEMS

Full citation: Bingyi Han, Simon Coghlan, George Buchanan, and Dana McKay. 2025. Who is Helping Whom? Student Concerns about AI-Teacher Collaboration in Higher Education Classrooms. *Proc. ACM Hum.-Comput. Interact.*, 9, 2, Article CSCW206 (April 2025), 32 pages, <https://doi.org/10.1145/3711104>.

The published paper begins on the next page.



Who is Helping Whom? Student Concerns about AI-Teacher Collaboration in Higher Education Classrooms

BINGYI HAN, The University of Melbourne, Australia

SIMON COGHLAN, The University of Melbourne, Australia

GEORGE BUCHANAN, The University of Melbourne, Australia

DANA MCKAY, RMIT University, Australia

AI's integration into education promises to equip teachers with data-driven insights and intervene in student learning. Despite the intended advancements, there is a lack of understanding of interactions and emerging dynamics in classrooms where various stakeholders including teachers, students, and AI, collaborate. This paper aims to understand how students perceive the implications of AI in Education (AIEd) in terms of classroom collaborative dynamics, especially AI used to observe students and notify teachers to provide targeted help. Using the story completion method, we analyzed narratives from 65 participants, highlighting three challenges: AI decontextualizing of the educational context; AI-teacher cooperation with bias concerns and power disparities; AI's impact on student behavior that further challenges AI's effectiveness. We argue that for effective and ethical AI-facilitated cooperative education, future AIEd design must factor in the situated nature of implementation. Designers must consider the broader nuances of the education context, impacts on multiple stakeholders, dynamics involving these stakeholders, and the interplay among potential consequences for AI systems and stakeholders. It is crucial to understand the values in the situated context, the capacity and limitations of both AI and human for effective cooperation, and any implications to the relevant ecosystem.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; • **Applied computing** → **Education** → Computer assisted instruction

Additional Key Words and Phrases: AI in Education, Human-Centered AI, Ethics, Fairness, Bias

ACM Reference format:

Bingyi Han, Simon Coghlan, George Buchanan, and Dana McKay. 2025. Who is Helping Whom? Student Concerns about AI-Teacher Collaboration in Higher Education Classrooms. *Proc. ACM Hum.-Comput. Interact.*, 9, 2, Article CSCW206 (April 2025), 32 pages, <https://doi.org/10.1145/3711104>.

1 Introduction

Higher education institutions have been eagerly deploying AI-enabled classroom analysis and intervention systems (e.g., [6, 126]) to work with teachers and students in teaching and learning. By collecting, analyzing, and interpreting student learning and performance data [10], AI can continuously observe and assess students' in-class behavior [17, 73]. This analysis can aid educators or instructors with tailored intervention suggestions (e.g., in engineering classrooms

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ACM 2573-0142/2024/4 - CSCW206

© Copyright is held by the owner/author(s). Publication rights licensed to ACM.

<https://doi.org/10.1145/3711104>

[11] and nursing education [31][92]) and offer students real-time personalized feedback [8, 14, 51, 115], thereby enhancing teaching efficiency and learning outcomes. Despite this promise, there is growing apprehension about AI's impacts, such as its effects on privacy and the magnification of existing biases within education [5], which may alter the dynamics in educational settings. However, AI's influence on the cooperative environment and interactions among various stakeholders, including students, teachers, and the AI system itself are not holistically understood. The CSCW community, with its expertise in navigating complexity among different user groups, is well positioned to shed light on possible future classroom dynamics influenced by AI, to identify problems, and to inform better AI in Education (AIED) design and adoption.

Existing studies examining the effects of AIED primarily focus on ethics. These discussions mainly revolve around theoretical and governance-related aspects [119], addressing traditional computational ethical issues such as data use and privacy [15, 75, 132], algorithm bias [84, 96, 161], and algorithm fairness [159]. This relatively narrow ethical focus is observed in relation to AI adoption in general. In the education context, there is a lack of attention given to the interaction between students and AIED. Students are vulnerable individuals [150] and can bear the direct impact of AIED, both positive and negative. Although AIED is explicitly designed to foster positive changes in students' behavior [57], students' perceptions of these potential changes remain unexplored. The way individuals perceive technology can profoundly shape their responses to and engagement with it: a student who views AI as beneficial may interact more openly with it (e.g., [33, 164]), whereas a student who is skeptical of AI might withdraw or disengage from it. A lack of understanding of students' perceptions may undermine successful system adoption [33].

HCI researchers have long recognized that purely conceptual notions of people's interaction with technology might diverge from actual practices. As Suchman [130] and Preece [109] highlight, circumstances around users are ever-evolving and often unanticipated; however, accommodating those unforeseeable contingent consequences is necessary for our designed digital tools. Suchman further posited that the dynamic interaction between humans and their social context constantly reconstructs human practices [130]. Therefore, to fully comprehend the AIED's implications, we must consider the broader educational ecosystem, encompassing students, teachers, AI, and the educational context – interwoven with social factors that shape how they interact.

AIED is becoming increasingly complex and pervasive, aiming to revolutionize education [23]. While AIED's intended outcomes are relatively well-defined, many potential unintended consequences remain uncertain. Unlike conventional educational technologies such as laptops or learning platforms like Zoom, AIED introduces autonomous decision-making systems that pose distinct concerns. For instance, AI systems often rely on copious sensitive data for training models, making them vulnerable to privacy breaches through data leakage, sharing, and hacking [15, 75, 132]. Another key concern involves opaque biases in deep learning AI that are related to race, ethnicity, or gender, and which can unfairly affect student outcomes [18, 79, 159]. Once these harmful unintended consequences arise, they can be persistent and difficult to remediate. Privacy violations, for example, may cause ongoing anxiety for those affected, and biased AI outputs might lead to affected students falling behind in their studies and feeling stigmatized. Although these concerns are recognized to some extent, the full spectrum of potential consequences of AIED implementation remains largely unexplored.

Yet identifying and addressing negative issues only post-implementation can be unethical and ineffective, especially in sensitive educational settings. The nuances of the educational context

and the possible invasiveness of AIEd require a proactive strategy to pre-emptively identify, confront, and mitigate potential unintended consequences. Thus, rather than asking about students' current experiences or exposing them to experimental treatments that could cause both short and long-term harm the nature of which is unclear, we chose to delve further into identifying and preventing potential concerns before they manifest negatively. To this end, we employed a user-centered speculative HCI method with a future perspective – The Story Completion Method (SCM) – to provoke participants to respond to AIEd scenarios designed based on current proposals from the AIEd community. In story completion tasks, participants are invited to create narratives, which reflect their socially and culturally constructed knowledge and values [74, 155]. The Story Completion Method (SCM) has been effectively used to probe likely responses and evaluate ethical implications regarding emerging or future technology [42].

Our study aims to uncover students' responses to AIEd adoption in higher education classrooms. We pose the following research question: *How do students perceive and respond to AI-assisted analysis and intervention in classroom learning?* Using SCM, we designed story completion tasks with futuristic classroom scenarios. These story prompts are informed by and reflect the advancements and proposals documented in recent AIEd research, as detailed in Section 2 and Section 3. Participants co-created these stories by imagining themselves as studying in such classrooms. By analyzing 65 stories, our study reveals students' concerns regarding AI in the education context, AI cooperating with teachers, and AI interacting with students. We also identified the interactions among various potential impacts of adopting AIEd. We summarize our contributions as follows:

1. Although our study is situated in the context of higher education, students' concerns and ethical perceptions regarding AIEd adoption remain largely unexplored [57], whether in higher education or K-12 education. This gap underscores the value of our work across different educational levels, as it provokes reflection not only in higher education, but also fosters ethics discussions in educational settings overall. Further, higher education students are typically adults (albeit young adults) with more agency than K-12 students. Ethical issues that arise in higher education, where participation in many elements of pedagogy is optional are likely to be more acute with younger, more vulnerable students whose educational participation is often compulsory—we are just beginning to address the ethical challenges wrought by introducing AI into the classroom.
2. We uncover student responses to the newly formed AI-infused classroom learning landscape, especially their concerns and potential impacts on key stakeholders – AI systems, teachers, and students. Our findings shed light on ethical and sociotechnical challenges introduced by AI, to inform better design and adoption of AIEd.
3. We reveal potential interactions between AI interventions and student behavior that are not reported in previous literature, especially the *Situated Changing Circle*. AI monitoring potentially leads to student behavioral change, which in turn can disrupt the AI system's effectiveness. Our findings demonstrate the dynamic nature of AI in education, underscoring Suchman's insights on the situated nature of technology adoption.
4. Our work underscores the significance of viewing AI, teachers, students and the educational environment as an integrated entity, rather than isolating AI within its educational context. This interconnected perspective, highlighting the mutual influence of each element, encourages the CSCW and HCI community to develop holistic, context-sensitive AIEd approaches.

2 Related Work

In this section, we first introduce the current approach of AI-assisted classroom analysis and intervention techniques. We then review existing ethical discussions about adopting these AIED tools.

2.1 Integrating AI Analysis and Intervention in Classroom

The AIED community has made significant effort to design and develop AI applications for higher education institutions, including in four popular areas: assessment and evaluation, profiling and prediction, adaptive personalization systems, and intelligent tutoring systems (ITS) [160]. Many AIED studies and technological advancement address more than one function area. For example, AI can be used to detect students in-class behavior, assess and predict students learning, diagnose their strengths and weaknesses, and provide customized learning suggestions and teaching strategies. These insights feed into ITS and allow teachers to offer tailored instructions ([11][31][92]), enriching teaching and learning experiences [58].

Recent HCI studies have proposed involving teachers in designing AIED tools or intelligent systems (e.g., [44, 63]). For instance, methods like the “co-orchestration” model [61, 91, 158] aim to synergize human capabilities and AI strengths, utilizing classroom sensing technologies to inform teachers with data-driven insights in classroom dynamics [58, 77].

Classroom sensing technologies, such as Lumilo [60], Edusense [3], ClassID [106], employ ambient sensors, such as cameras [10, 107], alongside AI techniques like computer vision and machine learning, to detect student movement and behavior, analyze students’ engagement, participation, and interactions in educational environments [146, 149]. Specifically, AI-enabled facial and emotion recognition techniques (e.g., [113, 163]) can capture student behavioral data such as head orientation (nodding, shaking, and head poses) [3, 107], body gestures [87], gaze behavior (e.g., eye direction) [4, 64, 83] during lectures or in-class practices [30], as well as facial emotional expressions (boredom, confusion, or frustration) [17, 43]. These data can be fed into learner’s affective and physiological state models with indicators of learning performance, such as body motions as learning evidence [87] and arousal in a collaborative learning environment [157]. Derived from quantified indicators [160], AIED tools interpret students’ in-class behavioral performance [10], cognitive-affective status, and real-time emotional and attentional engagement [17, 64, 113].

Based on these data-driven calculations of statistical parameters, students can be categorized such as ‘disengaged’ or ‘struggling’ [17, 39, 102] and ‘low-performing’ or ‘at-risk’ [101]. These insights about student conditions, performance or drop-out prediction [40, 49, 156] can be delivered to intelligent tutors [113] and human teachers [11, 17, 68] with technology such as TeachActive [11]. This helps to heighten teachers’ awareness of students’ emotional and cognitive status [131] and assist them in better adapting to learners’ emotions [38] or learning progress [39], prompt their reflections on teaching strategies [11, 12, 43, 50], and provide proactive intervention to support students learning [60, 64, 153].

AIED research has been diligently working on enhancing system functionality [9] through investigating system data and indicators [2] to improve the interpretation of students’ interactions and learning patterns in AI-mediated learning environments. A recent review study on AI applications in higher education found that most empirical studies have a quantitative focus [160], and that the current focus often lacks a “pedagogical grounding” [142]. Socio-technical considerations remain understudied in AIED development [89, 160].

While qualitative empirical studies exist, they typically explore stakeholders' expectations and functional preferences for adopting AIEd [82, 108, 151], concentrating on technical attributes to achieve intended outcomes such as successful integration [65, 103, 108]. Often lacking is an examination of the potential or actual impacts on students [114], especially the unintended consequences and a broader range of ethical risks associated with AIEd adoption [23]. However, the complexities and nuances of educational context warrant urgent critical consideration [122], particularly given the sensitive student characteristics involved, such as gender, sexual identity, and race. This is especially important as students, being a vulnerable group, may be disproportionately affected by these oversights. In the next section, we will briefly review existing concerns about AIEd adoption.

2.2 Concerns about Integrating AI in Education

In AIEd, established concerns that are intrinsic to AI techniques dominate the ethical conversation, especially surrounding data use and its computational approaches related to issues [57] such as privacy [143], algorithmic bias and fairness [67, 123], and transparency [112], as evidenced by recent reviews (e.g., [54, 57, 99, 144]). Empirical studies that explore stakeholder perspectives further narrow this scope, with a primary focus on specific themes such as privacy (e.g., [15, 75, 81, 132]) and bias (e.g., [85, 123]). Little attention has been given to the socio-technical impacts on teachers and students stemming from their direct interaction with AI.

This limited (though important) ethical focus has been extended to AIEd governance frameworks (e.g., [110, 120, 128]), which primarily address data-related concerns from a compliance standpoint, seldom considering broader socio-technical implications. This trend persists even in the SHEILA (Supporting Higher Education to Integrate Learning Analytics) framework [141], arguably the most comprehensive learning analytics framework grounded in empirical research. While SHEILA aims to secure stakeholder buy-in and identifies behavioral changes that learning analytics systems should facilitate, it overlooks potential unintended modifications to the learning environment and their impacts on the student body. This gap is acknowledged by SHEILA's architects, with Tsai et al. [140] critiquing the limited research on the impact of technology adoption on the student experience and learning journey. Neglecting these unintended influences risks undermining student trust and acceptance of AIEd, as legal compliance does not necessarily ensure a positive student experience with AIEd systems in practice.

The ethical considerations in AIEd should extend beyond AI and data-centric issues [57] to the real-world context of students-AIEd interactions. While the pedagogical community has raised theoretical critiques about the potential pedagogical impacts of AI-enabled classroom monitoring and intervention, such as affecting learner independence and learning autonomy [134] and the datafication of student behavior and education [90], these arguments often have limited empirical foundation. In contrast, the HCI and CSCW community, with its history of prioritizing ethics in technology design [47] and its emphasis on user-centered approaches [13, 24], advocates for a holistic and empirical understanding of user experience amongst stakeholders—especially, in this case, the students themselves, to truly grasp the nuances of their learning experiences when interacting with AIEd in the learning environment.

Recent HCI and CSCW research has shed light on the unintended socio-technical consequences of emerging technology in education. For instance, a study highlighted that data-driven classroom surveillance [84] can oversimplify children's complex psycho-social behaviors, reducing them to mere data points. Such methodologies risk reinforcing and amplifying existing

biases and inequalities in education, particularly those rooted in race, gender, and culture. That study focussed on repercussions from teachers' perspectives. However, it is crucial to recognize that students and teachers might perceive ethical values and impacts differently [158], especially given the inherent power differential between them (e.g., [105]). To date, AIED discourse has largely sidestepped students' views — a worrisome oversight given students will directly experience AIED's impacts. Ignoring these concerns could impede ethical and pedagogically useful AIED adoption and potentially harm the individuals it seeks to benefit.

We have noted that both HCI and CSCW underscore the complex situated nature of technology adoption (e.g., [45, 66]) and the importance of understanding its impacts not just on functionality but also on the broader social and emotional facets of people's lives (e.g., [70]). Yet it is well known that actual user interactions with technology can diverge from what designers anticipate [109, 130]. Indeed, the value promoted by the system might even contradict and undermine design intentions [46]. This is particularly concerning in the educational context, as any potential adverse effects of unintended impacts are significant and possibly harmful. However, as we have stressed, the intricate dynamics between students and the evolving AIED environment are underexplored.

Our research seeks to bridge this gap in understanding by focusing on the educational context from students' perspectives. While academia and industry are fervently pursuing advancements in AIED (as described in Sec 2.1), widespread adoption is still in its infancy, with only a handful of learners experiencing its full effects [97]. But given the complexity and potential invasiveness of AIED, it is crucial to preemptively identify and address its pitfalls before harm is done. To proactively address these concerns, we employ the user-centered speculative Story Completion Method (SCM), which allows us to delve into "what if" scenarios, offering insights that can guide the adoption of more ethically sound AIED designs.

3 Research Method

The overarching goal of this study is to explore students' concerns about AIED in higher education classrooms. We explore how students perceive classroom collaborative dynamics involving AI, teachers and themselves. Because there has been little consideration of broader sociotechnical implications in current ethical frameworks, we chose not to adopt any pre-existing ethical frameworks to guide our study. Instead, we suggest that our findings can enrich and supplement existing ethical frameworks. The research was approved by The University of Melbourne Human Research Ethics Committee (Ethics ID Number: 24525).

3.1 Why the Story Completion Method?

Speculative methods, such as story completion [42], and pre-mortems/post-mortems [145], are widely used to critically evaluate the potential ethical and moral implications of emerging technologies before their widespread implementation [42]. This approach is particularly relevant and ethical in educational contexts, where real-world experimentation might expose students to unintended negative consequences. By using speculative methods, we can proactively explore potential issues in sensitive settings, ensuring ethical considerations are thoroughly examined.

The Story Completion Method (SCM) incorporates elements of speculative design [42, 154] and research fiction [22, 138], inviting participants to create narratives based on provided Story Completion Tasks (SCT). Those story openings help with introducing and constraining the context of our research interest for participants [35]. Stories created by participants present their socially constructed knowledge and values [74, 155], indicating common assumptions about or

visions of AIED's impacts on students and their learning. Recent HCI work has shown the effectiveness of SCM in exploring other sensitive topics (e.g., VR pornography [155] and sex robots [139]), as the third-person storytelling approach with indirectness [35] encourages a more comfortable disclosure of concerns. These qualities make it an excellent candidate for revealing ethical challenges.

We acknowledge the important criticism that SCM cannot inform us about actual psychological states [74]. Hence, we interpret our results as potential outcomes ("might be") rather than definitive predictions ("will be"), aiming to uncover students' socially constructed visions of AI intervention in education.

3.2 Data Collection

We designed three Story Completion Tasks (SCT), as seen in the Appendix, presenting possible future scenarios of AI-enabled intervention in university classrooms. These scenarios were based on proposed or existing AIED technologies (Table 1) and were also informed by recent literature-based design fiction studies exploring the impacts of AI adoption in higher education (e.g., [37]). Unlike previous HCI studies exploring future technology design (e.g. [139, 155]) that used brief story openings consisting of only one or two sentences, our SCTs offer more immersive scenarios. Given our study's focus on the situated aspects of AIED adoption, we provided more descriptions to help participants understand the context. This approach aligns with a recent HCI study [137] that extended SCM with interactive storytelling gamification, further enhancing immersion through SCT.

Scenarios 1 and 2 (Appendix) depict the observation on students' in-class behaviors using AI tools such as Facial Recognition and Emotion Recognition, to help teacher understand students' emotional readiness and cognitive needs for learning so that to provide targeted support. Specifically, Scenario 1 focused on individual observation during lectures, while Scenario 2 emphasized observation during group activities to capture individuals' engagement in more extensive student interactions, so that teachers can better understand the learning progress of the class and adjust teaching strategies. In Scenario 3, the data captured by these systems, including students' engagements and affective states, are utilized for AIED functions. This scenario described the intended benefits to enhance students' learning with customized learning services, encompassing real-time analysis based on individual education profiles, personalized recommendations on learning pathways and resources, and automatic evaluation, prediction, and interventions.

We acknowledge that our use of story prompts could influence participants' responses. Terms such as "monitoring" might evoke certain narratives. To mitigate this limitation, we clarified in the introductory materials that the purpose of this research was not to cast AIED technology in a negative light. Rather, we seek to identify perspectives and concerns that can inform the ethical design and improvement of AIED systems. Our scenarios emphasized the objective of enhancing learning as we tried to maintain a neutral tone to avoid introducing any negative biases.

In addition, we intentionally described the technology artifacts and AIED functions with a degree of ambiguity, providing only the context, to encourage imagination and elicit a broader range of responses [28]. At the end of each scenario, we guided participants by asking what happens next and how the character would think and behave. Participants were asked to spend 5 to 10 minutes completing each story, to ensure the story quality.

Table 1. Story Prompts Justification – Proposed or Existing AIED Artifacts

Scenarios (S)	Functions & Techniques	Code in Story	Examples
S1&2	Classroom Facial & Emotional Monitoring	[CEM]	[113]
In-Class Observation,	Eye Direction / Gaze Tracking during Lecture	[EGT]	[4, 64, 83]
Real-Time Evaluation	Awareness & Emotional Engagement Analysis	[AEA]	[43, 59, 60]
	Real-Time Classroom Behaviour Analysis	[RBA]	[3, 17, 83, 107]
S3 Intervention,	Automatic Evaluation and Timely Intervention	[ATI]	[9, 17]
Prediction,	Prediction and Academic Early Warning	[PAW]	[1, 49, 100, 156]
Customization	Adaptive Personalized Recommendation	[APR]	[9, 48]

^aFull scenarios are available in the appendix. The functions and techniques in Table 1 are based on proposed or existing AIED technologies and recent literature. To facilitate referencing, a code has been assigned to each function, marked in the corresponding stories.

We targeted participants recently or currently experiencing higher education (undergraduate, postgraduate, PhDs, and recent graduates). To explore potential cultural influences on the perception of ethical concepts in AIED, we recruited participants from two different demographics (China and Australia). However, we did not ultimately observe substantive differences between them, and this allows us to conclude that our findings span cultures. Our recruitment did not filter participants based on their majors, level of AIED experience, or AI literacy, as our focus was on broader concerns encompassing all types of students, considering the diverse backgrounds of students in the actual university contexts.

We collected 71 stories anonymously via the Qualtrics online survey platform. Our survey presented story stems, and collected participant's demographic information including gender, age, and teaching experience, if any. Participants with teaching experience were denoted with a "T" (e.g., P1-T). Six incomplete stories were excluded from our analysis. Of the remaining 65 participants, 54% were male and 46% were female; none identified as non-binary.

3.3 Data Analysis

We employed reflexive Thematic Analysis (TA) to analyze the collected stories [26], as it allows us to form an understanding of the overarching shared viewpoints of our participants. The first author led the analysis following Braun and Clarke's [25] guidance. After carefully reviewed all the stories, we coded the stories and organized the codes inductively. These codes were either semantic, which captured the explicit ideas expressed by participants, or latent, which reflected the researcher's interpretations of the underlying meanings present within the data [133]. Then, we sorted the codes into initial themes and sketched their relationships on mind maps [25]. The research team reviewed the mind maps and refined the category structure through multiple rounds. This iterative process adheres to the principles of reflexive TA [27], which advocates for continuous iterative reflexivity and ensures that the analysis and interpretations evolve progressively and recursively [26, 135]. We finally answer our research question, *How do students perceive and respond to AI-assisted intervention in classroom learning*, through three main categories including AI in context, AI and teacher, AI and students themselves. We present those three categories in the following sub-sections, with identified themes highlighted in Bold.

4 Findings

This paper is part of an overarching study that aims to understand students' perceptions of AI interventions in higher education classrooms. The overall findings highlight students' concerns and expectations across various aspects of AIED within its adoption context: we have presented the individual dimension in [55], and the pedagogical dimension in [56]. In this paper, we focus

on three major concerns regarding the dynamics among AI in the context, teachers, and students: the first is the potential for AIEd's quantification approach to oversimplify intricate educational context and thereby disrupt fairness (Sec 4.1). The second is the collaboration between human teachers and AI, where we observe situated bias concerns and power disparities (Sec 4.2). The third is the fact that students may change their behavior as interactive strategies with AIEd, while these situated responses can intentionally or unintentionally disrupt AIEd's efficiency and effectiveness. (Sec 4.3). Finally, we introduce students' positive perceptions regarding educational processes and approaches (Sec 4.4).

4.1 Context: Disruptions to Accuracy and Fairness where AI is Context-Unaware

Participants' most frequently expressed concern was whether AI can fully represent the actual education context. They felt the **quantification approach** of AI-assisted evaluation **oversimplifies human behavior**. In particular, the nature of human beings and human experience, especially feelings and affective status, are "too delicate to be understood by quantitative indicators" (P36). "The polysemy inherent in human facial expressions adds complexity to their interpretation. Moreover, the connection between these expressions and an individual's inner thoughts remains ambiguous. So, the detection of such nuances by AI systems may be shallow and prone to inaccuracy" (P3). Hence, "the data may not be fully representative of the real-life situation" (P52-T). Here are some examples of **data misrepresentation**:

"... the system reported that he looked bored, but he was just really tired after working a late shift the night before." (P65-T)

"The real-time AI analysis detected his lack of confidence for today's lecture. Actually, it was caused by his frustration with his crush who sits nearby. Each glance in her direction intensified his nervousness..." (P5)

The story from P15 above also questions the correlation between the collected attributes of facial expressions and in-class behavior information with actual learning progress. Similar concerns frequently pointed to the quantitative representation's **decontextualizing the learning process and over-digitizing education**. Participants argue that learning outcomes are influenced by a multitude of factors, including individuals' learning motivations and various elements within the lecture environment such as "...the specific content of the lectures, the teaching method of the lecturers and the interactions between students" (P64). The system evaluation, in their view, oversimplifies the complexity of learning situations by solely focusing on surface-level outcomes while neglecting other essential pedagogical factors that shape students' experiences and achievements. An example below emphasizes the limitation of relying solely on quantitative data to comprehend the dynamics and intricacies of individual learning trajectories:

"The learning curves of most students follow a non-linear pattern, but more like an explosive growth during finals... AI analysis won't know that." (P33)

Another commonly raised concern pertains to the tendency of rigid **one-standard-based** AI systems to **over-generalize educational situations**. These concerns are closely linked to the presence of numerous complex **exceptions** within real-world educational settings, which the mechanistic fact-based nature of AI may struggle to handle in a fair manner. Handling student-related matters in educational settings requires flexible responses. AI systems that rely solely on

standardized metrics might neglect the inherent intricacies and diversity within education, and **unfair decision-making** may ensue:

Exceptional Case: “The student was sick, but he still made an effort to attend class every day and aced his grades. He didn’t participate in activities that much and seemed uncomfortable in class, which should’ve been understandable given the circumstances. But guess what? The system evaluation assigned him a lowest on-class performance score. Talk about unfair!” (P21-T)

Educational Diversity: “...bright students can appear disengaged because they are not challenged enough - which is also a reason for poor attendance... students may have covered some of his subject materials in previous study or may have transferred from another university where this material was covered earlier in the course. These factors all cause variations in student performance.” (P59-T)

The diversity within the student population is one obvious example that shows the complexity of education. The stories below described a particular advantaged personality of students when evaluating in-class performance by quantifying interactions and engagements. This mechanism potentially favor extraverts or “expressionists” against introverts, thereby disrupting **fairness**:

“This student is knowledgeable... He is one of the quieter ones in the class, which is likely why the evaluation system has given him a lower score... this assessment is unfair.” (P56-T)

“The system does not consider students' personality and traits that can affect students' patterns of participation in group discussion. While outgoing students who enjoy engaging in discussions may not be affected, those who feel uncomfortable expressing themselves within a group may face difficulties. The system becomes a burden for them, hindering their preferred mode of participation.” (P64)

In traditional classrooms, group work may already pose challenges for some students. The examples above demonstrate that AI-based performance evaluation, with its rigid algorithm, tends to be one-size-fits-all that can intensify these challenges for some students in group settings. Thereby, it potentially creates a less inclusive learning environment for diverse personalities.

Achieving fairness, as perceived by participants, necessitates the consideration of all possible situations for students, allowing for compassionate exemption and explanation opportunities. Participants' discontent with the inflexible nature of the AI mechanism and its oversimplification of educational scenarios during assessment often stems from their perception of AI's lack of compassion in decision-making. Particularly, the exceptional cases mentioned earlier were believed to be better handled by the human touch of compassion and subtle empathy. Those attributes are seen as essential in the educational context where AI can fall short, due to its unawareness of the intricacies of education. Interestingly, when participants perceive the inflexible AI evaluation as unfair, they tend to more frequently compare AI and human teachers. Their point revolved around the notion that learning evaluation should regard human factors where teachers' human experience and deliberation are vital to ensure fairness. This point features in the next category.

4.2 Teachers: AI Cooperation, Agency, and Power

In this section, we initially demonstrate how participants envisage teacher-AI cooperation and the respective roles of student and teacher in the classroom. We then introduce participants' concerns about, first, bias concerns and, second, power disparities.

4.2.1 Human-AI Cooperation and Roles. Participants advocated for integrating human judgment to counterbalance AI's inherent limitation: "AI analysis cannot be seen as the sole evidence or method for evaluating students, as it is unable to understand students' varying situations and circumstances" (P27-T). They see the role of human intervention in AI-mediated learning evaluation, as a critical way to add contextual insight to extend AI's analytical capability. Teachers, adept at discerning intricate classroom dynamics that impact academic outcomes, draw from "their observation of subtle student behaviors and affective states" (P42-T), as well as their understanding of cultural nuances and individual student backgrounds. This depth of teachers' situational awareness empowers them to formulate flexible strategies. The role of AI, therefore, should be positioned as a *supportive complement* to human teachers, furnishing valuable analytical insights without encroaching upon the authority inherently tied to the human facet of education.

Participants often describe a preferred partnership between humans and AI where **human teachers hold agency** while **AI serves as a supportive auxiliary tool**. Stories provide various examples of how human teachers exercise their agency in conjunction with AIED such as: Customizing the taught curriculum according to AI's analysis, such as adding scientific references based on class progress and student needs; adapting the instructional approach, by drawing on real-time analysis, automated feedback, and suggestions provided by AI; and providing targeted learning support by leveraging AI-generated cues to identify and address students' struggles.

Narratives commonly emphasized the need for **teachers' decision-making authority**. Teachers are expected to actively control the decisions made by AI systems as the analytical insights provided by AI always require **human oversight and interpretation**. One typical comment was:

"I'm worried if the AI system is given any more weight than what the teacher says. The tool should just only be treated as a reference to assist teaching but should never be given the ultimate say in anything or be the deciding factor in any major decisions (kicking someone out of a course, giving the final grade, marking written pieces of work)." (P60-T)

Especially when conflicting AI analyses occur, teachers are expected to overrule AI. Here are two examples where teachers would aim to ensure evaluation fairness:

"The professor becomes intrigued by the puzzling difference between the student's mediocre in-class performance, as rated by the AI system, and their impressive exam score. Driven by a strong sense of fairness, the professor carefully investigates how the system evaluates student work. Seeking to understand the truth, the professor kindly invites the student to share their perspective, allowing them to explain their performance. Recognizing their position of authority, the professor considers the option of reevaluating the student's work if it seems necessary." (P45-T)

Stories also mention **teachers' role changing** after introducing AI in education: transforming from observers and evaluators into educational strategists and designers, which requires more enhanced professional pedagogical skills (P13-T). Meanwhile, concerns were raised regarding the potential impact of role transformation on the personalization and individualization of the student-teacher relationship:

“The experience becomes less personal...Will teachers even bother getting to know him if the machine is taking the attendance and marking the class participation?” (P56-T)

“The professor tries to remember back over the semester about this student. Before the system was in place he used to take note of students who didn't seem to be engaging, and try to get a feel for what they were thinking. But he has become so used to relying on the system to find these students and offer them a guided path that he no longer has that close contact with his students...” (P65-T)

However, humans might not always ideally lead the Human-AI cooperation. Although the system evaluation might be “ruthless for lacking empathy ” (P40-T), some respondents perceived AI as objective and impartial, which is “perfect for playing a role as an **unbiased third party**” (P60-T). They view its fact-based approach as a means to mitigate human biases [125], ensuring fairness in the evaluation process, such as preventing teachers from playing favorites and instead promoting an equitable distribution of attention and learning materials (P67-T). Extending beyond evaluating students, another participant highlights the system's potential to provide a more comprehensive evaluation of teaching compared to student satisfaction surveys, which often reflect only extreme opinions (P65-T). Thus, AI evaluation is considered a valuable tool for a well-rounded assessment for both learning and teaching.

4.2.2 Bias Concerns. Nonetheless, AI analysis can be inherently biased [18], as our participants were well aware. Three participants (P30, P59-T, P65-T) express concern regarding unintended biases in AI systems. For example:

“All automatic systems are only as good as the concepts and algorithms programmed into them... the system may be biased by the ideologies and priorities of the designers. This represents a narrowing of the range of strategies and approaches to which students are exposed. Overall, it is likely that more students will not have their needs addressed.” (P59-T)

Participants recognize the potential of AI-provided evidence to address human bias (e.g. teachers' favorites) that may exist in conventional education. Yet, they also express concerns about the **unintended bias** that AI-generated evidence may introduce to teachers' perceptions. Participants highlight the potential **psychological impact on teachers** who may form preconceived opinions from AI analysis before themselves fully assessing students' abilities, akin to a “verdict before trial” (P30). This apprehension is centered around the negative consequences of predictive low grades or ranking, as participants fear that teachers may lose confidence in them (P29) and treat them differently (P43). These examples underscore the potential influence of AI predictions on teachers' biases and perceptions, emphasizing the need to carefully consider the unintended consequences of predictive models.

The empowerment of AI and the use of data then become a shared concern, given those potential implications and unintended consequences. Participants emphasize the unfairness of using past data to predict or prejudge students' future achievements, particularly considering the intricate and ever-evolving nature of the educational context. P30 argues that “even if the predictions are 100% accurate, they remain predictions rather than definitive outcomes.” P4 echoes that thought: “Data is power”. Further, the reliance of AI systems on historical data can perpetuate existing biases and inequalities present in educational systems. Using past performance as a basis for predictions carries the risk of reinforcing stereotypes and limiting opportunities for students based on their previous outcomes.

4.2.3 Power Disparities. Stories also reveal a **power dilemma**: The realization that AI is fallible prompted participants to call for caution and deliberation in leveraging data-driven tools and allocating the power of AI and human teachers. They grapple with questions such as “how much authoritative power should human teachers allocate to AI prediction and analysis when evaluating students?” (P15). We observed earlier that participants place greater trust in human teachers for making final decisions, as they perceive fairness is achieved through discerning special circumstances and possessing contextual understanding and discretion. Hence, teachers are expected to use their authority. However, teachers can be influenced by AI insights, creating a complex dynamic where the power appears to rest in human hands, but can be shaped by the implicit yet potent influence of AIEd; for example, when AIEd directs teachers to a student who is identified as needing more support in a class. Complicating matters further, AI insights themselves may be biased by past data and human designers.

Another example vividly portrays this power dynamic, where a teacher employs AI insights to support (and potentially rationalize) an already-made decision:

“The Prof doesn't feel the need to increase the scores... This belief is reinforced by the “empirical” data that the Evaluation System provided which he can use as evidence for his decision... He then provided the student with a few suggestions according to the AI-generated report... Ultimately Chris feels more justified in his decision to deny a score increase complaint as it is not solely based on his judgment, but an outside, unbiased, 3rd party.” (P60-T)

4.3 Students: Behavior Changes Challenge AI Effectiveness

All AI systems rely on data sets that are representative of the interactions they wish to model, yet adopting AI in classrooms will directly impact students' behavior due to their interaction with AIEd. We have demonstrated students' potential behavior changes in detail in [55]. This disruption of behavior means the AI may be based on a model that does not accurately reflect changed circumstances in the classroom, thus reducing the efficacy of the model. This section is not going to demonstrate all possible impacts on individual students, but to reveal how students' potential reactions might further challenge AIEd effectiveness.

The cooperation of AI and teachers in the classroom creates a learning environment perceived by our participants as having ubiquitous monitoring. This perception instills fear in students that their poor performance will be recorded and reported, leading them to believe more unnecessary effort is needed for learning. To meet the perceived stringent requirements for learning and their wish to gain good performance scores, students may adopt interactive strategies and behavior changes, both intentional and unintentional. For example: pretending, acting, and pandering to the surveillance, avoiding making mistakes, being extra cautious of their expression, and forcing active engagement and participation. Detailed impacts on individual dimensions have been presented in [55]. We shall now demonstrate its challenges for AIEd effectiveness.

Most directly, commonly expressed **unintentional unauthentic behaviors**, **intentional performativity**, and **gaming system behaviors** will all influence the quality of data. However, there are issues that contaminated data will be used to train AI; those altered behaviors are result from AI. These **distorted data inputs** of identification and recognition will further challenge the **technical effectiveness** of the following data-driven functions such as evaluation, customization and intervention. Student participants acknowledged the problems:

“He pretends that he really understands the newly delivered knowledge. He shows a confident face (Pretending and Acting) ... He will work harder at home (Privacy-Seeking) to figure out the hard part rather than allowing the system to detect his struggle (Avoid Making Mistakes) ... All the data are generated based on the fact that all students hide their true selves in front of the camera. Those data are pointless.” (P52) “Students’ reactions were real in the conventional class. Now everyone deliberately presents their perfect self rather than being what they really are (Performativity). The analysis won’t reflect the true situation.” (P21-T)

“I finally managed to achieve high scores by Gaming the System... The system knows exactly where my study weakness lies, I believe it will reduce the workload for teachers and enhance my learning efficiency by targeting my struggles. Oh wait! My data and information profile are no longer “accurate” starting from this semester, what should I do...” (P18)

One participant pointed out that measures can become poor indicators when they are targeted, referencing Goodhart’s Law. It questions the situation where students likely game the system to meet AI expectations and the evaluation will become ineffective:

“However, as Charlie is a clever university student, he quickly figures out that it is not hard to pretend to be engaged without actually engaging. One can pretend to be reading materials without attempting to understand and one can say many things without actually meaning anything. Not only does he do this, but he also finds out that many other people are doing the same in order to trick the system. A classical phenomenon as described by the Goodhart’s Law – ‘When a measure becomes a target, it ceases to be a good measure.’ Sometimes people have to remember that the system of human behaviours is not invariant with outside observers. The very motion of trying to evaluate people change people. It is hard to think and say how exactly can the AI evaluate without interfering.” (P39-T)

Even **hostile behaviors** are expected, with intentionally rebellious strategies to subvert the system’s surveillance. For instance, quite a few stories described students’ intention to disrupt the accuracy of the system analysis, by deliberately presenting flawed data and “*contaminating the data model*” (P30). These behaviors target **attacking the essential value of AIED systems** in a covert indirect way. There are also overt direct aggressions such as “...he walked to the front of the classroom and unplugged the power source of the monitoring system with anger” (P2), or even retaliating by attacking the database. Further, stories also describe a mutinous or collective resistance response e.g., “rally classmates to make trouble against the system” (P38).

In addition to disrupting the system outcome, students may also seek to **benefit themselves** by “manipulating the system instead of being manipulated” (P38):

“The system can benefit him if he uses it well: this professor only read slides in his lecture... He should see those recorded bored faces and improve his teaching. He calls his classmates to exaggerate their facial expressions of boredom...” (P37)

“He had a strong dislike for algebra. With the system capable of predicting a student’s failure, he saw an opportunity to use it to his advantage. If teachers and parents lose confidence in him, he could easily get away with dropping algebra. He began to behave poorly and felt satisfied with his plan.” (P48)

“If AI monitoring tracks my eyeballs, I will deliberately prolong my gaze stare on the illustrations. The analysis might suggest improvements in textbooks by incorporating more pictures...” (P29)

The phenomenon of students engaging in strategic manipulation of the analysis also gives rise to a deliberate **restructuring of student-teacher communication dynamics** with a **new medium**. With the awareness of being under constant surveillance, students become more cautious in their behaviors and expressions, leading to a decrease in direct interaction with both their peers and teachers, particularly when they encounter difficulties in their learning process. Numerous narratives depict students’ desire for the system to assist students, particularly introverted individuals who perceive themselves as benefiting the most from this approach. Notably, some stories portray students intentionally gaming the system by amplifying their weaknesses, showing or even feigning struggles, anticipating that the system will prompt teachers to intervene, thereby capturing their attention and securing much-needed assistance. Moreover, once students **trust the efficacy** of the system, they may forego seeking help directly from teachers, relying instead on the system to allocate teacher attention to address their academic needs.

“This is too challenging for me, but it seems to be easier for my classmates. Perhaps the system can observe this and inform the teacher.” (P40)

“He knows that the system might help the teacher assist him so he might play up his weaknesses to make the AI believe that he needs additional help...Because he is aware of the system tracking his progress, he feels less need to seek direct help from the teacher ...the academic profile will inform the teacher either way.” (P60-T)

However, participants frequently voice a **diminished trust** in the accuracy and effectiveness of evaluation across the wider education system. They are skeptical due to their awareness that the system’s function is based on “...fabricated data resulting from students pretending and acting” (P45-T); “those inconsistent datasets cannot truly represent the actual learning situation, rendering the analysis untrustworthy” (P52). This distrust extends to the educational system itself, which is perceived to rely on these flawed AI evaluations. For instance, “... the education bases student criticism on quantified method with low fault tolerance, particularly when the system itself might be inaccurate... I will quit the school like this.” (P3)

4.4 Positive Perceptions: Enhancing Pedagogical Processes

In addition to the positive aspects mentioned in previous sections - such as being perceived as introvert-friendly by providing active support and directing teachers to assist, as an unbiased third party to provide objective evaluations, and creating a new communication medium between students and teachers - participants frequently noted the positive reconstruction on educational process and approaches.

For instance, students foresee AI intervention targeting on individual weaknesses, thereby improving learning effectiveness. Students perceive that AIEd system can provide “undivided, equal attention to each student” (P31), complementing teachers who have “limited time and energy per student... This allows teachers to utilize their energy more effectively” (P36). Additionally, students expect AIEd to assist both teachers and students by providing analytical insights to inform education, thereby increasing the effectiveness of educational process and approach.

4.5 Summary

Our findings revealed students' concern that AI's quantification and data representation approach may overlook dynamic and evolving educational contexts to the detriment of teaching and learning (as seen in examples presented in Sec 4.1). This simplifying of learning situations potentially caused by AI evaluating students can also lead to unfairness. Students expressed an expectation for human teachers to use their power and authority to overrule AI, to mitigate oversimplification, and provide contextual knowledge for ensuring fairness (Sec 4.2). We also observed the potential underlying issues of bias and power disparity for AI-human cooperation. However, while students expect teachers to collaborate with AI to ensure fairness, this cooperation in the classroom can lead students to perceive something like a surveillance panopticon. Narratives frequently depicted student behavior changes aimed at either appeasing or resisting AI-powered surveillance. This effect may not only further challenge technical functionalities but also diminish students' trust in the AIEd (Sec 4.3). Students' concerns about the three aspects of AI-mediated education – AI in the educational context, AI and teacher, AI and student – are often related to ethical issues such as fairness, bias, and power. Nevertheless, students also perceive AIEd can positively enhance pedagogical processes (Sec 4.4). In the next section, we will discuss these perceptions, issues, and concerns.

5 Discussion

A recent empirical HCI study on AI in future cities [98] highlighted four characteristics of urban intelligence that involve both humans and AI: *Contextual*, emphasizing the integration of knowledge rooted in human understanding of local contexts; *Conscious*, advocating for grounding in real-world experiences and interactions over mere quantification and data reliance; *Collaborative*, highlighting the synergy between tools and information, and between humans and AI; and *Controlled*, emphasizing the importance of human agency in governing AI interactions. As we shall further see below, these characteristics align with our findings on students' concerns about AI's cooperative role in education. Beyond that, our study probes specific challenges with detailed examples. For instance, the quantification approach might risk decontextualizing learning situations (Sec 4.1), and there is an anticipated blending of AI capabilities with human-derived insights, with teachers expected to exert control and make overriding decisions when collaborating with AI tools (Sec 4.2). We further revealed how students' potential behavior change might further challenge AIEd efficacy (Sec 4.3). In this section, we further discuss our findings about the possible effects and challenges of such AIEd.

5.1 Understanding the Context of AI Adoption

Our findings reveal student concerns about AI decontextualizing educational situations. To discuss these concerns, in Section 5.1.1, we examine the importance of understanding human values to help AI grasp the situated context, ensuring that the delivered outcomes align with user expectations. In Section 5.1.2, we discuss the challenges associated with contextualizing AI in education.

5.1.1 Understanding Situated Context and Human Values. HCI and CSCW research has long emphasized the importance of understanding the situated context of technology use [45] to ensure that its adoption aligns with the social and cultural milieu (e.g., [65, 117, 118]), particularly in high-stakes and sensitive areas [21]. Kaye's [70] exploration of low bandwidth devices for couples in long-distance relationships to communicate intimacy, for example, illustrates how simple devices can foster users' rich interpretations of simple interaction behavior on a single-bit

medium of communication, since it is situated in emotionally and socially rich pre-existing relationships.

Such studies remind us to delve deeper into the human values we aim to support, extending far beyond the mere productivity and efficiency that technology could bring. The HCI community has also advocated for solutions that transcend narrow, data-driven approaches (e.g., reflexive AI systems [98]). For instance, the “Whereabouts Clock” [29], which displays family member locations, not only supports family coordination but also contributes to the emotional aspects of family life (though it might also create other problems of its own (e.g., [94])). It provides a compelling example of how technology can enrich lives beyond functional utility. Thus, technology design and adoption should be aware of and responsive to the social and cultural contexts.

Extending these arguments to education, teaching and learning activities are recognized as processes steeped in social, cultural, and emotional capital [152]. Introducing AI into this nuanced educational context goes beyond helping with mere task completion, especially integrating AIED which relies on quantitative indicators to provide insights on evaluation, teaching strategies, and learning recommendations. Our data reveal the nuanced ways educational experiences might be influenced (e.g., students' performativity) or enriched (e.g., the newly created communication channel between students and teachers). However, the challenge remains in measuring these changes and determining their value within the educational context.

Our findings highlight student concerns that AI itself may oversimplify education and sometimes represent over-digitalization. These arguments from the student perspective in our study align with teacher concerns in a recent HCI study [85] that data-driven classroom surveillance measures codify and simplify the nuanced psycho-social factors that drive students' behavior and performance. Similarly, the education community provides theoretical criticism that behavioral evaluation systems can transform students' learning behavior and its complex social environment into statistical information - a form of pedagogical reductionism [90]. This approach can result in the underrepresentation or inaccurate portrayal of the social context, masking the ambiguity of actual social situations and limiting the interpretation of complexities [121]. Consequently, it decontextualizes [120] and oversimplifies [7] the actual learning process, governing students through classifications and comparisons based on numerical data. However, the potential behavior change observed in our stories may sometimes derail AIED from its intended objectives.

Such unanticipated consequences (e.g., Suchman [130] and Preece [109]) and value disparities are not new to HCI. An early study on designing education games [46] revealed a discrepancy between the actual and intended values promoted by the designed game, where the design reward structure inadvertently promoted a competitive style of caregiving, rather than the proposed value of cooperation and collaboration. Similarly, smart home technology has been found to inadvertently facilitate control and abuse to intimate partners [94], and electricity information sharing systems can lead to unexpected privacy breaches for household activities [127].

These examples of unintended negative consequences of well-intentioned technologies highlighted the need for extra caution in the educational contexts due to its complexity and the vulnerability of students. Any contradictions between the actual values and intended ones could potentially harm students. Therefore, we must consider the roles and implications of social and emotional aspects of AI intervention for students, teachers, and the educational environment. Understanding potential changes and their impacts can guide us to optimize technology use, align

AIED with students' values and educational practices, and develop more effective and contextually appropriate design solutions.

5.1.2 Better contextualizing AIED to Promote Accuracy and Fairness. A recent HCI study on a student-school assigning algorithm [112] revealed a clash between algorithm modeling assumptions and real-world situations, as the algorithm overlooks the intricacies of existing socioeconomic inequalities. This misalignment can compromise stakeholder values and hinder fair allocation. Similarly, our participants' concerns provide vivid examples of how the educational landscape is littered with complex, multifaceted situations where a simple quantification and data-driven approach could lead to students being unfairly treated. Therefore, a more holistic and context-aware approach to AIED is essential to ensure the technology does not inadvertently exacerbate existing inequalities or misunderstandings.

Contextualizing AI requires a more comprehensive student data profile [72] including contextual factors [116] such as student background, learning environment, psychological conditions, and individual differences, which can lead to more effective learning experiences. However, this increased contextualization, achieved through extensive data gathering and analysis, could introduce greater intrusions into the educational process and students' lives, raising issues of surveillance, privacy, and student autonomy [134]. This 'privacy paradox' [143] in education – the inherent tension between full data utilization and privacy – calls for a delicate balance between acceptable data utility and ethical use.

In addition to the need for more nuanced data, the contextual understanding of AIED should be supported by pedagogical evidence and social theory. AIED designers should raise awareness of the social dynamics within which the collected data sits, providing social and emotional contexts for data use [85] and making better sense of the digital data [121]. Our findings emphasize student expectations that human involvement will complement AI's limitations in contextual understanding. This resonates with other work on AI-human collaboration and HCI research [84], which stresses the vital roles of teachers in undertaking data work. Such work recontextualizes data to mediate the data-driven educational technology reducing complex students' experiences into simple behavior data. In the next section, we will further discuss approaches that promote ethical and effective human-AI cooperation.

5.2 Understanding Human-AI Cooperation

CSCW research has extensively explored human-AI cooperation in high-stakes areas. Similarly to AI in education, collaborative AI in healthcare includes AI assisting radiologists [111] and clinicians [31, 147] with diagnostic insights, enhancing practices for therapists in rehabilitation assessment [80], and collaborative decision-making between humans and AI. These studies highlight the transformative promises of AI, revealing healthcare practitioners' perceptions of AI automating [111] and supporting routine tasks [31], enriching human work by providing richer information artifacts [80, 111], and allowing human professionals to focus on complex intellectual tasks such as "reasoning" and "interpreting" or translating AI results [111] and execution of AI. By synthesizing human-based and algorithmic insights, healthcare professionals perceive AI positively as a collaborative assistant or colleague (e.g., [147] [111]).

In fields beyond healthcare, AI is also seen as a complementary actor in human work, never eliminating the need for human input [148]. Sometimes, AI is perceived not merely as a tool for collaboration but as a subject [162] or an agent with its own agency (e.g., [62, 66]), influencing organizational structures and stakeholders relationships [66]. These insights resonate with our findings where students identified a potential restructuring of student-teacher communication

dynamics and teachers' future roles and tasks. Understanding the dynamic and complex roles of humans and AI in its situated context is crucial for configuring effective human-AI partnerships [66].

This effective human-AI collaboration requires strategic insights into task allocation between AI and humans, optimizing interactions, and developing new collaboration paradigms [111]. It has been revealed that human-AI collaborative systems can enhance decision-making accuracy [80] while human and AI can learn from each other's strengths: therapists report greater confidence in their decisions when supported by AI's analysis and feedback, while their comments on AI analysis help improve AI systems. This cooperative dynamic has also been seen in AIED where AIED and human can augment each other's interpretations [58, 84]. Despite this, clinicians maintain that AI systems must be verified by human professionals, who remain accountable for the patient treatment outcomes [147]. This resonates with the student preference in our study that teachers hold decision-making authority and overrule AI analysis results to ensure fairness in special cases or pedagogical circumstances. It also resonates with teachers' preferences in another recent study that they should have the final say and take control over AI when AI suggested student pairing plans. These teachers insisted that they know their students' needs better than AI analysis does [77], and that they are responsible for any actions in the classroom when humans and AI share control.

Most importantly, humans must be equipped with an understanding of AI, since the effective cooperation requires a bidirectional exchange of information between humans and AI. It is important to manage expectation in AI and the potential pitfalls of overpromising AI capabilities in human-AI collaboration, especially overhyping the competence of the AI systems [71]. Transparency then becomes critical. For instance, recent CSCW studies ([31] [147]) emphasize transparency and trust in collaborative decision making in healthcare, finding that clinicians desired to understand not only the specific reasoning behind AI decisions (e.g., how the AI processes the inputs and arrives at a decision) but also the basic global properties of the models, including its strengths, theoretical limitations, expertise viewpoints, and design objectives [31]. Enhancing trust involves making AI capability and limitation transparent [147], and providing training for humans for effective use.

5.2.1 Understanding AI: The Need for Teacher Awareness of AI Bias. Our participants frequently perceived AI evaluation as an unbiased third party providing objective insight to mitigate teacher preference and ensure fairness, unaware that algorithms, data collection, and interpretation processes can inherently carry and amplify human bias [41, 85]. In turn, this bias can reinforce social power differentials and discrimination [88] and exacerbate inequality in education [85]. This echoes previous studies where people often assume algorithmic decision-making to be fairer and more robust than human judgment [16], perceiving the algorithms as impartial [41]. However, these unintended systematic algorithmic biases are not yet fully understood [18]. While existing studies (e.g., [18]) explore and propose strategies for reducing algorithm bias in education, they often overlook how teachers should manage biased insights when algorithm bias cannot be fully erased, as is often the case. Hence, potential machine biases should be made transparent to teachers. This knowledge can inform collaboration, such as how much to trust AI outputs versus human judgment when they conflict.

5.2.2 Understanding Humans: Teacher Bias and Attention. Teacher-AI collaboration needs to be aware not only of the inherent AI bias issues but also of how AI insights can bias human perceptions. Our participants expressed concerns that AIED analyses might further prejudice teachers by influencing their expectations of students. Unfortunately, these students' concerns

are corroborated by teachers' concerns reported in another empirical study [85]. That study revealed that data-driven insights from classroom behavior management systems can reinforce and reproduce teachers' bias, and even institutionalize bias, especially against marginalized groups, perpetuating existing inequality in education. Ironically, we have demonstrated earlier that students often perceive AIED as an impartial party that can mitigate teacher bias, a perception that could lead to disappointment when confronted with the reality of teachers' perspectives.

There is another striking contrast between students' and teachers' perspectives: Students in our study generally expect that AI can direct teachers' attention to them when they struggle, particularly those who identify as introverts. They view this system as a savior that can help them secure the attention they desire. However, a previous study [85] discovered that teachers often overlook quieter "invisible" students and focus on students with hypervisibility who frequently gain rewards or penalties. Interestingly, some of our participants expressed the view that they would exaggerate their struggles to turn the system into a communication medium, which might actually work to make them noticeable. This, however, raises the question: does this new system encourage students to reshape their behavior and inadvertently penalize quieter students for their natural disposition? The HCI and CSCW communities need to establish measures of potential system bias and raise awareness among teachers about the extent to which they might be influenced by the system if we are to mitigate unintended sociotechnical influences on teachers.

We have discussed how participants acknowledged the potential for AI and human teachers to counteract each other's limitations within a collaboration. HCI and CSCW research must discern how algorithms can be leveraged to mitigate disparities caused by human bias and how human oversight can critically scrutinize and address biases introduced by AI [125], especially biases related to race, gender, sexuality, among others, to prevent injustices. Additionally, it is important to consider other vulnerable groups such as neurodivergent students who may be misread by AI systems.

5.2.3 Exploring Optimal Frequency of Timely Intervention among Agencies. While AI intervention, such as shared gaze visualization [131] and "co-orchestration" [61] where AI identifies struggling students in real-time and notifies teachers to direct assistance [69], can be beneficial, it may also create tensions between intervening to help the student and disrupting their learning autonomy [143]. Our previous work [56] reveals that students' perceptions of intervention might vary based on their past achievement: confident students often view intervention negatively as disrupting learning independence, whereas students with less confidence typically see it as beneficial timely support. Given the dynamic and diverse nature of the education system, designers need to consider the possible impacts in each specific situation and the effects of individual student learning styles, preferences and needs [30, 95]. HCI and CSCW researchers need to explore how and when interventions occur among AI, teachers, and students, and determine the frequency that proves most beneficial for learning, while taking care of the bias and inclusivity in the design [96], considering the student population is complex [77], sensitive and vulnerable.

5.3 Students Behavior Change Challenges AI Effectiveness

The previous section raised questions about potentially reconstructed students' practices (e.g., to ensure fair attention). As we said earlier when drawing from Suchman's argument [130], the dynamic interaction between humans and technology, situated in their social context, can continually reshape their behavior. In this section, we will discuss how these reshaped human practices can further impact AIED functions.

5.3.1 The Potential Impact of Continuous Observation on Students Behavior. The “Hawthorne Effect” [76], a well-known phenomenon in the field of Human-Computer Interaction (HCI) and psychology, posits that observing a person can result in a change in their behavior. This effect has been considered in the design of many HCI methods (e.g. [20, 78]) - particularly as people tend to behave in a more idealized way under observation, meaning that the observer has to wait for a longer period of time until the monitored people revert to their natural behavior. Our stories show potential behavioral changes (both deliberate and unintended) in student responses to AIED’s observation and evaluation. In the context of AIED, the observers are obviously not humans but digital agents who will help make decisions. Nonetheless, they may affect behavior. Even if this effect is not guaranteed, its possibility requires careful consideration.

Narratives provided by our participants reveal descriptions of student resistance to the learning environment with AI monitoring. This resistance manifests in students gamifying the evaluation mechanism to achieve higher scores or pass tests, not necessarily to learn but rather to navigate the system and attain their degree. Such gamification and rebellion are seen in other areas where technology is imposed on others by people in authority. For example, children and young adults whose whereabouts are geo-tracked by their parents via smartphone apps, have deployed various clever measures to trick or thwart the digital surveillance, and have poured scorn on both parents and tracking app companies [93].

In our study, projections about students’ potential deliberate behavior raises questions about the integrity of the learning process and the potential disconnect between AI-driven evaluations and genuine learning outcomes. AIED’s evaluation approach to quantifying students’ performance enables calculations, easier comparisons, classifications, and rankings [36] among students. This facilitates control and governance, where individuals start to reflect on their behaviors and choices [52] and reconstruct their identities by simply accounting for the values [86]. A culture of performativity arises as students adjust their behaviors according to quantifiable measurements of performance success [90]. We saw these potential inauthentic behavior changes of pandering to the evaluation in our participants’ stories.

To be sure, the occurrence and permanence of these potential behavioral changes in students as they adapt to AIED adoption remains unclear. This raises critical questions for the HCI and CSCW communities: Are these changes acceptable? Will they persist in individuals or arise among future student cohorts when and if the technology becomes commonplace? Can AIED deliver the promised learning outcomes amidst potential disruption of students’ learning processes? While adaptive systems have been developed to detect gaming behaviors [19], we must consider the broader implications of these technologies on the educational landscape. This leads us to question the envisioned future of education: Are AIED systems designed in a way that respects students’ autonomy and aligns with their educational values and goals?

Moreover, we lack measures of to what degree students are likely to engage in artificial behaviors to satisfy or undermine the system. Existing AIED measures primarily focus on system utility and efficiency (e.g., [32]). Future HCI and CSCW work is thus needed to capture the degree of users’ involvement in inauthentic behaviors driven by their perception of the system’s implicit expectations. For instance, it is essential to understand how much trust students may place in the system’s fairness and to what extent they may pretend to exhibit certain behaviors (including dishonestly) to receive positive assessments from the system.

However, here we encounter a dilemma: conducting trials to examine the long-term effects of AIED interventions on student behavior and learning outcomes could be unethical due to potential unintended but harmful impacts. Thus, there should be, we submit, proactive and

comprehensive consideration of potential unintended consequences of AIEd before it is designed and implemented.

5.3.1 Impact of Students Behavior Change on AI Datasets. What further complicates matters from our stories is this: students' potential changed behaviors, when used as further data inputs to AI systems, could further threaten the accuracy of AIEd analysis and functions. Here, changes in behavior can directly impact data quality since datasets contain noisy, incomplete, misleading, and unrepresentative data. A flawed dataset can undermine the robustness of AI tools [16] and lead to misinformed decisions [144]. Unfortunately, inaccurate analysis based on erroneous data might be treated as valuable evidence for directing students' learning [90] while teachers and institutions accept that as an objective fact [104]. It could thereby create profound negative implications for education and fairness. This issue is akin to challenges in medical AI (e.g., [124]), where flawed data leads to costly misdiagnoses and recommendations. In classroom AIEd, these behavior changes can occur subtly, without students' or teachers' awareness, making it challenging to ensure data authenticity and evaluate dataset quality.

In addition to unintended behavior change, participants' narratives exhibit the possibility of deliberate acts of resistance against AIEd, including sabotaging the datasets. In particular, some stories highlighted students' resistance stemming from their doubts about AIEd's decision-making criteria and perceived as procedural injustice: they view the over-simplified evaluation approach and misrepresentation of data as unfair. This finding resonates with a previous study exploring the relationship between student perceptions of classroom justice and student aggression [34]. The study argues that students' perception of procedural fairness (evaluating the fairness of decision-making processes) positively correlates with motivation and effective learning. To restore equity in the classroom, students may alter or cognitively distort their inputs for being evaluated (e.g. performativity), leave the field (e.g. quitting school), or act against the distributor of evaluation outcomes (e.g. hostile intent towards AIEd system). Our data provides concrete examples of these potential behavior changes, extending the discussion to the classroom learning environment with AI.

Importantly, student resistance, aggression, and hostility toward AIEd, particularly any efforts to disrupt the system by manipulating datasets, may be subtle and difficult to identify. The previous study in the context of traditional classroom [34] suggests that individuals who are currently enrolled in the educational system are more likely to engage in moderate, covert, and indirect forms of aggression towards perceived unfairness, due to the lower risk of negative consequences of such acts. However, in the context of AIEd, students' acts of aggression might be more hidden. In our stories, students' aggression targets the system's core values by passively withholding real information or data during the data collection process. This can manifest in various ways, as shown in our data: faking facial emotions or performing learning behaviors that don't reflect their true engagement of understanding; conducting antisocial strategies such as "contaminating the data model" (P30), or even mutinously rallying classmates to disrupt the system collectively. This covert resistance underscores the impact of student behavior changes on AI dataset, an unanticipated challenge that needs to be considered in the AIEd design and adoption.

When our participant P39-T referred to Goodhart's law by saying, "When a measure becomes a target, it ceases to be a good measure" [129], they highlighted how system can be easily manipulated by students altering their behavior to achieve high scores, thereby misleading evaluation progresses. Similar behavior changes to distort data and meet metrics for accomplishing individual's situated goals have been observed in organizations using analytics for

decisions making and assessing worker performance [136]. This organizational study [136] applies signaling theory to describe Goodhart's law: when data produced by workers are used to measure their performance or inform decisions, these data as signals are not neutral. They are often manipulated to convey a specific message or achieve particular objectives. Consequently, measures become ineffective due to the faulty assumption that the signals are honest, reliable, static, or interpreted accurately. The validity of metrics and analytics insights can be thus rendered doubtful.

These signaling acts (either gaming, conforming, or resisting the evaluation mechanism) can potentially empower workers or students, allowing them to assert agency. However, they can also create a feedback loop: technology adoption influences individual behavior, which in turn affects the system performance—as seen in Fig.1. This interplay raises pressing concerns: an overly invasive system like AIEd might not only bring unintended impacts on students and the pedagogical landscape, but also jeopardize system effectiveness. This concern has been similarly raised in the workplace, where, under the influence of monitoring systems, work may become a “strategic, iterative cat-and-mouse game” between organizations attempting to quantify worker performance and workers who continuously adjust their behaviors to manipulate the metrics [136]. Nevertheless, unlike AI used to evaluate work performance, the intended outcome of classroom observation AI is not only to assess students but to help teachers be aware of classroom situations so that to provide targeted assistance to help students to learn.

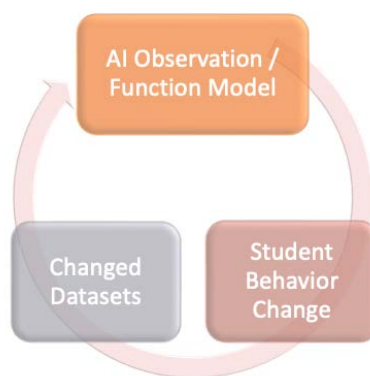


Fig. 1. Recreation of the Repeating Changing Circle.

As HCI and CSCW researchers, we must assess the possible implications on both individuals and ecosystems like education in which AI, students, teachers, and the educational environment are situated, especially when the impacts can be negative, and the outcomes can misalign with the intended objectives.

6 Limitations and Future Work

The Story Completion Method (SCM) is a relatively novel method, and the extent to which the framing of the story stems influences participants is not yet fully understood. As noted earlier in the method section, our scenario framing, particularly the use of terms like ‘monitoring’, may have unintentionally steered responses towards more negative perceptions, eliciting specific emotional and behavioral reactions in narratives. While this approach can help highlight negative feedback and better prepare designers for potential worst-case scenarios, we recognize the importance of careful language use. This insight is valuable for future speculative research, as it

emphasizes the need to refine scenario crafting to minimize biases and ensure a diverse array of responses is captured.

As a speculative method, SCM is used to explore commonly held assumptions from a social constructivist perspective. Our study has shown the effectiveness of SCM in detecting a range of possible implications of future technology adoption. The third-person perspective, rather than asking “your perspective”, enables participants to project their fears and anticipated behavioral changes when addressing their considerations in sensitive contexts like education. The HCI and CSCW community could leverage this method to understand the potential ethical and educational impacts of other emerging Ed-tech, such as ChatGPT in education, from the viewpoint of students, teachers, and institutions alike. Of course, such speculative methods need to be complemented by studies of actual contexts where that is possible and, furthermore, ethical. Our study focuses on the context of AIEd’s adoption in higher education where student groups are university students. As students in different age groups may interact with and perceive the system differently, future studies could extend the participant age range to include under-18 teenagers or younger children. Additionally, future work should expand to other stakeholder groups such as institutions and educators.

7 Conclusion

Integrating AI in education transcends mere technological endeavor. It is a complex process deeply rooted in the social, cultural, and emotional landscape of the educational environment, and involving the diverse backgrounds and learning needs of students. Our study explores how students perceive AI in this intricate setting, revealing the dynamic interplay between AI, teachers, students, and the broader educational system. We spotlight critical concerns: AI’s potential to decontextualize education settings; AI-teacher collaborations fraught with biases and power disparities; and AI’s potential impact on student behavior, which can in turn challenge AI’s technical efficacy. Our findings highlight the possibility that changes in one component of this ecosystem can lead to unintended (negative) implications in others. Since education settings are sensitive and students are vulnerable, there is a need for AIEd designs and implementation to consider these contextual nuances. We would therefore re-emphasize to the CSCW and HCI communities the importance of a holistic approach, one that considers AI, teachers, students, and the learning environment as an interconnected whole, rather than isolating AI in a technical vacuum [53]. Such an approach should help ensure that AIEd systems not only enhance learning outcomes but also integrate ethically into the entire educational experience. This balanced consideration of AI’s potential alongside a range of concerns paves the way for more effective, ethical, and contextually sensitive AI-facilitated education.

References

- [1] Amjed Abu Saa, Mostafa Al-Emran and Khaled Shaalan. 2019. Factors affecting students’ performance in higher education: a systematic review of predictive data mining techniques. *Technology, Knowledge and Learning* 24, 4, 567-598.
- [2] Atezaz Ahmad, Jan Schneider, Dai Griffiths, Daniel Biedermann, Daniel Schiffner, Wolfgang Greller and Hendrik Drachler. 2022. Connecting the dots—A literature review on learning analytics indicators from a learning design perspective. *Journal of Computer Assisted Learning*.
- [3] Karan Ahuja, Dohyun Kim, Franceska Khakaj, Virag Varga, Anne Xie, Stanley Zhang, Jay Eric Townsend, Chris Harrison, Amy Ogan and Yuvraj Agarwal. 2019. EduSense: Practical classroom sensing at Scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3, 1-26.
- [4] Karan Ahuja, Deval Shah, Sujeath Paredy, Franceska Khakaj, Amy Ogan, Yuvraj Agarwal and Chris Harrison. 2021. Classroom Digital Twins with Instrumentation-Free Gaze Tracking. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama, Japan, Article 484. <http://dx.doi.org/10.1145/3411764.3445711>

- [5] Ashraf Alam and Atasi Mohanty. 2022. Foundation for the Future of Higher Education or ‘Misplaced Optimism’? Being Human in the Age of Artificial Intelligence. In *International Conference on Innovations in Intelligent Computing and Communications*. Springer, 17-29.
- [6] Bryan Alexander, Kevin Ashford-Rowe, Noreen Barajas-Murph, Gregory Dobbin, Jessica Knott, Mark McCormack, Jeffery Pomerantz, Ryan Seilhamer and Nicole Weber. 2019. Horizon report 2019 higher education edition. 2019.
- [7] William Alonso and Paul Starr. 1987. *The politics of numbers*. Russell Sage Foundation,
- [8] Nayef Shaie Alotaibi and Awad Hajran Alshehri. 2023. Prosper and obstacles in using artificial intelligence in Saudi Arabia higher education institutions—The potential of AI-based learning outcomes. *Sustainability* 15, 13, 10723.
- [9] Daniela Altun, Christopher Krauss, Alexander Streicher, Christoph Mueller, Daniel Atorf, Lisa Rerhaye and Dietmar Kunde. 2022. Lessons Learned from Creating, Implementing and Evaluating Assisted E-Learning Incorporating Adaptivity, Recommendations and Learning Analytics. In *International Conference on Human-Computer Interaction*. Springer, 257-270.
- [10] Nese Alyuz, Eda Okur, Utku Genc, Sinem Aslan, Cagri Tanriover and Asli Arslan Esme. 2017. An unobtrusive and multimodal approach for behavioral engagement detection of students. In *Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education*. Association for Computing Machinery, Glasgow, UK, 26–32. <http://dx.doi.org/10.1145/3139513.3139521>
- [11] Dana Alzoubi, Jameel Kelley, Evrim Baran, Stephen B. Gilbert, Aliye Karabulut Ilgu and Shan Jiang. 2021. TeachActive Feedback Dashboard: Using Automated Classroom Analytics to Visualize Pedagogical Strategies at a Glance. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama, Japan, Article 312. <http://dx.doi.org/10.1145/3411763.3451709>
- [12] Pengcheng An, Saskia Bakker, Sara Ordanovski, Ruurd Taconis, Chris LE Paffen and Berry Eggen. 2019. Unobtrusively enhancing reflection-in-action of teachers through spatially distributed ambient information. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1-14.
- [13] Cecilia Aragon, Clayton Hutto, Andy Echenique, Brittany Fiore-Gartland, Yun Huang, Jinyoung Kim, Gina Neff, Wanli Xing and Joseph Bayer. 2016. Developing a research agenda for human-centered data science. In *Proceedings of the 19th ACM conference on computer supported cooperative work and social computing companion*. 529-535.
- [14] Kimberly E Arnold and Matthew D Pistilli. 2012. Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge*. 267-270.
- [15] Kimberly E Arnold and Niall Sclater. 2017. Student perceptions of their privacy in learning analytics applications. In *Proceedings of the seventh international learning analytics & knowledge conference*. 66-69.
- [16] Onur Asan, Alparslan Emrah Bayrak and Avishek Choudhury. 2020. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *Journal of Medical Internet Research* 22, 6, e15154. <http://dx.doi.org/10.2196/15154>
- [17] Sinem Aslan, Nese Alyuz, Cagri Tanriover, Sinem E Mete, Eda Okur, Sidney K D'Mello and Asli Arslan Esme. 2019. Investigating the impact of a real-time, multimodal student engagement analytics technology in authentic classrooms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1-12.
- [18] Ryan S Baker and Aaron Hawn. 2022. Algorithmic bias in education. *International Journal of Artificial Intelligence in Education* 32, 4, 1052-1092.
- [19] Ryan S. J. D. Baker, Albert T. Corbett, Kenneth R. Koedinger, Shelley Evenson, Ido Roll, Angela Z. Wagner, Meghan Naim, Jay Raspat, Daniel J. Baker and Joseph E. Beck. 2006. Adapting to When Students Game an Intelligent Tutoring System Springer Berlin Heidelberg, 392-401. http://dx.doi.org/10.1007/11774303_39
- [20] Kathy Baxter, Catherine Courage and Kelly Caine. 2015. *Understanding Your Users : A Practical Guide to User Research Methods*. Elsevier Science & Technology, San Francisco, UNITED STATES.
- [21] Ann Blandford. 2019. HCI for health and wellbeing: Challenges and opportunities. *International Journal of Human-Computer Studies* 131, 41-51. <http://dx.doi.org/10.1016/j.ijhcs.2019.06.007>
- [22] Mark Blythe, Enrique Encinas, Jofish Kaye, Miriam Lueck Avery, Rob McCabe and Kristina Andersen. 2018. Imaginary design workbooks: Constructive criticism and practical provocation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1-12.
- [23] Melissa Bond, Hassan Khosravi, Maarten De Laat, Nina Bergdahl, Violeta Negrea, Emily Oxley, Phuong Pham, Sin Wang Chong and George Siemens. 2024. A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education* 21, 1, 4.
- [24] Alan Borning and Michael Muller. 2012. Next steps for value sensitive design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1125-1134.
- [25] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2, 77-101.
- [26] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4, 589-597. <http://dx.doi.org/10.1080/2159676x.2019.1628806>
- [27] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology* 18, 3, 328-352. <http://dx.doi.org/10.1080/14780887.2020.1769238>
- [28] Virginia Braun, Victoria Clarke, Nikki Hayfield, Hannah Frith, Helen Malson, Naomi Moller and Iduna Shah-Beckley. 2019. Qualitative story completion: Possibilities and potential pitfalls. *Qualitative Research in Psychology* 16, 1, 136-155.

- [29] Barry Brown, Alex S Taylor, Shahram Izadi, Abigail Sellen, Joseph Jofish' Kaye and Rachel Eardley. 2007. Locating family values: A field trial of the Whereabouts Clock. In *International Conference on Ubiquitous Computing*. Springer, 354-371.
- [30] Christine Buchanan, M Lyndsay Howitt, Rita Wilson, Richard G Booth, Tracie Risling and Megan Bamford. 2021. Predicted influences of artificial intelligence on nursing education: Scoping review. *JMIR nursing* 4, 1, e23933.
- [31] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox and Michael Terry. 2019. "Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. *Proceedings of the ACM on Human-computer Interaction* 3, CSCW, 1-24.
- [32] Ben Carterette. 2011. System effectiveness, user models, and user utility: a conceptual framework for investigation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in information retrieval*. 903-912.
- [33] Cecilia Ka Yuk Chan and Wenjie Hu. 2023. Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education* 20, 1, 43. <http://dx.doi.org/10.1186/s41239-023-00411-8>
- [34] Rebecca M Chory-Assad and Michelle L Paulsel. 2004. Classroom justice: Student aggression and resistance as reactions to perceived unfairness. *Communication Education* 53, 3, 253-273.
- [35]
- [36] Patricia Cline Cohen. 2016. *A calculating people: The spread of numeracy in early America*. Routledge,
- [37] A. M. Cox. 2021. Exploring the impact of Artificial Intelligence and robots on higher education through literature-based design fictions. *International Journal of Educational Technology in Higher Education* 18, 1. <http://dx.doi.org/10.1186/s41239-020-00237-8>
- [38] Scotty Craig, Arthur Graesser, Jeremiah Sullins and Barry Gholson. 2004. Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of educational media* 29, 3, 241-250.
- [39] Sidney D'mello and Art Graesser. 2013. AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 4, 1-39.
- [40] Thiago Nazareth De Oliveira, Flavia Bernardini and José Viterbo. 2021. An overview on the use of educational data mining for constructing recommendation systems to mitigate retention in higher education. In *2021 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1-7.
- [41] Rachel Douglas-Jones, Antonia Walford and Nick Seaver. 2021. Introduction: Towards an anthropology of data. *Journal of the Royal Anthropological Institute* 27, S1, 9-25. <http://dx.doi.org/10.1111/1467-9655.13477>
- [42] Anthony Dunne and Fiona Raby. 2013. *Speculative everything: design, fiction, and social dreaming*. MIT press,
- [43] Mohamed Ez-Zaouia, Aurélien Tabard and Elise Lavoué. 2020. EMODASH: A dashboard supporting retrospective awareness of emotions in online learning. *International Journal of Human-Computer Studies* 139, 102411.
- [44] Gloria Milena Fernandez-Nieto, Roberto Martinez-Maldonado, Vanessa Echeverria, Kirsty Kitto, Dragan Gašević and Simon Buckingham Shum. 2024. Data storytelling editor: A teacher-centred tool for customising learning analytics dashboard narratives. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*. 678-689.
- [45] Geraldine Fitzpatrick and Gunnar Ellingsen. 2013. A review of 25 years of CSCW research in healthcare: contributions, challenges and future agendas. *Computer Supported Cooperative Work (CSCW)* 22, 609-665.
- [46] Mary Flanagan, Daniel C Howe and Helen Nissenbaum. 2005. Values at play: Design tradeoffs in socially-oriented game design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 751-760.
- [47] Kenneth R Fleischmann, Sherri R Greenberg, Danna Gurari, Abigale Stangl, Nitin Verma, Jaxsen R Day, Rachel N Simons and Tom Yeh. 2019. Good systems: Ethical ai for cscw. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing*. 461-467.
- [48] Guillaume Fontaine, Sylvie Cossette, Marc-André Maheu-Cadotte, Tanya Mailhot, Marie-France Deschênes, Gabrielle Mathieu-Dupuis, José Côté, Marie-Pierre Gagnon and Veronique Dubé. 2019. Efficacy of adaptive e-learning for health professionals and students: a systematic review and meta-analysis. *BMJ open* 9, 8, e025252.
- [49] Dragan Gašević, Shane Dawson, Tim Rogers and Danijela Gasevic. 2016. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education* 28, 68-84.
- [50] Dragan Gasevic, Jelena Jovanovic, Abelardo Pardo and Shane Dawson. 2017. Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics* 4, 2, 113-128-113-128.
- [51] Víctor González-Calatayud, Paz Prendes-Espinosa and Rosabel Roig-Vila. 2021. Artificial Intelligence for Student Assessment: A Systematic Review. *Applied Sciences* 11, 12, 5467. <http://dx.doi.org/10.3390/app11125467>
- [52] Ian Hacking. 1991. How should we do the history of statistics? *The Foucault effect: Studies in governmentality*.
- [53] Thilo Hagendorff. 2020. The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines* 30, 1, 99-120.
- [54] Laura Hakimi, Rebecca Eynon and Victoria A Murphy. 2021. The Ethics of Using Digital Trace Data in Education: A Thematic Review of the Research Landscape. *Review of Educational Research*, 00346543211020116.
- [55] Bingyi Han, George Buchanan and Dana McKay. 2022. Learning in the Panopticon: Examining the Potential Impacts of AI Monitoring on Students. In *Proceedings of the 34th Australian Conference on Human-Computer Interaction*. 9-21.
- [56] Bingyi Han, Sadia Nawaz, George Buchanan and Dana McKay. 2023. Ethical and Pedagogical Impacts of AI in Education. In *Springer Nature Switzerland*, 667-673.

- [57] Wayne Holmes, Kaska Porayska-Pomsta, Ken Holstein, Emma Sutherland, Toby Baker, Simon Buckingham Shum, Olga C Santos, Mercedes T Rodrigo, Mutlu Cukurova and Ig Ibert Bittencourt. 2021. Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, 1-23.
- [58] Kenneth Holstein, Vincent Aleven and Nikol Rummel. 2020. A conceptual framework for human-AI hybrid adaptivity in education. In *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part I* 21. Springer, 240-254.
- [59] Kenneth Holstein, Bruce M McLaren and Vincent Aleven. 2017. SPACLE: Investigating learning across virtual and physical spaces using spatial replays. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. 358-367.
- [60] Kenneth Holstein, Bruce M. McLaren and Vincent Aleven. 2018. Student Learning Benefits of a Mixed-Reality Teacher Awareness Tool in AI-Enhanced Classrooms. In Springer International Publishing, 154-168.
- [61] Kenneth Holstein, Bruce M. McLaren and Vincent Aleven. 2019. Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms Springer International Publishing, 157-171. http://dx.doi.org/10.1007/978-3-030-23204-7_14
- [62] Ming-Tung Hong, Jesse Josua Benjamin and Claudia Müller-Birn. 2018. Coordinating agents: Promoting shared situational awareness in collaborative sensemaking. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 217-220.
- [63] Evey Jiaxin Huang, Daniel Rees Lewis, Shubhanshi Gaudani, Matthew Easterday and Elizabeth Gerber. 2023. Intelligent Coaching Systems: Understanding One-to-many Coaching for Ill-defined Problem Solving. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1, 1-24.
- [64] Stephen Hutt, Kristina Krasich, James R. Brockmole and Sidney K. D'Mello. 2021. Breaking out of the lab: Mitigating mind wandering with gaze-based attention-aware technology in classrooms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [65] Seray Ibrahim, Asimina Vasalou and Laura Benton. 2022. Understanding the Situated Practices of School Technology Leaders in the Early Stages of Educational Technology Adoption. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New Orleans, LA, USA, Article 312. <http://dx.doi.org/10.1145/3491102.3502120>
- [66] Azra Ismail, Divy Thakkar, Neha Madhiwalla and Neha Kumar. 2023. Public Health Calls for/with AI: An Ethnographic Perspective. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 354. <http://dx.doi.org/10.1145/3610203>
- [67] Weijie Jiang and Zachary A Pardos. 2021. Towards equity and algorithmic fairness in student grade prediction. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 608-617.
- [68] Laura P Jiménez-Mijangos, Jorge Rodríguez-Arce, Rigoberto Martínez-Méndez and José Javier Reyes-Lagos. 2023. Advances and challenges in the detection of academic stress and anxiety in the classroom: A literature review and recommendations. *Education and Information Technologies* 28, 4, 3637-3666.
- [69] Shamyia Karumbaiah, Conrad Borchers, Ann-Christin Falhs, Kenneth Holstein, Nikol Rummel and Vincent Aleven. Teacher Noticing and Student Learning in Human-AI Partnered Classrooms: A Multimodal Analysis.
- [70] Joseph 'Jofish' Kaye. 2006. I just clicked to say I love you: rich evaluations of minimal communication. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, Montréal, Québec, Canada, 363-368. <http://dx.doi.org/10.1145/1125451.1125530>
- [71] Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey T Hancock and Michael S Bernstein. 2020. Conceptual metaphors impact perceptions of human-AI collaboration. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2, 1-26.
- [72] DALLA Khairy, RANIA A Abougalala, MARWA F Areed and SAFAA M ATAWY. 2020. Educational robotics based on artificial intelligence and context-awareness technology: a framework [J]. *Journal of Theoretical and Applied Information Technology* 98, 1817-3195, 2227-2239.
- [73] Abirami Kirubarajan, Dylan Young, Shawn Khan, Noelle Crasto, Mara Sobel and Dafna Sussman. 2022. Artificial intelligence and surgical education: a systematic scoping review of interventions. *Journal of Surgical Education* 79, 2, 500-515.
- [74] Celia Kitzinger and Deborah Powell. 1995. Engendering infidelity: Essentialist and social constructionist readings of a story completion task. *Feminism & Psychology* 5, 3, 345-372.
- [75] Priya C Kumar, Marshini Chetty, Tamara L Clegg and Jessica Vitak. 2019. Privacy and security considerations for digital technology use in elementary schools. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1-13.
- [76] Henry A Landsberger. 1958. Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry.
- [77] LuEttaMae Lawrence, Vanessa Echeverria, Kexin Yang, Vincent Aleven and Nikol Rummel. 2024. How teachers conceptualise shared control with an AI co-orchestration tool: A multiyear teacher-centred design process. *British Journal of Educational Technology* 55, 3, 823-844.
- [78] Jonathan Lazar, Jinjuan Heidi Feng and Harry Hochheiser. 2017. *Research methods in human-computer interaction*. Morgan Kaufmann,
- [79] Hakeoung Hannah Lee and Emma Gargroetzi. 2023. "It's like a double-edged sword": Mentor Perspectives on Ethics and Responsibility in a Learning Analytics-Supported Virtual Mentoring Program. *Journal of Learning Analytics* 10, 1, 85-100.

- [80] Min Hun Lee, Daniel P Siewiorek, Asim Smailagic, Alexandre Bernardino and Sergi Bermúdez Bermúdez i Badia. 2021. A human-ai collaborative approach for clinical decision making on rehabilitation assessment. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1-14.
- [81] Xiaotian Vivian Li, Mary Beth Rosson and Jenay Robert. 2022. A Scenario-based Exploration of Expected Usefulness, Privacy Concerns, and Adoption Likelihood of Learning Analytics. In *Proceedings of the Ninth ACM Conference on Learning@ Scale*. 48-59.
- [82] Phoebe Lin and Jessica Van Brummelen. 2021. Engaging Teachers to Co-Design Integrated AI Curriculum for K-12 Classrooms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama, Japan, Article 239. <http://dx.doi.org/10.1145/3411764.3445377>
- [83] Adam Linson, Yucheng Xu, Andrea R English and Robert B Fisher. 2022. Identifying Student Struggle by Analyzing Facial Movement During Asynchronous Video Lecture Viewing: Towards an Automated Tool to Support Instructors. In *International Conference on Artificial Intelligence in Education*. Springer, 53-65.
- [84] Alex Jiahong Lu, Tawanna R. Dillahunt, Gabriela Marcu and Mark S. Ackerman. 2021. Data Work in Education: Enacting and Negotiating Care and Control in Teachers' Use of Data-Driven Classroom Surveillance Technology. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 452. <http://dx.doi.org/10.1145/3479596>
- [85] Alex Jiahong Lu, Gabriela Marcu, Mark S. Ackerman and Tawanna R Dillahunt. 2021. Coding Bias in the Use of Behavior Management Technologies: Uncovering Socio-technical Consequences of Data-driven Surveillance in Classrooms. In *Designing Interactive Systems Conference 2021*. Association for Computing Machinery, Virtual Event, USA, 508-522. <http://dx.doi.org/10.1145/3461778.3462084>
- [86] Kathleen Lynch. 2014. New managerialism, neoliberalism and ranking. *Ethics in science and environmental politics* 13, 2, 141-153.
- [87] Qianru Lyu, Wenli Chen, Junzhu Su, Kok Hui John Gerard Heng and Shuai Liu. 2023. How Peers Communicate Without Words-An Exploratory Study of Hand Movements in Collaborative Learning Using Computer-Vision-Based Body Recognition Techniques. In *Artificial Intelligence in Education*. Springer Nature Switzerland, 316-326.
- [88] Mark MacCarthy. 2014. Student privacy: Harm and context. *International Review of Information Ethics* 21, 11-24.
- [89] Katerina Mangaroska and Michail Giannakos. 2018. Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies* 12, 4, 516-534.
- [90] Jamie Manolev, Anna Sullivan and Roger Slee. 2019. The datafication of discipline: ClassDojo, surveillance and a performative classroom culture. *Learning, Media and Technology* 44, 1, 36-51.
- [91] Roberto Martinez-Maldonado, Andrew Clayphan and Judy Kay. 2015. Deploying and Visualising Teacher's Scripts of Small Group Activities in a Multi-surface Classroom Ecology: a Study in-the-wild. *Computer Supported Cooperative Work (CSCW)* 24, 2, 177-221. <http://dx.doi.org/10.1007/s10606-015-9217-6>
- [92] Roberto Martinez-Maldonado, Vanessa Echeverria, Gloria Fernandez-Nieto, Lixiang Yan, Linxuan Zhao, Riordan Alfredo, Xinyu Li, Samantha Dix, Hollie Jaggard and Rosie Wotherspoon. 2023. Lessons learnt from a multimodal learning analytics deployment in-the-wild. *ACM Transactions on Computer-Human Interaction* 31, 1, 1-41.
- [93] Jane Mavoa, Simon Coghlan and Bjørn Nansen. 2023. "It's About Safety Not Snooping": Parental Attitudes to Child Tracking Technologies and Geolocation Data. *Surveillance & Society* 21, 1, 45-60.
- [94] Dana McKay and Charlynn Miller. 2021. Standing in the way of control: A call to action to prevent abuse through better design of smart technologies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [95] Danielle S McNamara. 2024. AIED: From Cognitive Simulations to Learning Engineering, with Humans in the Middle. *International Journal of Artificial Intelligence in Education* 34, 1, 42-54.
- [96] Danaë Metaxa-Kakavouli, Kelly Wang, James A Landay and Jeff Hancock. 2018. Gender-inclusive design: Sense of belonging and bias in web interfaces. In *Proceedings of the 2018 CHI Conference on human factors in computing systems*. 1-6.
- [97] Fengchun Miao, Wayne Holmes, Ronghuai Huang and Hui Zhang. 2021. *AI and education: A guidance for policymakers*. UNESCO Publishing.
- [98] Jakub Mlynar, Farzaneh Bahrani, André Ourednik, Nico Mutzner, Himanshu Verma and Hamed Alavi. 2022. AI beyond Deus ex Machina-Reimagining Intelligence in Future Cities with Urban Experts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1-13.
- [99] Damian Murchan and Fazilat Siddiq. 2021. A call to action: a systematic review of ethical and regulatory issues in using process data in educational assessment. *Large-scale Assessments in Education* 9, 1. <http://dx.doi.org/10.1186/s40536-021-00115-3>
- [100] Sungjin Nam and Perry Samson. 2019. Integrating Students' Behavioral Signals and Academic Profiles in Early Warning System Springer International Publishing, 345-357. http://dx.doi.org/10.1007/978-3-030-23204-7_29
- [101] A. Namoun and A. Alshanqiti. 2020. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences* 11, 1.
- [102] Sadia Nawaz, Gregor Kennedy, James Bailey, Chris Mead and Lev Horodyskyj. 2018. Struggle town? Developing profiles of student confusion in simulation-based learning environments. In *35th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education, ASCILITE*. 224-233.
- [103] Rebecca Nicholson, Tom Bartindale, Ahmed Kharrufa, David Kirk and Caroline Walker-Gleaves. 2022. Participatory Design Goes to School: Co-Teaching as a Form of Co-Design for Educational Technology. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New Orleans, LA, USA, Article 150. <http://dx.doi.org/10.1145/3491102.3517667>

- [104] Rosemary Papa and Shadow WJ Armfield. 2018. *The Wiley handbook of educational policy*. John Wiley & Sons,
- [105] Hyanghee Park, Daehwan Ahn, Kartik Hosanagar and Joonhwan Lee. 2022. Designing fair AI in human resource management: Understanding tensions surrounding algorithmic evaluation and envisioning stakeholder-centered solutions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1-22.
- [106] Prasoon Patidar, Tricia J Ngoon, John Zimmerman, Amy Ogan and Yuvraj Agarwal. 2024. ClassID: Enabling Student Behavior Attribution from Ambient Classroom Sensing Systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2, 1-28.
- [107] Yanyi Peng, Masato Kikuchi and Tadachika Ozono. 2023. Development and Experiment of Classroom Engagement Evaluation Mechanism During Real-Time Online Courses. In *International Conference on Artificial Intelligence in Education*. Springer, 590-601.
- [108] Sara Polak, Gianluca Schiavo and Massimo Zancanaro. 2022. Teachers' Perspective on Artificial Intelligence Education: an Initial Investigation. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New Orleans, LA, USA, Article 431. <http://dx.doi.org/10.1145/3491101.3519866>
- [109] Jenny Preece, Yvonne Rogers and Helen Sharp. 2002. *Interaction Design: Beyond Human-Computer Interaction*. John Wiley,
- [110] Paul Prinsloo and Sharon Slade. 2016. Student vulnerability, agency, and learning analytics: An exploration. *Journal of Learning Analytics* 3, 1, 159-182.
- [111] Mohammad Hosein Rezazade Mehrizi. 2023. Pre-framing an emerging technology before it is deployed at work: the case of artificial intelligence and radiology. *Journal of Computer-Mediated Communication* 28, 4, zmad029.
- [112] Samantha Robertson, Tonya Nguyen and Niloufar Salehi. 2021. Modeling assumptions clash with the real world: Transparency, equity, and community challenges for student assignment algorithms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [113] Xingran Ruan, Charaka Palansuriya and Aurora Constantin. 2023. Affective Dynamic Based Technique for Facial Emotion Recognition (FER) to Support Intelligent Tutors in Education. In *International Conference on Artificial Intelligence in Education*. Springer, 774-779.
- [114] Tracii Ryan, Michael Henderson, Kris Ryan and Gregor Kennedy. 2021. Designing learner-centred text-based feedback: a rapid review and qualitative synthesis. *Assessment & Evaluation in Higher Education* 46, 6, 894-912.
- [115] Sdenka Zobeida Salas-Pilco and Yubin Yang. 2022. Artificial intelligence applications in Latin American higher education: a systematic review. *International Journal of Educational Technology in Higher Education* 19, 1, 21.
- [116] Olga C Santos, Milos Kravcik and Jesus G Boticario. 2016. Preface to special issue on user modelling to support personalization in enhanced educational settings. *International Journal of Artificial Intelligence in Education* 26, 3, 809-820.
- [117] Laura Scheepmaker, Kay Kender, Christopher Frauenberger and Geraldine Fitzpatrick. 2021. Leaving the field: Designing a socio-material toolkit for teachers to continue to design technology with children. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [118] Morgan Klaus Scheuerman, Alex Hanna and Emily Denton. 2021. Do datasets have politics? Disciplinary values in computer vision dataset development. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2, 1-37.
- [119] Daniel Schiff, Justin Biddle, Jason Borenstein and Kelly Laas. 2020. What's next for ai ethics, policy, and governance? a global overview. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. 153-158.
- [120] Niall Sclater. 2016. Developing a Code of Practice for Learning Analytics. *Journal of Learning Analytics* 3, 1, 16-42. <http://dx.doi.org/10.18608/jla.2016.31.3>
- [121] Neil Selwyn. 2015. Data entry: Towards the critical study of digital data and education. *Learning, Media and Technology* 40, 1, 64-82.
- [122] Neil Selwyn. 2016. *Is technology good for education?* John Wiley & Sons,
- [123] Lele Sha, Mladen Rakovic, Angel Das, Dragan Gasevic and Guanliang Chen. 2022. Leveraging Class Balancing Techniques to Alleviate Algorithmic Bias for Predictive Tasks in Education. *IEEE Transactions on Learning Technologies*, 1-12. <http://dx.doi.org/10.1109/tlt.2022.3196278>
- [124] Arun Shastri. 2021. What 400 Flawed Healthcare AI Models Can Teach Us. Retrieved from <https://www.forbes.com/sites/arunshastri/2021/06/10/what-400-flawed-healthcare-ai-models-can-teach-us/?sh=545fc3f84355>.
- [125] Jake Silberg and James Manyika. 2019. Notes from the AI frontier: Tackling bias in AI (and in humans). *McKinsey Global Institute* 1, 6.
- [126] Max Smolaks. 2019. Chinese university deploys facial recognition tech to monitor student behavior. Retrieved 13th April, 2023 from <https://aibusiness.com/verticals/chinese-university-deploys-facial-recognition-tech-to-monitor-student-behavior>
- [127] Stephen Snow, Kenneth Radke, Dhaval Vyas and Margot Brereton. 2014. Privacy in the new era of visible and sharable energy-use information. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*. 342-351.
- [128] Erica Southgate, Karen Blackmore, Stephanie Pieschl, Susan Grimes, Jessey McGuire and Kate Smithers. 2019. Artificial intelligence and emerging technologies in schools.
- [129] Marilyn Strathern. 1997. 'Improving ratings': audit in the British University system. *European review* 5, 3, 305-321.
- [130] Lucy A Suchman. 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press,

- [131] Gahyun Sung, Tianyi Feng and Bertrand Schneider. 2021. Learners Learn More and Instructors Track Better with Real-time Gaze Sharing. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 134. <http://dx.doi.org/10.1145/3449208>
- [132] Arnout Terpstra, Alwin De Rooij and Alexander Schouten. 2023. Online Proctoring: Privacy Invasion or Study Alleviation? Discovering Acceptability Using Contextual Integrity. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-20.
- [133] Gareth Terry and Nikki Hayfield. 2020. Reflexive thematic analysis. *Handbook of qualitative research in Education*. Edward Elgar Publishing Limited, 430-441.
- [134] Terrie Lynn Thompson and Paul Prinsloo. 2022. Returning the data gaze in higher education. *Learning, Media and Technology*, 1-13. <http://dx.doi.org/10.1080/17439884.2022.2092130>
- [135] Lisa R Trainor and Andrea Bundon. 2021. Developing the craft: Reflexive accounts of doing reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 13, 5, 705-726.
- [136] Jeffrey W Treem, William C Barley, Matthew S Weber and Joshua B Barbour. 2023. Signaling and meaning in organizational analytics: coping with Goodhart's Law in an era of digitization and datafication. *Journal of Computer-Mediated Communication* 28, 4, zmad023.
- [137] Giovanni M Troiano, Matthew Wood, Mustafa Feyyaz Sonbudak, Riddhi Chandan Padte and Casper Hartevelde. 2021. "Are We Now Post-COVID?": Exploring Post-COVID Futures Through a Gamified Story Completion Method. In *Designing Interactive Systems Conference 2021*. 48-63.
- [138] Giovanni Maria Troiano, John Tiab and Youn-Kyung Lim. 2016. SCI-FI: shape-changing interfaces, future interactions. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. 1-10.
- [139] Giovanni Maria Troiano, Matthew Wood and Casper Hartevelde. 2020. "And This, Kids, Is How I Met Your Mother": Consumerist, Mundane, and Uncanny Futures with Sex Robots. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1-17.
- [140] Yi-Shan Tsai and Dragan Gasevic. 2017. Learning analytics in higher education---challenges and policies: a review of eight learning analytics policies. In *Proceedings of the seventh international learning analytics & knowledge conference*. 233-242.
- [141] Yi-Shan Tsai, Pedro Manuel Moreno-Marcos, Kairit Tammets, Kaire Kollom and Dragan Gašević. 2018. SHEILA policy framework: informing institutional strategies and policy processes of learning analytics. In *Proceedings of the 8th international conference on learning analytics and knowledge*. 320-329.
- [142] Yi-Shan Tsai, Diego Rates, Pedro Manuel Moreno-Marcos, Pedro J Muñoz-Merino, Ioana Jivet, Maren Scheffel, Hendrik Drachslar, Carlos Delgado Kloos and Dragan Gašević. 2020. Learning analytics in European higher education—Trends and barriers. *Computers & Education* 155, 103933.
- [143] Yi-Shan Tsai, Alexander Whitelock-Wainwright and Dragan Gašević. 2020. The privacy paradox and its implications for learning analytics. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. Association for Computing Machinery, Frankfurt, Germany, 230-239. <http://dx.doi.org/10.1145/3375462.3375536>
- [144] Dimitrios Tzimas and Stavros Demetriadis. 2021. Ethical issues in learning analytics: A review of the field. *Educational Technology Research and Development* 69, 2, 1101-1133.
- [145] Shannon Vallor. 2018. An Ethical Toolkit for Engineering/Design Practice. Retrieved from <https://www.scu.edu/ethics-in-technology-practice/ethical-toolkit/>.
- [146] Michael Vierhauser, Clemens Sauerwein, Tobias Antensteiner and Sebastian Hatmanstorfer. 2024. Learning Analytics Support in Higher-Education: Towards a Multi-Level Shared Learning Analytics Framework. In *Proceedings of the 16th International Conference on Computer Supported Education*.
- [147] Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Wang, Haiyi Zhu, Yvonne Gao, Xiangmin Fan and Feng Tian. 2021. "Brilliant AI doctor" in rural clinics: challenges in AI-powered clinical decision support system deployment. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1-18.
- [148] Dakuo Wang, Justin D Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz and Alexander Gray. 2019. Human-AI collaboration in data science: Exploring data scientists' perceptions of automated AI. *Proceedings of the ACM on human-computer interaction* 3, CSCW, 1-24.
- [149] Zhifeng Wang, Longlong Li, Chunyan Zeng and Jialong Yao. 2023. Student learning behavior recognition incorporating data augmentation with learning feature representation in smart classrooms. *Sensors* 23, 19, 8190.
- [150] Jenny Waycott, Greg Wadley, Stefan Schutt, Arthur Stabolidis and Reeva Lederman. 2015. The Challenge of Technology Research in Sensitive Settings: Case Studies in sensitive HCI. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*. 240-249.
- [151] Alexander Whitelock-Wainwright, Yi-Shan Tsai, Hendrik Drachslar, Maren Scheffel and Dragan Gašević. 2021. An exploratory latent class analysis of student expectations towards learning analytics services. *The Internet and Higher Education*, 100818.
- [152] Ben Williamson. 2019. Datafication of education: a critical approach to emerging analytics technologies and practices. *Rethinking pedagogy for a digital age*, 212-226.
- [153] Billy Tak-Ming Wong, Kam Cheong Li and Simon K. S. Cheung. 2022. An analysis of learning analytics in personalised learning. *Journal of Computing in Higher Education*. <http://dx.doi.org/10.1007/s12528-022-09324-3>
- [154] Richmond Y Wong and Vera Khovanskaya. 2018. Speculative design in HCI: from corporate imaginations to critical orientations. In *New Directions in Third Wave Human-Computer Interaction: Volume 2-Methodologies*, Springer, 175-202.

- [155] Matthew Wood, Gavin Wood and Madeline Balaam. 2017. " They're Just Tixel Pits, Man" Disputing the Reality of Virtual Reality Pornography through the Story Completion Method. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5439-5451.
- [156] Wanli Xing, Xin Chen, Jared Stein and Michael Marcinkowski. 2016. Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in human behavior* 58, 119-129.
- [157] Lixiang Yan, Roberto Martinez-Maldonado, Linxuan Zhao, Xinyu Li and Dragan Gašević. 2023. Physiological Synchrony and Arousal as Indicators of Stress and Learning Performance in Embodied Collaborative Learning. In *Artificial Intelligence in Education*. Springer Nature Switzerland, 602-614.
- [158] Kexin Bella Yang, Vanessa Echeverria, Zijing Lu, Hongyu Mao, Kenneth Holstein, Nikol Rummel and Vincent Aleven. 2023. Pair-Up: Prototyping Human-AI Co-orchestration of Dynamic Transitions between Individual and Collaborative Learning in the Classroom. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Hamburg, Germany, Article 453. <http://dx.doi.org/10.1145/3544548.3581398>
- [159] Mireia Yurrita, Tim Draws, Agathe Balayn, Dave Murray-Rust, Nava Tintarev and Alessandro Bozzon. 2023. Disentangling Fairness Perceptions in Algorithmic Decision-Making: the Effects of Explanations, Human Oversight, and Contestability. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-21.
- [160] Olaf Zawacki-Richter, Victoria I. Marín, Melissa Bond and Franziska Gouverneur. 2019. Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education* 16, 1. <http://dx.doi.org/10.1186/s41239-019-0171-0>
- [161] Qiyang Zhang and Amanda Neitzel. 2023. Choosing the right tool for the job: Screening tools for systematic reviews in education. *Journal of Research on Educational Effectiveness*, 1-27.
- [162] Rui Zhang, Nathan J McNeese, Guo Freeman and Geoff Musick. 2021. " An ideal human" expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3, 1-25.
- [163] Wei Zhang, Mingxuan Cai, Hong Joo Lee, Richard Evans, Chengyan Zhu and Chenghan Ming. 2024. AI in Medical Education: Global situation, effects and challenges. *Education and Information Technologies* 29, 4, 4611-4633. <http://dx.doi.org/10.1007/s10639-023-12009-8>
- [164] Zhe Zhang and Ling Xu. 2022. Student engagement with automated feedback on academic writing: a study on Uyghur ethnic minority students in China. *Journal of Multilingual and Multicultural Development*, 1-14. <http://dx.doi.org/10.1080/01434632.2022.2102175>

Appendix

Appendix. Story Completion Tasks

* The codes presented throughout the stories correspond to the functions and techniques outlined in Table 1, which are based on existing research proposals and AI-driven educational tools. Each code, such as [ATI] for Automatic Evaluation and Timely Intervention, is used to reference specific AI functionalities integrated into classroom settings.

Scenario 1: During this summer break, Charlie's University has implemented an intelligent AI education system to assist teaching. It is the first lecture of this semester. Charlie walked into the classroom. He heard his classmates whispering: There will be an In-Class Surveillance by using the Classroom Behavior Analysis System:

- When students walk into the classroom, the system records their attendance by Facial Recognition Technology [CEM]. It will then be generated as engagement data stored in the Classroom Attendance System.
- During the class, the system will capture students' affective (emotional) states, such as when a student is either confused or confident, using an Emotion Recognition Technology [CEM]. Then the analysis report will be created for teachers to understand students' emotional readiness and cognitive needs for their learning [AEA][RBA].
- Each student's learning material and feedback will be generated via real-time analysis in class [RBA], to build a profile of the student's progress in each class. Across a term, this data will generate a personal "Education Profile".

Now the class bell rings, Charlie feels ...

Scenario 2: The lecturer said: "Let's have a group discussion". Charlie then heard another whisper from his classmates:

- During the in-class groupwork practice, the system will also identify what each student is doing and saying, by using Facial Recognition and Voice Recognition. Eye Tracking Technology will be

used to capture the amount of time that each student is focused on each learning resource [EGT]. Then the analysis report will be created by Eye Tracking for the teachers to understand the learning progress of the class and the efficacy of the chosen learning resources.

Now the classroom becomes noisy and full of conversations, Charlie ...

Scenario 3: On another day, Charlie learns that there is a digital “Education Profile” that tracks all education-related data for every student (e.g., data of in-class emotion, voice, facial expressions, gesture, academic scores). This profile provides real-time analysis during class and identifies each students’ academic weaknesses and strengths [ATI]. Based on this analysis, it suggests personalized learning pathway and generates tailored recommendations for learning resources targeting areas of improvement [APR].

- The Education Profile will work with an Academic Early Warning System [PAW]: It has a high accuracy of predicting when a student is likely to drop-out or fail from a subject. It captures the learner’s mindset and how it changes over time, such as a student’s attitude towards a lecture changed from positive to negative and unwillingness to engage increased. It has strong indicators to track students’ progress and identify changes in each student’s learning confidence and motivation toward a course. The system can automatically trigger an alert to ask for timely interventions to support students’ learning, such as providing targeted feedback to teachers [ATI].
- The Education Profile will also work with an Intelligent Evaluation System, which automatically assesses homework and provides customized learning assistance when students’ struggles are identified [ATI].

Later that day Charlie is in a group exercise on algebra, on a topic he finds quite hard. He knows the Education Profile is going to be used to see how he and his class are doing What happens next?

Received January 2024; revised July 2024; accepted October 2024.

4.7 ADDITIONAL REFLECTIONS

While generating the findings, I noticed that students' ethical concerns and perceptions varied significantly between the different AIED tools presented in the scenarios. As mentioned briefly in the methodology section, there was a notable shift in attitudes from Scenario 1 and Scenario 2, which both emphasised AI monitoring, to Scenario 3, which focused on functions like customised learning suggestions derived from data collected through AI facial and emotional analysis. While students found the system more acceptable in Scenario 3, this does not diminish the discomfort or concerns expressed in the first two scenarios regarding being monitored while studying. These differing attitudes informed my decision to design the Study 3 experiment, which examines students' specific preferences and attitude changes toward distinct functions and elements of AIED - this will be discussed in detail in Chapter 6.

How I presented the findings of this study has evolved through multiple rounds of feedback and revisions, shaped by reviewers' revision comments or rejections. A recurring critique was the perceived negativity of the results, with reviewers often stating that "the results paint a very negative picture." Many suggested toning down the claims to avoid positioning the research as opposing technological advancements. However, there were also contrasting viewpoints, with one reviewer urging a more critical stance and suggesting that certain technologies should be outright discouraged in education. Throughout this process, I have remained steadfast in my commitment to presenting findings that are faithful to the data, focusing on what participants' narratives told me, and deriving implications that prioritise the well-being of users. The negative results, though speculative, can be seen as worst-case scenarios, serving as critical guideposts for responsible AIED design.

Chapter 5: Powerful Stakeholders’ Perspectives: Ethical Gaps and Power Dynamics in AIEd Decision-Making

The previous chapter (Chapter 4) explored students’ ethical concerns and perspectives on AIEd adoption. This chapter presents the findings from Study 2, which examines whether other key stakeholders that are more powerful in AIEd adoption decision-making have considered and addressed students’ concerns. Study 2 addresses RQ2: “*How do stakeholder dynamics shape ethical risks in AIEd adoption?*” 30 in-depth interviews were conducted with parents, developers, school governors and teacher representatives. Through reflexive thematic analysis, Study 2 uncovers three critical gaps in ethical AIEd design and adoption and the underlying power dynamics shaping these issues.

Viewed in relation to Study 1, the findings from Study 2 reveal a striking disconnect between students’ concerns and considerations of decision-makers, exposing a fundamental misalignment in AIEd adoption (discussed further in Chapter 7). This gap underscores how differing ethical considerations and value systems can exacerbate the existing challenge of AIEd adoption. This finding marks a crucial turning point in this thesis to explore the ethical challenges of AIEd adoption. It shifts the focus from individual stakeholder concerns to the structural problems of AIEd governance. On the one side, students’ values are overlooked; on the other, current accountability mechanisms fail to protect students from potential harm. This tension sets the stage for Study 3, which investigates whether students align with the value priorities and trade-offs expected or imposed by more powerful stakeholders.

This chapter begins with a brief recap of the study’s approach (Section 5.1) and its contribution to the thesis (Section 5.2). It then presents the under-review manuscript arising from Study 2 (Section 5.3), followed by additional reflections which has not been covered in the paper (Section 5.4).

5.1 APPROACH FOR STUDY 2

As detailed in Chapter 3 Section 3.5, Study 2 employed semi-structured interviews to interrogate more powerful stakeholders' lived experiences, ethical concerns, and decision-making processes in AIEd adoption. 30 participants were interviewed (10 school governors and teacher representatives, 10 parents, and 10 developers), with each session lasting between 45 and 70 minutes.

To ensure both depth and consistency, interviews begin with general questions about participants' roles and experiences with AIEd before transitioning to more specific questions about ethical considerations and responsibilities. To maintain continuity with Study 1, participants were introduced to the same AIEd scenarios used in Study 1's story prompts. This approach also offered a consistent and clearer understanding of AIEd artefacts and their application among different participants, allowing for more informed and focused reflection on navigating AIEd adoption in their perspective.

The interview data were analysed using Reflexive Thematic Analysis (RTA) to uncover patterns, themes, and tensions across stakeholder groups. This analysis allowed the study to identify shared concerns, knowledge gaps, conflicting priorities, and deepen the understanding of structural challenges in ethical AIEd adoption.

5.2 FINDINGS AND CONTRIBUTION TO THE THESIS

Study 2 contributes to this thesis's overarching research question – “What are the ethical challenges in adopting artificial intelligence in education (AIEd) from the perspectives of different stakeholder groups?” – by turning to the perspectives of the more powerful stakeholders in decision-making in AIEd design and adoption. By focusing on decision-makers' responses, this study uncovered three critical ethical gaps that emerge in AIEd adoption processes: 1) systemic exclusion of students – arguably the most affected yet least powerful group – in decision-making processes about AIEd adoption; 2) developers' pragmatism under significant project pressures including project delivery demand and deadlines that relate to financial burden, and limited engagement with ethical considerations post-adoption; 3) school's limited technological literacy necessary to critically evaluate AIEd systems or foresee their potential consequences. Underpinning these ethical gaps are deeply rooted power dynamics involving knowledge asymmetries and authority imbalances. Developers, as technical experts with knowledge power, wield significant control over the design and implementation process, while schools and parents

- despite being decision-makers - often defer to this technical expertise. Students, as the most vulnerable group, have the least agency but bear the brunt of AIED's impacts, positive or negative. These dynamics entrench disadvantages and contribute to flawed decision-making that neglects ethical considerations in favour of other priorities (e.g., perceived educational outcomes or institutional pressures).

Study 2 contributes this power dynamic model of AIED adoption (Fig. 4). This model made visible how different forms of power, including knowledge and authority, are distributed among stakeholders, providing a stakeholder-centred framework for addressing the ethical challenges of AIED adoption.

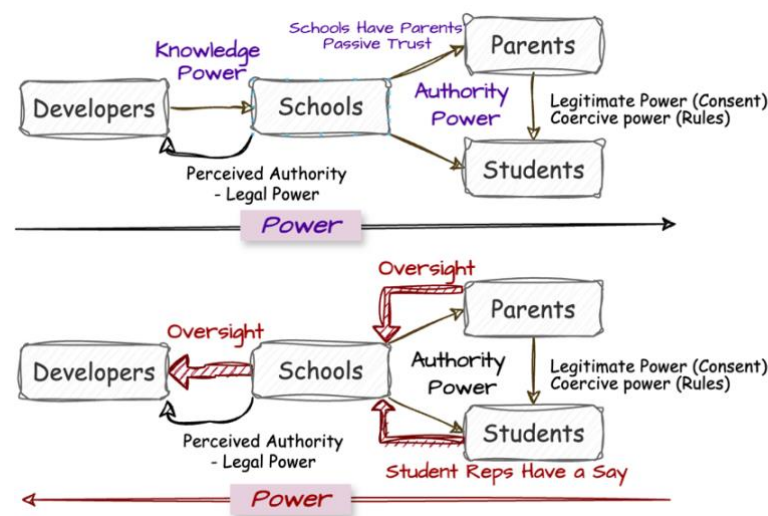


Fig. 4 Power dynamics that may facilitate ethical gaps in AIED adoption and ways of addressing them. (This figure is presented in its original form from the manuscript in Section 5.3 of this thesis.)

Study 2 directly addresses *Problem 2: Limited Understanding of Stakeholder Dynamics in AIED Adoption*, by examining how power is distributed across stakeholder groups, and how this shapes the design, adoption, and governance of AIED systems. It offers a nuanced account of how institutional pressures, varying levels of AI literacy, and role-specific priorities influence stakeholders' engagement in ethical decision-making. Crucially, the study reveals the structural exclusion of students from these processes, highlighting a critical gap between those who decide on AIED systems and those most affected by them. By surfacing these dynamics, Study 2 contributes empirical insights to inform more effective governance mechanisms and foster more informed, balanced, and ethical adoption processes. For instance, the findings advocate improving the AI literacy of key decision-makers (e.g., schools and parents), and integrating student representatives

into decision-making structures, to help mitigate power imbalances and strengthen ethical oversight.

This study also contributes to addressing *Problem 4, Limited Understanding of Ethics in AIEd*. It shifts the ethical focus from merely evaluating the impacts of AIEd technologies to interrogating the broader sociotechnical systems and institutional conditions through which they are designed and adopted. The findings demonstrate that ethical risks do not reside solely in the technology itself but emerge from the relational and procedural dynamics of governance. Study 2 offers a systemic perspective to enrich existing ethical frameworks by underscoring the importance of context and power in shaping ethical outcomes.

More broadly, Study 2 advances the thesis's overarching aim of promoting ethical, context-sensitive AIEd design and adoption. It challenges the assumption that user-centred design approaches (e.g., participatory design) alone can ensure ethical outcomes - particularly when applied without consideration of broader social and institutional dynamics. This study provides a vivid case showing how decision-makers prioritise teachers as direct users, while treating students as passive recipients, reinforcing existing power hierarchies. This contribution extends beyond AIEd, offering critical lessons for other multi-user systems such as AI in healthcare or hiring (further discussed in Chapter 7). It emphasises the need to situate design practices within their wider sociotechnical context and power-related complexities, to ensure more equitable and responsible technology implementation.

5.3 ETHICAL GAPS IN AIED DESIGN AND ADOPTION: STUDENTS' SYSTEMATIC EXCLUSION, DEVELOPERS' PRAGMATISM, SCHOOLS' LIMITED AI LITERACY

Study 2 findings have resulted in a manuscript that was submitted for publication to IJHCS in March 2025.

The paper manuscript begins on the next page.

Ethical Gaps and Power Dynamics in Decision-Making about AI Adoption: The Case of AI for Monitoring Student Learning in Education

Adopting AI in multi-stakeholder contexts often involves complex ethical challenges, particularly when decision-makers are not directly impacted by the technology. Ethical challenges arise not only from the design of the AI technology but also from the social-technical systems within which it is developed and deployed. If these issues are not addressed in the decision-making, disempowered stakeholders may be harmed. This paper examines how stakeholder dynamics in AI adoption contribute to negative outcomes, using a case study of AI-based student monitoring in K-12 education. These systems analyse students' emotions, concentrations, and classroom performance to inform teaching. We conducted 30 semi-structured interviews with key stakeholders: parents, school representatives, and developers. Our findings identify three key factors with potential to negatively affect decision-making: exclusion of vulnerable stakeholders (students) from decisions, pragmatism of and pressure faced by AI developers, and deficiencies in technological literacy limiting due diligence among school governors. These factors arise from power dynamics related to knowledge and authority, resulting in insufficient accountability mechanisms to mitigate the ethical risks of AI adoption in education, such as negative and enduring impacts on student well-being and learning. The paper suggests addressing these power imbalances by increasing AI ethics literacy of key stakeholders to ensure effective oversight. By focusing on systemic gaps throughout AI design and adoption processes, this work shifts the ethical focus from the impacts of AI adoption to the processes that create negative outcomes. Beyond education, these insights inform various other multi-stakeholder AI adoption contexts such as digital health and public services.

CCS CONCEPTS • Human-centred computing -> Empirical studies in HCI; Empirical studies in collaborative and social computing.

Keywords: Ethics, AI in Education, User-centred Design, Accountability, Power Relations, Stakeholder Relationships, Multi-User Systems

Article Highlights:

- Power dynamics (knowledge, authority) create accountability gaps in AIEd adoption decisions.
- Students' values and needs are systematically excluded in AIEd design and adoption processes.
- Developers' pragmatism, shaped by organisational pressure and knowledge power, overshadows ethics.
- Schools' AI illiteracy in AIEd decisions fails to shoulder the ethical responsibility.
- Mitigate power imbalance by enhancing AI literacy to enable effective ethical oversight.

1 INTRODUCTION

AI adoption can be an ethically complex and intricate undertaking, especially where multiple stakeholders make decisions about the technology's design and implementation. Ethical problems with AI may be exacerbated when decision-makers do not directly experience unintended negative impacts. This occurred, for instance, when Amazon's AI hiring tool for streamlining job candidate selection discriminated against female applicants due to biases in historical data [13], and when a healthcare algorithm in US hospitals exhibited racial bias in its patient risk predictions [54]. Insufficient ethical reflection and action in the design and development process of technologies is a problem the HCI community has long sought to address (e.g., [37, 75, 132]). However, we still lack a detailed understanding of how certain decisions shape an AI system's design, implementation, and governance in various real-world circumstances.

HCI research has explored human factors influencing the ethical design and adoption of AI systems (e.g., [7, 24, 37]), especially in multi-user settings and high-stakes domains (e.g., [25, 70, 88]). HCI scholars have acknowledged the conflicts between the values of different stakeholder groups and inherent power imbalances, such as in student-teacher [93], patient-healthcare provider [28, 29], and employee-employer relations [88]. These conflicts can compromise the ethical deployment of AI and may disproportionately affect disempowered users (e.g., [88]). Effective solutions for resolving value compromises, however, remain elusive. What is currently lacking is an understanding of how stakeholder dynamics affect AI adoption and address or contribute to its ethical challenges. In this paper, we explore how power dynamics among stakeholders can lead to ethical gaps, taking the example of monitoring-focused AI in Education (AIEd).

Monitoring AIEd encompasses AI technologies such as facial recognition, gaze tracking, and emotional detection to observe, monitor, and analyse students' behaviours, activities, and performance in educational settings [76]. These technologies, which are the focus of our study, have already been deployed in educational institutions (e.g., [104]). There they have raised ethical issues such as data privacy [15, 23, 57, 63, 110] and algorithmic bias [68, 77] related to race, ethnicity, nationality, or gender that can unfairly affect student outcomes [11, 132]. However, broader sociotechnical implications may arise including, synthetic student behaviour [43], impacts on wellbeing and disruption of learning autonomy [46], magnification of teacher biases [44, 68], and exacerbation of educational inequality [11]. When such challenges emerge, they can be very challenging to address. While much is known about various negative effects of technology adoption [90], empirical studies mainly focus on concerns from individuals directly interacting with technology, such as students [43, 52] and teachers [58, 92] who are already dealing with any consequences of decisions made during the setup process. Less attention has been paid to the decision-making processes behind implementation, and how stakeholders navigate decisions which can contribute to these ethical consequences. This study aims to address that gap.

Our study asks: *How do stakeholder dynamics shape ethical challenges in the adoption of multi-user systems like AIEd?* We conducted 30 semi-structured interviews with key stakeholders involved in monitoring-style AIEd rollout in K12 schools: parents, school representatives, and developers. The findings reveal three systematic ethical gaps: the exclusion of vulnerable stakeholders (students) in decision-making; pragmatism and pressure felt by AI developers; and limited technological literacy and failure of 'due diligence' amongst AI customers (schools). Underpinning these gaps are two types of power dynamics, involving knowledge and authority. These power imbalances can entrench disadvantages for the most vulnerable stakeholder group—students—and result in ineffective accountability mechanisms to protect students from unintended ethical harm. We further uncover limitations of conventional user-centred design approaches which often marginalise disempowered user groups as passive subjects of system impacts, both positive and negative. Finally, we discuss how flawed decision-making due to power imbalances might be mitigated, such as by improving AI literacy amongst certain decision-makers and empowering them to make more informed ethical choices and effective oversights. Our findings have ramifications beyond AIEd for multi-user systems where power imbalances are salient.

This research contributes a critical understanding of how ethical considerations are negotiated in real-world AIEd design and adoption processes, revealing how power dynamics and value priorities shape ethical decision-making. It expands our understanding of the AI adoption challenges, by shifting the focus from the impacts of technology design alone to a more systemic view of how ethical systems are created and implemented. By exposing how stakeholder dynamics can obscure or override ethical concerns, this study offers empirical evidence of the structural barriers to inclusive and ethical AIEd governance. The study complements existing normative ethical frameworks for AI, by broadening the ethical scope, advocating for comprehensive "organisational and systemic approaches" to ethical issues

[88] that address both traditional AI ethics concerns (e.g., algorithmic fairness and data security) and the deeper decision-making processes that drive negative outcomes.

2 RELATED WORK

This section first reviews HCI studies on the challenges in AI development or adoption involving multiple stakeholders (Section 2.1). It then introduces AI-enabled classroom performance management systems and reviews possible ethical impacts associated with their adoption (Section 2.2). Finally, it briefly discusses existing ethics models for AI and AI in education, which our approach of exploring ethical challenges seeks to complement (Section 2.3).

2.1 Challenges in AI Adoption: Stakeholders Dynamics

AI systems involve multiple stakeholders each of whom play critical roles in shaping the development, deployment, and governance of the technology. These stakeholders – developers and designers, adoption decision-makers, and users – can either influence or be influenced by the ethical challenges that arise throughout the AI adoption process. This section first reviews the challenges faced by AI practitioners, particularly developers and designers, who are key players in shaping AI systems [40] (Section 2.1.1). It then examines the value tensions among various stakeholders including multiple users and developers within the broader ecosystem of AI adoption (Section 2.1.2).

2.1.1 Challenges for AI practitioners

A significant body of HCI research has focused on understanding and addressing the challenges faced by AI practitioners in developing ethical systems, sometimes offering tools designed to support creating more ethical AI [14, 51]. For example, some of these tools help to assess and mitigate unfairness for machine learning algorithms. However, while these tools empower developers to address ethical challenges arising from technical aspects, they do not fully address the needs of practitioners in real-world industry settings [51]. There are broader challenges that practitioners face in the ecosystem of AI development.

Studies reveal that individual practices, personal values, professional expertise, and ethical beliefs play a crucial role in guiding practitioners such as designers and developers in their decision-making processes (e.g., [41, 95, 125]). For instance, individuals often draw on their personal values to evaluate design decisions, creating a strong connection to ethical stances in their work [41, 95]. In some cases, personal or reputational motivations drive ethical actions [125].

However, these individual values, motivations, and practices can come into conflict with organisational structures, pressures and goals such as budget limitations, tight project schedules, client expectations, and the organisation's overarching focus on profitability or usability [26, 41, 95]. While AI practitioners may recognise ethical concerns in designing and developing AI systems, organisational pressures often hinder their execution of ethical decisions. Aligning individuals' responsible AI initiatives with organisational business goals is often difficult [95]. Furthermore, practitioners' ethical responsibility and engagement are often constrained by their professional roles and identities [26]. For instance, UX designers and engineers might focus on user-centred goals but lack the authority to challenge broader efficiency-driven organisational decisions [26]. HCI scholars have highlighted this tension between individual practices and the organisational environment, emphasising the need for organisational support to enable ethical practices [78, 95].

It has been revealed that industrial organisations often adopt a reactive approach to creating responsible AI systems, driven by external pressures such as media scrutiny, reputational risks, or regulatory requirements, rather than proactively addressing ethical concerns [95]. This reactive stance, combined with the lack of accountability and unclear roles within organisations, results in individuals shouldering the burden of initiating responsible AI efforts without

sufficient structural support [95]. Such conditions compound the challenges discussed earlier, as practitioners are left to navigate ethical concerns in isolation. Similarly, [125] highlights the presence of fragmented accountability in AI development. In modular AI systems, different teams or organisations develop interconnected software modules, leading to divided responsibility. Within this environment, individuals often view their role as confined to specific technical tasks, leaving broader ethical concerns to others, which further reduces accountability for the overall system [125].

Gray and colleagues [40] stress the inherent ethical complexity faced by technology designers, particularly the power dynamics between stakeholders. Designers may inadvertently reinforce structural inequalities or employ deceptive design techniques (e.g., "dark patterns") to meet organisational or client goals. While these practitioners may initiate such decisions, their work is nonetheless deeply influenced by and has far-reaching impacts on various stakeholders across the ecosystem. Gray and colleagues [40] emphasise that designers must navigate and balance competing ethical considerations from multiple stakeholders, a challenge that becomes even more pronounced as AI adoption intersects with diverse interests and values. This dynamic not only complicates the challenges practitioners face within their organisations, but also sets the stage for broader tensions between values in AI development that involve not only designers but also other users and stakeholders within the ecosystem.

2.1.2 Stakeholder Tensions

Ethical complexities and challenges also extend beyond practitioners to involve a wider range of stakeholders. HCI research has long investigated various stakeholders' perspectives to promote ethically informed AI systems. This effort seeks to ensure that deployed AI systems align with diverse stakeholder values, thereby ensuring fairness (e.g., [25]), acceptance by all groups (e.g., [88]), protection for vulnerable groups (e.g., [59]) and successful uptake (e.g., [24]). However, this research frequently reveals complex tensions and divergences between group interests. For instance, in multi-user contexts like HR systems, firms and employees have differing interpretations of the term "fairness" [88]. This disparity, coupled with the research gap in effectively engaging less powerful stakeholders, leads to design decisions that disproportionately favour dominant groups [88, 98]. This may disadvantage underrepresented cohorts and potentially perpetuate power inequalities. Similar disparities manifest in education, where, for example, teachers and students have different preferences for control levels for students' interactions with intelligent tutoring systems [130]. This unaddressed tension creates design challenges and potentially unfair and unequal collaborative opportunities for students [130].

Value conflicts in AI systems extend beyond user groups, to AI practitioners. One study on AI recommendation systems revealed tensions between developers and users in value interpretation and practices regarding privacy, transparency, and fairness [24]. In these ways, design decisions can mismatch with user expectations. Developers also face challenges in addressing differing stakeholders' needs, as seen, for example, in health chatbot design [55]. In education, algorithms used to assign students to public schools have been found to overlook socioeconomic inequalities, causing a misalignment between intended values and real-world harms [97]. HCI has advocated for ongoing stakeholder engagement [32], value trade-offs [55, 97], and safeguards to protect vulnerable groups from powerful stakeholders [97].

While significant efforts have been made to identify challenges in creating ethical AI systems, such as the tensions faced by practitioners within industry organisations over ethical actions (Section 2.1.1) and value conflicts between stakeholders in real-world AI design and adoption (Section 2.1.2), there is a limited understanding of how these challenges interact across different stakeholders. Specifically, studies rarely explore how practitioners navigate tensions between internal organisation goals and external, competing ethical considerations from various stakeholders within specific adoption contexts. Furthermore, little is known about the interactions between decision-makers – such as users

or clients – and AI practitioners in shaping the design and adoption of AI systems, and how the values of these stakeholders ultimately influence the decisions made. There is a notable gap in empirical research on the inter-stakeholder dynamics within the real-world AI design and adoption ecosystems, particularly regarding how these dynamics may obscure or affect ethical concerns. These dynamics, in turn, can lead to adoption decisions that either mitigate or exacerbate ethical challenges. Understanding how AI adoption decisions are framed is vital to ensuring that positive intentions behind ethical AI development are realised. This paper aims to explore these complexities within the AI in Education (AIEd) context, where, as we shall see, power imbalances are prevalent [74].

2.2 Monitoring-Focused AIEd and its Potential Ethical Impacts

This study involves AI-enabled systems for classroom monitoring and detection of students’ in-class performance [4, 9, 49]. This section first outlines the current systems in this field (Section 2.2.1) and then reviews their potential ethical impacts (Section 2.2.2).

2.2.1 Monitoring-Focused AI Systems in Classrooms

AI-enabled classroom performance management systems heavily rely on data about students’ behaviour and attitudes obtained through in-class monitoring cameras [4] which embed AI-enabled facial and emotion recognition techniques [100]. These systems detect students’ behaviours from head movements [1, 89], emotional expressions [9, 31], and eye direction [2, 53]. This data allegedly gives real-time insights into student performance and engagement [9, 53, 100] and facilitates predictions of performance or dropouts [62, 129], early warning [82], automated intervention [50], and tailored learning recommendations [3] (see Table 1).

Most work on designing classroom AIEd tools positions teachers as the primary users [9]. For instance, using the information collected on students, some AIEd seeks to guide teachers on improving teaching strategies [2, 6, 31] and allocating their attention based on student needs [6].

Table 1: Existing Monitoring-Focused AIEd Artifacts and Related Functions

Category	Application Examples
Classroom Facial / Emotional / Eye Gaze Monitoring	[2, 53, 64, 100]
Awareness & Emotional Engagement Analysis	[31, 49, 50]
Real-Time Learning Performance Analysis	[1, 9, 64, 89]
Automatic Evaluation and Timely Intervention	[3, 9]
Prediction and Academic Early Warning	[38, 62, 82, 129]
Adaptive Personalised Recommendation	[3]

Despite the intended benefits of this technology, existing HCI studies on AIEd development often lack ethical awareness. For instance, some studies underscore how AI helps teachers to scaffold learning [9] and minimise students’ off-task behaviours [49], while overlooking possible negative impacts.

2.2.2 Ethical Impacts of Monitoring AIEd

Recent HCI studies have revealed students’ responses to classroom monitoring AI. Many students perceive AI monitoring as a classroom panopticon [43]. In response, they may change their behaviours to avoid AI-determined penalties. The perception of a panopticon may foster a culture of acting rather than genuine learning. It may also induce fears and anxieties, with significant and potentially enduring effects on student wellbeing, and lower their trust in education and their active learning [43, 44, 46]. Resulting behaviour changes might become long-lasting and potentially

affect other aspects of students' lives. A quantified approach to students' learning and social interactions [67] might promote obedience in classroom practices [43, 126] but reduce learning autonomy [45, 69]. AIED monitoring could also have broader implications. For example, labelling certain students as 'disengaged' [83] could perpetuate stereotypes [60] and amplify race, gender, ability, culture biases amongst teachers [68]. The compounding social injustice that results may be extensive yet subtle and require significant redress.

2.3 Existing Ethical Frameworks for AI and AIED

There has been growing ethical attention to AI in academia (e.g., [33, 56]), global institutions (e.g., [117, 127]), government departments (e.g., [10]), and AI ethics initiatives (e.g., [22]). Various frameworks (e.g., [34, 56]), guidelines (e.g., [8]), and checklists (e.g., [70]), have been developed to guide ethical practice, addressing concerns such as privacy, consent, bias, transparency, explicability and fairness [94]. Computational or data-centric issues, like bias and fairness in data and algorithms, can sometimes be measured, mitigated, or addressed through indicators or metrics. These approaches primarily focus on the technical aspects of the problem. Yet UNESCO's [117] global standards for AI ethics emphasise the "profound and dynamic" influences of AI, and advocate to consider its broader situated context, particularly in high-stakes domains such as healthcare, justice, welfare, and education [22]. To develop more effective and context-specific solutions, we need a deeper understanding of specific ethical considerations and contexts in which AI is developed and implemented. This study seeks to support these ongoing ethical discussions by focusing on the adoption of AI in educational settings.

In the field of AI in Education (AIED), the ethical focus has predominantly been on computational and data-centric issues such as bias, data privacy, and algorithm fairness [48]. Specific checklists (e.g., Greller & Drachsler's [42], DELICATE [30], SHEILA [115]) aim to address these ethical issues, particular in the context of institutional implementation and adoption processes. Ethical frameworks (e.g., EEAI framework [105], IEAIED's [112]), along with policy guidelines developed by global institutions (e.g., UNESCO [79, 116], UNICEF [119], World Economic Forum [127]), offer higher-level principles to guide the ethical use of AI technology in schools. While these frameworks are crucial, knowledge of ethical principles does not necessarily avoid real-world harm. For instance, bias and fairness are core concepts of AIED ethical frameworks [48], but AI may subtly amplify teachers' biases without their awareness.

Despite offering normative guidance on ethical AI, these frameworks tend to focus on the technical design of AI systems. This leaves a significant gap in our understanding of how these ethical issues emerge during the design and adoption stages within broader sociotechnical systems. The ethical consequences of AI adoption do not merely result from technical flaws or features, but also stem from the interactions between stakeholders, which influence how AI systems are designed, implemented and used. How these human factors contribute to ethical challenges has yet to be fully explored.

This study aims to fill this gap, by shifting the focus from examining technical issues such as AI bias and the impacts of AI implementation, to a systemic view of the design and adoption processes that shape various outcomes. It examines how AI systems are created, implemented, and governed within multi-stakeholder contexts, using an example from AIED. It seeks to expose how values embedded in this process and stakeholder dynamics can lead to unethical outcomes. In this way, the study aims to broaden existing ethical discussion about AI adoption, and to complement existing normative frameworks with a more comprehensive understanding of potential ethical failings and their implications in educational contexts.

3 METHOD

We take the case of adopting monitoring-focused AIED in K12 settings to examine ethical gaps stemming from stakeholder decision-making processes. We focus on AIED designers and developers, school and teacher representatives, and parents (as agents of their children studying in K-12 education). While students are obviously central stakeholders, our study design did not involve children under 18 years, for two reasons. First, we focused on the viewpoints of stakeholders who have input into or oversee the education system and make decisions about AIED. Second, we have specifically examined student perspectives on this technology in two previous studies [43, 44, 47][one under review]. Nonetheless, the vital interests of students come through clearly in this study.

We conducted 30 semi-structured interviews with the above stakeholders. A semi-structured approach allowed participants to bring their own experiences into the conversations [16]. We then conducted the Reflexive Thematic Analysis (RTA) approach outlined by Braun and Clarke [18, 19] to analyse data. Rather than using coding frames, RTA necessitates in-depth reflective engagement in data interpretation, informed by the researchers' values and experiences [111]. Notably, the lead author's unique experiences – being involved in industrial system design and implementation practices, and as a current graduate student – provide valuable insights into this analysis. Through iterative rounds of generating meaning from the data, we developed themes that capture patterns of shared meaning relevant to the ethical gaps and the power differentials that exist in the AIED design & adoption practices.

3.1 Participant Recruitment and Justification

We recruited 30 participants from school representatives, parents, and AIED developers, all from China. The primary reason for focusing on China is because the country has been actively pursuing, promoting and supporting the adoption of AI in education, as evidenced by its national development plans such as the “National Artificial Intelligence in Education Action Plan (2019-2022)” [80]. Unlike many other countries, numerous Chinese schools have actively adopted AIED and a wealth of AIED applications are available there for children's at-home learning. This enabled the recruitment of participants with varying AIED experiences.

We deliberately sought diversity (explained further below) when recruiting school representatives and parents, to capture a wide range of experiences and perspectives and mitigate group homogeneity and biased perspectives. We employed the principle of “information power” [72], prioritising the depth and relevance of obtained information. This approach aligns with our commitment to reflexive thematic analysis, focusing on the depth of data rather than saturation [20]. We determined that the data possessed sufficient information power after interviewing 10 people in each group. The sample specificity was informative for the research question; participants engaged in high-quality dialogues, especially the group of developers, who conveyed extensive insights from their experiences [72]. Detailed participant demographics are included in Appendix 1.

School representatives comprised 10 participants from different schools, including principals and head teachers. Participants were deliberately drawn from diverse school types, sizes, and locations, from kindergarten to high school, from rural to urban settings, and school sizes from 800 to 3000 students. This diversity offered varied insights into the potential attitudes or concerns of AIED in different contexts. Six schools where the participants work had prior AI or AIED experiences, including facial recognition security checks and customised learning resources for students. The remaining four had no previous experience with AIED.

The parent group comprised 10 participants who had children across various school levels and living in different-sized cities, allowing for socio-economic diversity. Eight participants' children had previous AIED experience, including

facial recognition attendance checks and interactive teaching agent bots, either at school or at home. The remaining two participants' children had no previous AIED experience.

The developer group comprised 10 participants, from 10 different companies, with diverse roles including IT solutionists, project and product managers, client relationship and service delivery managers, business analysts, database developers, algorithm researchers and developers. The companies ranged from medium-sized to leading AI enterprises. This diversity provided broad expertise and insights into AIED development and adoption. Three participants had AIED design and development experience and seven had only general AI experience.

In our data collection and analysis, we adopted a constructionist and experiential orientation, where we prioritised participants' own accounts of their experiences [21], especially considering their varied roles as school representatives, parents, and AIED developers. These diverse perspectives, informed by each participant's identity, and situated within their specific geographical, cultural, educational and technological environment, were critical in understanding ethical gaps in AIED.

3.2 Procedure

We conducted pilot interviews to trial our protocol and questions with three participants from three groups. In these interviews, we intentionally refrained from providing any AIED examples for classroom use, to avoid influencing participants' initial perceptions of AIED. However, three pilot participants expressed difficulties visualizing classroom AIED scenarios. Hence, to ensure participants' consistent understanding of our research context, we crafted specific scenarios. These were developed in line with our previous studies [redacted for review], which employed a speculative research method to explore student perspectives on AIED adoption, through scenario probing. We then centred on their opinions to investigate if other stakeholders – developers, parents, and school representatives - take their concerns into account. Thereby in this study, we created similar conceptual references for these three stakeholder groups. Detailed in Appendix B, these scenarios rephrased the monitoring-focused AIED artifacts (as presented in Table 1) from our literature review (section 2.2) into plain language. These scenarios served only as conceptual references to what AIED could look like for participants.

Furthermore, by framing these monitoring-focused AIED artifacts, we likely primed participants to consider specific types of AIED, particularly those akin to AI-enabled classroom performance management systems. However, it is important to clarify that our research focus was not on examining the AIED artifacts themselves, but on understanding the social factors and stakeholder dynamics that contribute to the ethical dimensions of AIED adoption. For this reason, we chose to probe stakeholder considerations using scenarios that represented seemingly more intrusive and monitoring-focused AIED applications, areas rife with ethical questions (e.g., [76]).

By doing so, we embraced the notion of “Imagaries”, a technique employed in recent HCI studies (e.g., [81]) to probe stakeholders' sociotechnical perceptions. This approach not only captures participants' knowledge and experience but also uncovers deeper socioculturally constructed perceptions, e.g., assumptions. Our focus was not to evaluate these hypothetical applications but to understand stakeholders' views and how they potentially influence AIED design and adoption. These scenarios were provided to participants at the start of the interview.

We briefed participants on the research aim and obtained informed consent for anonymous data collection. We clarified that responses from schools and developers would not represent their institution's official stance, encouraging open and comfortable expression about conflict with the company's interests.

We conducted semi-structured interviews, with each one-on-one session lasting from 37:25 to 63:03 minutes (average 43 minutes). We started by asking about related experiences in designing and developing AIED (for developers);

deploying AIED (for school staff); and if their children are using AIED (for parents). We asked developers to provide detailed information on their roles and occupations during the AI project. Then we asked participants what problems they foresee when they design or develop AI systems (developers); consider deploying or using the artifact (school staff); learn that their children's school is implementing the artifact (parents). We next asked about the precautions developers and schools want to employ before the adoption and during design and development in practice. We also asked participants with previous experience about any lessons learned from past AIED practice. Lastly, we asked developers, schools and parents about issues they would pay attention to after the launch of the service. All interviews were conducted in Chinese and were audio recorded and transcribed before analysis.

3.3 Data Analysis

We used reflexive thematic analysis to analyse our data, as it allows us to understand the overarching shared viewpoints of our stakeholder groups and the differences between groups by analysing the data as a single set. Following Braun and Clarke's [17] 6-phase guidance, the first author, a native Mandarin speaker, led the analysis.

In phase 1, the first author reviewed transcripts and generated initial ideas about each group's ethical focus and possible tensions based on all aspects of the data. In phase 2, the first author used NVivo to code interviews at the sentence level and organise the open codes inductively. Codes were either semantic, capturing participants' explicit ideas, or latent, reflecting the researcher's interpretations of underlying meanings in the data [111]. In phase 3, the first author sorted the codes into initial themes and sketched their relationships on mind maps [17]. In phase 4, the research team reviewed the mind maps and reflected on the themes through multiple rounds. In phase 5, we categorised and defined different themes of ethical gaps and stakeholders' power dynamics during the AIED adoption practices. Phase 6 then involved drafting the initial report which further revised the themes. Following multiple rounds of discussions and feedback, the first author revisited phases 3 to 6, and refined the themes. This iterative process aligns with RTA principles [19], which advocate for ongoing iterative reflexivity and ensure the analysis and interpretations are progressively recursive [18, 113].

3.4 Limitations in Context

Our study is more reflective of AIED in a policy-encouraged environment like China, which might not reflect other cultural contexts. Furthermore, perceptions of the potential harms of monitoring AI may also be influenced by culture. Investigations into other countries, policies, and cultures can help us understand variations and identify common concerns, and provide a more comprehensive understanding of ethical considerations in AI design and adoption.

4 FINDINGS

Our findings identify in detail three ethical gaps: exclusion of students' perspectives during design and adoption (4.1); developers' pragmatism due to project pressures, often limited time, and resources (4.2); and schools' limited technological literacy and failure of 'due diligence' in an ethical sense (4.3). Relatedly, we uncover 'knowledge power' and 'authority power' dynamics that promote these ethical gaps.

Below when presenting our detailed findings, we use two notations throughout the reporting. For individual participants, we label quotes from school representatives with 'S', developers with 'D', and parents with 'P' (e.g., P10 is the tenth parent). For reporting the total number of participants whose data indicate a finding, we use the "number + stakeholder group label". For instance, "Participants of all groups frequently perceived AIED systems as tools designed for teachers' use (7p, 8s, 6d)," indicates that the claim was supported by 7 parents, 8 school staff, and 6 developers.

4.1 Exclusion of Student Values, Experiences, and Needs

The first ethical gap is the systematic exclusion of students throughout the AIED design and adoption lifecycle. This starts with an initial design process that is teacher-centric (4.1.1). This focus then shifts during deployment to technical realization, often sidelining user experience (4.1.2). Finally, the cycle culminates post-deployment in coercing student compliance (4.1.3) and compromising their privacy for presumed educational gains (4.1.4). **Authority power dynamics** sit beneath this coercion and compliance, where user acceptance at this stage is meant to ensure the designed product meets user needs. This sequence widens the gap between actual and ideal user-centred design and adoption practices.

4.1.1 Design: Prioritising Teacher Needs while Ignoring Student Values and Needs

Participants from all groups frequently perceived AIED systems as tools designed for teachers' use (7p, 8s, 6d), with scant mention of students' learning experience with AIED (2p, 3s, 3d). One school representative highlighted this design emphasis on teachers, stating:

"My focus of the system design overall is the ease of operation, simplicity, and convenience, to ensure teachers' smooth adoption..." (S10)

Most participants from the three stakeholder groups shared a belief that AIED primarily serves teachers to improve teaching practices, thereby enhancing student learning. Though developers noted the merits of user-centred design, their actual practice in design workshops contrasted sharply with this ideal, lacking any mechanism for ensuring the representation of students. Meanwhile, they frequently framed students as beneficiaries but not *users* (8d):

"Decision-makers from schools are our requirement providers, direct clients and service recipients. We primarily engage with them to efficiently gather requirements... We often skipped over students to avoid inflating our workload or contract specifications. Schools' objectives typically centre on enhancing teaching quality with our system, the feedback we receive mainly stems from teachers, even though we know that students' perspectives will be very different..." (D4)

Developers revealed that the design process is driven by functional specifications reflecting user expectations and needs. This design approach frames teachers as the key end-users and schools as the decision-makers. This approach streamlines and expedites the design process for developers, quickly transitioning to development to ensure efficient project delivery. Students are mostly overlooked.

4.1.2 Deployment: Technical Focus Overshadowing Stakeholder Experience

Although developers highlighted user experience as a critical factor at various stages of the AIED design and adoption process, it was often relegated to a secondary priority during actual project execution. Below is the typical design and adoption process as outlined by and agreed upon developers:

- **Project Scoping:** Before building the business case, developers assess the product expectations of clients (schools) and define the product scope with input from sales teams, product architects, project and product managers (PMs) and IT solutionists (D1, V6).
- **Requirements Gathering and Product Design:** During delivering the business case, business analysts or PMs collect requirements from school representatives, and refine, translate, and deliver them to the technical team (D6). Here, they aim to streamline the entire user operation process and function journey as user experience considerations (D1).

- Product Development: Algorithm engineers develop models and pass them on to the development team for final product packaging; the product team oversees cost and revenue (D8). Development and deployment are seen as the most important events, as they directly determine product quality (D1).
- Product Testing and User Acceptance: The project delivery concludes with the user acceptance phase, involving the evaluation of user feedback through the User Acceptance Test (D1). Throughout the entire project, each cycle often involves iterative refinement and testing, resembling a sprint. Following the feedback collection on technical performance, the project typically proceeds with either refining or releasing the product (D7).

This evaluation process always bypassed student feedback. For example:

“The client conducts UAT. In this context, our client is the school, which has nothing to do with the user (teacher) of the monitoring system or the monitored students. The school’s Quality Assurance as a tester, usually a teacher representative, will conduct the UAT.” (D5)

Developers acknowledged that “a successful User Acceptance Test (UAT) is a critical milestone in system delivery” (D1), which focuses on client acceptance. Developers exclusively engaged teachers as primary users in assessing the technical performance against developer-prescribed goals. Ironically, when queried about who should assess the positive and negative outcomes of AIEd adoption, only the developer group said students (0p, 0s, 5d), even as they excluded students from UAT. Furthermore, schools and developers prioritised technical requirements, rarely examining user experiences. As some developers stated:

“The service delivery is primarily focused on functionality, encompassing aspects such as data representation accuracy, system continuity, and teacher-focused accessibility, rather than broader user engagement.” (D7)

“The technical team focuses on algorithm accuracy and performance, our (product team’s) main focus is to assess functionality, identify any missing scenarios, and ensure the feasibility of requirements. After implementation, we prioritise bug fixes, not specifically gathering user feedback. Users (teachers) give feedback to the client (school decision-makers), who can request second-round changes.” (D1)

This technology-centric approach risks producing systems that are functionally robust yet lack user-friendliness or responsiveness to specific user (both student and teacher) needs. Such a skewed focus can ignite user resistance, from “veteran teachers overwhelmed by the perceived complexity of the new systems” (S4), to “students feel[ing] uncomfortable with constant monitoring” (P3). Alarming, no participant group considered post-adoption impact assessments. Later we explain how developers may deflect responsibility for such evaluation (Section 4.2), while schools perceive AIEd mainly as an opportunity, often overlooking associated risks (Section 4.3).

4.1.3 Compliance: Persuading and Coercing Students into Acceptance

In our study, all three stakeholder groups expected students to comply with AIEd adoption decisions. This expectation persisted despite some participants acknowledging students’ potential resistance to monitoring (4p, 6s, 3d), and predicting adverse emotional responses (3p, 6s, 3d), behavioural changes (3p, 3s, 4d), and disrupted learning autonomy (3p, 3s, 0d). All groups suggested that guidance and explanation would mitigate student resistance (6p, 8s, 4d), emphasising anticipated learning benefits as the convincing argument. They unanimously identified persuasion as the only path to secure student ‘acceptance’ and viewed resistance as a mere transitional phase (3p, 4s, 1d), irrespective of students’ actual willingness or comfort. As a school principal explained:

“Students may initially feel uneasy and resistant under monitoring, but by explaining that data collection enhances their learning, they can feel better and gradually accept the *fait accompli* of the adoption decision. Despite an adjustment period, students will eventually overcome and become accustomed to it, as is the case with any new technology... My biggest concern is how we (schools and parents) can conduct the proactive communication to persuade effectively...” (S4)

Participants across all groups agreed that the responsibility for persuasion falls on parents and ultimately schools (6p, 8s, 4d). This delegation makes some sense since, as a developer said, “...students and parents inherently trust the schools’ authority.” (D9). We will demonstrate this parental trust in detail in section 4.3.1. Strikingly, even parents who fundamentally prioritised their children’s well-being seemed to subscribe to the view that students’ discomfort or resistance to AIEd should be managed and ultimately subdued:

“If I observe any negative mental or psychological impacts on my children, I will gently guide them towards positive well-being. Technology is neutral, its impact depends on its usage and how users perceive it. Positive mental guidance is vital... I will explain that the system is here to help.” (P10)

This coercive-compliance model underscores **authority power** over students that could lead to ethical gaps in AIEd adoption. The lack of meaningful consultation or consent from students reflects a top-down approach where students’ concerns are secondary to institutional decisions. This model both potentially endangers student well-being and overrides student autonomy. Next, we provide a demonstrative example.

4.1.4 Trading Student's Privacy for Perceived Educational Gains

All three groups mentioned privacy as a potential chief cause of students’ resistance to AIEd adoption (4p, 6s, 4d). However, around one-third of the participants, mainly parents and school representatives, viewed “schools as public spaces” (S4). They argued that students’ in-school activities are “...not private...as educational information could be surveilled for educational advancement” (P7). Disturbingly, two parents and one school principal suggested that the monitoring should be covert, to ‘protect’ students from any negative emotional responses to their loss of privacy, while simultaneously demanding detailed insights into their children’s school experiences. One parent explained:

“I hope the devices can monitor my child so discreetly that they’re unaware of its presence. This way, so they won’t feel uncomfortable being watched, we can still stay informed and intervene when necessary.” (P2)

This suggestion of covert or clandestine monitoring provides another instance of authority power, where students’ privacy is neglected in favour of adults’ desire for oversight. This approach bypasses the need for transparency and consent, further entrenching the power imbalance between students and parents and schools.

School representatives and parents generally accepted data collection practices (7p, 6s) as a necessary trade-off for improving learning. Their primary privacy concern was not the data collection itself, but rather the risk of data leakage and unauthorised data retention. Attitudes like the following were common among school representatives and parents (6p, 6s):

“Since all data is solely collected for teaching purposes, privacy is generally not a major concern... we will assure students that their data is protected. As long as the stored data remains untouched, there should be no significant worries about privacy.” (S10)

This stance not only downplays students' right to privacy but also prioritises institutional objectives over individual needs or rights. Some participants expressed resignation and pessimism regarding privacy due to cultural change (4p, 2s, 2d). They noted that "pervasive data collection and privacy intrusion are daily realities", citing examples like "... contact tracing during the pandemic and mobile payment information." (P3) Some felt cornered into acceptance, stating they "have no choice but to accept it because there is no alternative: Should we change to another school? It is impossible." (P2) Such decisions can leave parents feeling powerless and faced with a binary choice: acquiesce to the school's authority or face the impracticality of opt-out by switching schools. Previous research shows that students' resistance to data sharing varies with cultural and contextual expectations [123]. However, this resigned attitude disregards students' valuation of their privacy.

4.2 Developers' Pragmatism under Project Pressures

The next ethical gap concerns the behaviour of developers working under the pressure of completing projects. As we shall show, a critical undercurrent shaping this scenario is **knowledge-based power dynamics**. Developers' superior technical knowledge relative to school authorities positions them as de facto decision-makers in designing AIEd (4.2.1). Furthermore, a stark contradiction is seen to exist between developers' laudable ethical intentions and their actual practices (4.2.2). Once again, ethical considerations and student well-being are frequently overshadowed. The power imbalance not only marginalises students but also enables developers to limit accountability for the enduring impacts of their systems.

4.2.1 Developer's Pragmatism Facilitated by Knowledge Power

Of all the stakeholders, developers exhibited the greatest awareness and understanding of ethical issues. This comprehension covered not only traditional aspects of AI ethics like bias, transparency, user consent, and privacy but also sociotechnical implications including learning autonomy and students' psychological and cognitive status:

"In designing the system, we (the product manager and business intelligence) decide what data is presented and how it's displayed ethically. Transforming data into effective insights involves designers' ability with subjectivity. We need to avoid bias... and make assumptions that negative impacts can affect involved stakeholders. These decisions require careful reflection on the educational context and an awareness that data can both help and harm. For example, I won't display detailed individual comparisons between students and parents as it might disrupt students' learning autonomy and increase parents' anxiety from seeing detailed performances of their children...Broader implications on the educational environment also concern me..." (D5)

Developers often related delicate decision-making and responsible data utilisation and presentation (5d). For instance: "avoid displaying rankings and prediction results to students and parents to prevent negative emotions." (D1) and "anonymise other individuals if ranking is necessary" (D9). Developers proposed solutions to protect individual privacy, such as "increasing data report granularity by focusing on macro analysis" (D8). That could include "presenting the percentage of students experiencing confusion during a class for informing teaching adjustments, rather than revealing specific distressed individuals." (D4) Developers often expressed a sensitivity to students' perspectives (6d):

"Cultivating empathy is crucial for a product manager to truly understand user needs and ensure the designs are user-friendly...I always put myself in the shoes of the service receiver (students)." (D7)

Developers also frequently reported that schools possess limited technical understanding (6d). This restricted the requirements schools communicated to developers. These requirements were often general, vague concepts:

“Clients often come with a broad vision but struggle to articulate specific requirements. They may have high expectations, unaware of the technical challenges involved. We need to decode their needs, manage their expectations, explain our proposed options from a technical perspective, and guide them to reach a realistic solution...” (D7)

Developers often interpreted schools' expressed needs and assessed the feasibility of desired features (5d). They thereby exercised tacit control over the data use and representation, a point echoed by D5 earlier. Their power over AIED design was fully acknowledged (6d):

“Even when faced with impractical requests, we can steer the conversation toward a feasible decision, since our expertise gives us the leading potential in the negotiation process...But still, we serve as constructors since schools give the final word on product direction.” (D5)

The knowledge gap between developers and schools creates inherent power dynamics in AIED design and adoption, which (possibly inadvertently) puts developers in a decisive role. However, D5's quote above illustrates a disconnection: developers, despite their technical capacity and decisive influence, often viewed themselves as mere executors of schools' directives (8d) or as “builders and conveyors of technical functions, not controllers” (D3). Developers perceive school decision-makers as having decisive authority over them (7d), reflecting their “contractual agreement” (D7). While developers recognised their power in shaping data use, they interpreted this as an ability to “remind or highlight issues and suggest ethical alternatives” (D1). This discrepancy is evident even when developers confidently asserted their leadership on technical topics (4d):

“Our power and initiative vary depending on the project type: During the R&D stage, we lead with greater design autonomy; in regular non-innovative commercial projects, we act purely as executors adhering to the contracting party's directions and focusing on funding, and our authority is limited... Given schools' limited technical understanding, we guide and influence their viewpoint.” (D7)

Notably, this disconnection between developers' self-perception and their actual influential role often emerged when discussing the ethical implications of AIED features (4d):

“We are able to identify potential (ethical) issues in the given requirements. But we can only provide suggestions, within our capabilities... To what extent our suggestions are taken into account by the school decision-makers is beyond our control... We are just technology providers.” (D6)

Developers perceived this subtle power as a discretionary responsibility (6d). The exercise of this responsibility was further hampered by other pressures, which we now illustrate.

4.2.2 Developers' Ethical Sensitivity Overwhelmed by Time Pressures

Developers often eloquently articulated ethical concerns and described mitigating actions. Yet, when questioned about actual practices, their tone sometimes shifted dramatically. For example, in one conspicuous example a participant declared: “You know what, actually... I don't give a shit.” (D6) This kind of abrupt response highlights the challenging reality developers face from constraints like tight schedules. Developers prioritised meeting delivery deadlines, since

“any delay beyond the deadline could lead to financial deduction per business day” (D5). This urgency and the limited energy available for technical realisations can dampen ethical concerns. For example, despite emphasising empathy, D7 suggested that pragmatism overrides ethics:

“Completing all planned requirements within the given timeframe is already a formidable task. To be honest, we simply don’t have the luxury of time to contemplate the unintended impacts like students’ well-being after adoption...Ethical risks, as a service provider, are not within my purview. While my personal ethics and sense of social responsibility can influence the time and effort I put into ethical considerations, they aren’t seen as necessary in the workplace. We just need to get the work done and get paid under time pressure.” (D7)

Developers also acknowledged another misalignment: school representatives sometimes request AIED features that differ from frontline teachers’ needs (5d). Developers often operate under the assumption that school representatives fully comprehend teachers’ needs, even though the evidence and their own awareness suggest otherwise (4d). Developers admitted that design workshops could bridge this gap. However, this approach was often omitted due to time and cost constraints. User-centred design, while understood by developers, remained theoretical and largely unimplemented in practice:

“Direct interaction with frontline teachers to understand their needs often sidelines due to considering project timelines and human resource limitations. Instead, we mainly engage with decision-makers, assuming they comprehend teachers’ needs. Clients, especially pre-contract, are reluctant to allocate resources (teachers’ time) for us to gather information, viewing it as a waste of staff time. This is a common industry problem.” (D7)

Developers frequently disclaimed responsibility for the ethical consequences of system adoption (8d). Before a project starts, their legal teams craft contracts that explicitly define legal responsibilities, excluding any accountability for negative consequences upon adoption. During project delivery, project managers’ ethical considerations focus mainly on regulatory compliance, primarily regarding data use. After project completion, developers view their obligations as fulfilled and disengage from post-adoption impacts, relying on pre-established exemption clauses:

“Before launching any products, we conduct internal legal reviews to ensure privacy compliance, with necessary adjustments made in close collaboration between the legal audit and PM team. Sometimes clients’ legal teams also help...” (D8)

“We will seek exemptions early on and avoid taking on this accountability (for impacts of adopting AIED) since it only complicates things and might lead to financial penalties.” (D6)

Developers used legal approaches strategically to evade accountability. These contracts were designed to protect developers from additional ethical liabilities and to preserve their reputation after projects concluded.

4.3 Schools’ Limited Technological Literacy and Failure to Shoulder Ethical Burden

All three stakeholder groups believed that schools are key decision-makers and should shoulder the ethical consequences of AIED adoption (4.3.1). They all acknowledge the schools’ **authority power** over parents and students, which can mean bypassing explicit consent from these groups. However, we expose a troubling reality: schools often lack the AI literacy and associated foresight to navigate this complex responsibility and exercise due ethical diligence (4.3.2). This is another ethical gap related to **knowledge** imbalance. It is also an issue that can be magnified by reliance on reactive, trial-based AI adoption strategies (4.3.2) and a failure to anticipate potential ethical ramifications.

4.3.1 Shifted Responsibility: Developers Pass Buck to Schools; Parents Defer to Schools

All stakeholder groups held developers responsible for technical details, primarily data-related, while assigning ethical accountability to schools, since schools ultimately “approve design and adoption” (D8). After project completion, developers viewed their obligations as solely technical, effectively distancing themselves from broader ethical concerns:

“As a commercial entity, my priority is to meet my paid obligations by delivering technical solutions. Schools take charge of the ethical assessment of the system, its function, and its potential sociotechnical implications for students. Schools should have this overall consideration prior to project initiation.” (D6)

Developers expected schools to manage the broader implications of AIEd systems. Yet, as we demonstrated above, developers both felt ethically responsible for the impacts on students and recognised that schools often possessed limited knowledge of AIEd. We further elaborate on the latter in Section 4.3.2.

Parents frequently expressed trust and expected schools to ethically manage AIEd implementation and protect their children (8p):

“I trust schools’ choice of developers, systems, and data handling. They know what’s best for my kid and won’t do anything harmful. I don’t care about the details of data use; I just trust the schools’ good intentions.” (P7)

However, this parental trust was sometimes passive:

“If I don’t agree with the school’s decision (on AIEd adoption) ... I have to trust them and how they manage data.” (P8)

Parental trust, rooted in a power dynamic between schools and parents, allows schools to sidestep explicit consent requirements (4s), as this school representative observed:

“We didn’t collect user consent for the recent facial recognition attendance systems. Everyone - students, parents, and teachers, naturally complied and defaulted to this process. Especially parents obey the arrangements made by the school. They trust us with data protection and aren’t concerned about data privacy. It’s a natural progression. There is no need for explicit written consents.” (S4)

Nonetheless, some parents interpreted this acquiescence as compromise rather than agreement. For instance, parents would often defer to the majority, as they perceived their “individual dissenting voice is inconsequential” (P3). This led to “reluctant compliance with the schools’ decision” (P8). Although schools expressed respect for parent’s opinions and acknowledged their decision-making power, they also recognised that “most parents, confused with technology, tend to follow the crowd.” (S7)

Most school governors did not indicate that they would seek parent feedback. S8 mentioned that they would “observe parents’ reactions after implementation”, acknowledging that parents often defer to the school’s decision. Still, two principals from large cities (S1, S10) specifically noted the importance of obtaining parent consent to pre-empt complaints. One stated: “Parents in Beijing are often highly aware of privacy issues and know their power of complaints” (S1). The other principal said: “We would only move forward deployment after obtaining signed consents from all parents to prevent any complaints and legal issues.” (S10)

All three groups (4p, 5s, 3d) recognised the school’s authority over parents and students. Nonetheless, while parents’ authority is recognised in theory, it can be undermined in practice. These dynamics prompt critical ethical questions

about the nature of consent and trust within education, particularly given the power schools have over parental rights and student autonomy.

4.3.2 Limited AI Acumen in Schools: Exacerbating Ethical Blind Spots

Schools frequently perceived AI as an opportunity, attracted by its perceived benefits (6s), while underestimating its potential pitfalls. They demonstrated limited ability to recognise and anticipate possible negative implications of AIED and often admitted that they would address ethical issues once they surface (3s). However, education is intrinsically complex, and addressing ethical concerns only after trials means that students may already have been impacted by potentially difficult-to-reverse harms on the learning and wellbeing.

Schools' limited understanding of AI and its potential impact manifested in viewing issues such as privacy, fairness, and equity as arising from human misuse of data, particularly teachers' misconduct (6s). Parents also exhibited relatively little awareness of the AI's technical flaws, thinking of the concerns only relevant to teachers' misuse:

"My concern is that teachers shouldn't over-rely on automatic insights without their own thinking." (S2)

"My biggest privacy concern is teachers' access to student information. This needs transparency." (P4)

Given the limited awareness of ethical issues, parents and schools placed responsibility for these directly on teachers, proposing training to avoid data breaches and misuse as mitigation. Even within the focus on data misuse, there were conflicting perspectives among parents and schools. Parents were worried about teachers' misusing AIED tools in ways that exacerbate existing unfairness (3p). They believed that, with an inherent focus on academic KPIs, schools and teachers might disproportionately favour high-scoring students who have higher chances of getting into top universities. Consequently, teachers could exploit AI data insights to amplify pre-existing biases. As one parent noted:

"Predicting students' failure can induce biases among teachers. They might then put more attention on high-achieving students, skew resource allocation, and thereby worsen educational inequality." (P8)

This sentiment is seemingly validated by a teacher representative's statement on how they can use AI insights to cater to specific student groups:

"The schools' objective is to boost college admission rates by focusing on high-achievers. It will be vital for us if AIED helps identify areas these 'seed' students struggle with. We used to conduct regular one-on-one interviews with those selected students and their families, to understand their learning pain points, so we could provide tailored assistance. I believe AIED customised analysis can save lots of energy for us." (S4).

These differing perspectives from P8 and S4 reveal the conflicting values of different stakeholders, demonstrating a lack of consensus. This value tension highlights the power dynamics within this multi-user system, which could lead to inequality in the treatment of students. Furthermore, as demonstrated in section 4.1, the teacher-centred AIED design process risks an uneven benefit and value distribution within the system, deepening ethical disparities.

4.4 Summary of Findings

Our findings reveal three significant ethical gaps, each driven by underlying power dynamics, that could hinder the ethical AIED adoption:

Gap 1 - Excluding Students. Teacher-centric design and technical-focused deployment overlooks students' needs, wellbeing, and experiences. The post-deployment primarily enforces compliance rather than fostering genuine

acceptance by students. The **authority power dynamics** between students, parents, and schools facilitate this systematic exclusion of students in AI adoption.

Gap 2 - Developer's Pragmatism. Developers often prioritise efficiency and financial gains over ethical considerations. Their extensive technical knowledge and ethical awareness grant developers **knowledge power** with critical influence over AIED design and deployment. However, their pragmatic focus disconnects their actual practices from their ethical intentions or ideals. They view themselves as executors rather than decision-makers. This perception allows them to shift responsibility for ethical outcomes to schools, and avoid full accountability, shielded by legal frameworks.

Gap 3 - School's Failure of 'Due Diligence'. Schools, as primary AI decision-makers, lack the necessary AI literacy to handle ethical implications effectively. They focus on perceived benefits while underestimating risks and relying on reactive solutions to emerging issues. The trust parents place in schools due to schools' **authority power** allows for AI adoption without explicit consent. However, schools may fail to be fully accountable due to limited ethical foresight.

Our study found that key ethical responsibilities within this system are not effectively understood or shouldered by key stakeholders. This absence of effective responsibility, while sometimes understandable, exacerbates the ethical risks.

5 DISCUSSION

Our findings highlight crucial gaps in stakeholder inclusion, ethical awareness, responsibility and capability in AI development and adoption. In Section 5.1, we discuss how these ethical gaps contribute to ineffective accountability mechanisms, which fail to mitigate possible harmful outcomes for users—especially students—in the context of AIED. In Section 5.2, we discuss factors associated with power dynamics that could be avoided to mitigate ethical challenges. Finally, in Section 5.3, we use an HCI lens to unpack ethical complexities in ethical AI design and adoption applicable across multi-stakeholder contexts.

5.1 Ethical Gaps and Inadequate Accountability Mechanisms

Recent HCI research has identified potential impacts on students resulting from their interaction with data-driven classroom behaviour management systems [58]. Beyond established privacy concerns [58], classroom AI monitoring can disrupt learning and impact students' well-being [43], including their individual identity and social interactions [44]. Indeed, monitoring-focused AI may create a new educational panopticon [43], triggering possibly longer-term changes to students' learning behaviour. Fearful of being constantly monitored and assessed, students may behave artificially to satisfy idealised learning standards [43], in turn promoting performative and score-oriented learning [45]. These negative experiences can accrue over time and have an enduring impact on student development, such as eroding the social relationships and autonomy [124] required for growth and critical thinking, ultimately diminishing the educational journey. Despite these potential impacts, they are often overlooked by powerful stakeholders in the AIED design and adoption process in our study. We have highlighted how student values and experiences are frequently disregarded (Gap 1), which can contribute to the occurrence of these adverse effects.

Powerful stakeholders in our study, particularly parents and schools—two groups that usually prioritise students' well-being—do acknowledge students' potential resistance to AI monitoring. These groups do not, however, always take student concerns seriously. Instead, they frame them as obstacles to be overcome or managed rather than ethical issues requiring deeper reflection. HCI research has highlighted students' potential emotional and behavioural changes [43] and has called for studies into long-term effects. However, stakeholders in our study largely view these changes as a natural, transitional discomfort that students will eventually acclimatise to. In fairness, some related HCI studies, such as in-home monitoring, have demonstrated that individuals adapt surprisingly quickly to even significantly intrusive

monitoring while acknowledging that negative feelings do surface initially during adoption [87]. It is possible that students will also adapt better than we might think to monitoring-focused AIED. However, the critical question remains whether the possible behavioural adaptations or changes in students genuinely align with the values that we seek to uphold in the design of AIED and education itself, and whether these effects on students truly are only transient and minor. Importantly, schools and parents often perceive their “protective” role as guiding students to accept AI by overcoming emotional discomfort during AIED deployments, viewing enhanced learning as a worthwhile trade-off for, say, lost privacy. Yet this stance ignores possible long-lasting and complex social and educational consequences. This issue results from the lack of literacy about ethical protection and scrutiny where AI is involved (Gap 3).

The limited understanding of ethical risks among parents and schools not only hinders the ethical use of AIED but also its governance. Parents and schools in our study often attribute bias, unfairness, and privacy risks solely to teachers’ misuse of data. They thus fail to acknowledge the inherent risks posed by AI models to student wellbeing, identity, and learning. This insufficient AI literacy among stakeholders with protective roles exacerbates the ethical risks, especially in terms of determining who is responsible and accountable for safeguarding students from harms like bias and unfairness. HCI research, with its focus on human factors in technology design and ethical usage, has highlighted how ethical gaps are tied to how humans perceive ethical issues [68]. This underscores the urgent need for a more holistic understanding of AIED’s ethical challenges—one that extends beyond technical concerns to the broader governance structures that oversee these practices.

Our study reveals significant gaps in addressing ethical concerns post-AI adoption, and the absence of suitable accountability mechanisms among stakeholder groups. This finding resonates with previous research that highlights the reactive approach of industrial organisations to ethical concerns in AI systems [95] and the lack of accountability within these organisations, where developers primarily focus on technical details [125]. Our findings also echo these challenges: developers are aware of ethical issues, but their actions are hindered by their roles or organisations’ profit-driven focus [26, 41, 95]. However, our research situates these tensions within a broader ecosystem. We demonstrate that the accountability gap is not limited to developers and their organisations, since other key stakeholders, such as parents and school representatives, are also unable to ensure accountability. This lack of accountability deepens the ethical risks, as no party is truly responsible for potential harms.

We also reveal how developers’ knowledge power influences their actions and interactions with clients (schools), adding nuance to existing research on developers-client dynamics. This power imbalance is another critical factor in the accountability gap, as developers’ technical expertise often positions them in decisive roles in system design, while they sideline their ethical responsibility. Previous studies have discussed the constraints developers face due to clients’ expectations during system design and development processes (e.g., [41]). In our study, we reframe these product expectations as part of the prevailing power dynamics, specifically as “perceived authority”, stemming from the legal contracts developers are bound by. We further reveal the knowledge power developers have over clients, where developers’ expertise implicitly gives them influence over decision-making. This insight extends the conversation about developers’ constraints, by showing that the power is not just one-directional. These dynamics complicate the ethical landscape and decision-making process, ultimately shaping the ethical outcome of AI design and adoption. We will explore these power dynamics in more detail in the next subsection.

Our findings reveal an assumed division of labour between stakeholders that should provide comprehensive ethical protection. All three groups in our study assign technical responsibility for data security to developers, while schools handle administrative duties like preventing data leaks from human errors. Yet this arrangement neglects broader ethical issues. As highlighted in Section 2.2, prevailing AIED governance frameworks focus primarily on data management and

privacy (e.g. [58, 102]). Developers, prioritise adherence to these narrow guidelines and legal or organisational policies as a framework for ethical decision-making [26]. As seen in our study, they perceive their obligations as confined to data-related compliance (resonate to [95]), thus evading broader ethical responsibility. Schools, on the other hand, lack the expertise to proactively tackle non-data-related ethical issues. This, of course, is not new: prior research highlights educational institutions' limited ethical awareness, resources, and capabilities in acknowledging and evaluating the ethical implications of adopting educational technologies [5, 114]. When these concerns are not addressed in planning, the ethical risks of AIED adoption are heightened.

This accountability gap mirrors the 'Swiss Cheese Model' [96], where multiple defence layers (cheese slices) have holes representing vulnerabilities, and system failures occur when all holes align. In AIED, these aligned vulnerabilities – developers avoiding post-adoption responsibility (Gap 2), schools' limited expertise in ethical scrutiny (Gap 3), and an overly narrow focus on data management by schools, parents, and developers (Gaps 2, 3) – culminate in potent ethical failures. Across a sector, these recurring ethical risks may well lead to significant negative outcomes for the most vulnerable stakeholders and intended beneficiaries of the technology, namely students.

5.2 Addressing Ethical Challenges Created by Power Dynamics

We have demonstrated the power dynamics that can engender ethical gaps. Power imbalances are, of course, not always problematic and sometimes they are desirable for protecting interests and rights, such as when an accountable and knowledgeable stakeholder has the power to decide whether to implement or reject a technology for the protection of vulnerable groups. Even so, decisions must sometimes be navigated carefully to avoid ethical pitfalls. Furthermore, some factors can cause ethical problems which may be relatively easily avoidable when certain conditions are enacted. This is especially true, we would suggest, of the knowledge power imbalance between AI developers and schools (See in purple in Fig.2). Schools' lack of AI literacy leads to a failure to consult, ask questions, and source the right knowledge to assist ethical decision-making.

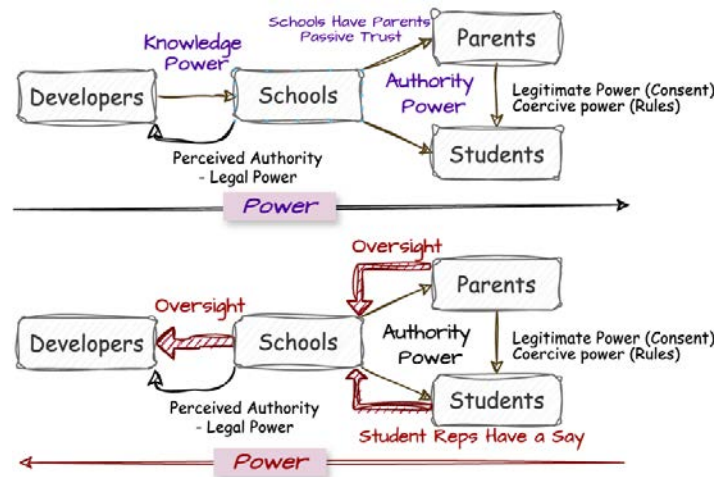


Fig.2 Power dynamics that may facilitate ethical gaps in AIED adoption and ways of addressing them.

Our data highlights how power dynamics contribute to problematic accountability mechanisms in AIED design and adoption decision-making. Fig.2 summarises our discussion of power dynamics underpinning ethical gaps and proposed

strategies for mitigating or avoiding them. We contend that the practical and effective stewardship of AIEd [79] [86] requires key decision-makers (e.g., schools, teachers, parents) to be well-informed and, especially in the case of schools, capable of critically assessing AI adoption. This involves not just passive oversight but proactive, knowledgeable engagement in making ethical decisions [65, 118]. Meanwhile, developers should be encouraged to broaden their ethical responsibility beyond data-centric issues to include anticipation of long-term impacts on students and their educational experience post-launch. They could then raise these concerns with schools who are contracting the technologies from them. Furthermore, students and their representatives should be granted a genuine say, perhaps including the power to veto the school's decisions on AIEd adoption. In the following section, we will discuss these specific strategies for addressing the ethical challenges and power imbalances.

Empower School Governors for Effective Stewardship: Our findings highlight the need to empower school decision-makers or educational leaders with the knowledge to understand the possible effects of AIEd. They must be better equipped to make informed ethical decisions and more effectively steward the design and deployment of AIEd [79, 86]. This approach helps to mitigate the knowledge power imbalance between them and developers (Seen in red in Fig.2). This shift in power, in turn, addresses the accountability gap and reduces ethical challenges. Various tools can assist in this, including AI ethics education, ethical adoption guidelines (e.g., [84]), checklists (e.g., [112]), or conceptual thinking tools (e.g., [105]), to empower school decision-makers to make sound decisions on design, procurement, and the use of AI, while also contemplating, questioning, and addressing the ethical implications of adopting AI in schools.

Support Practitioners for Undertaking Ethical Actions: Our study, along with previous research [95], reveals that practitioners' awareness and actions are often hindered by organisational pressures, such as tight project timelines and profit-driven goals. To mitigate these challenges, organisations must cultivate a supportive internal culture and provide structural support for individual practitioners, enabling them to navigate ethical complexities more effectively [26, 78]. This includes fostering organisational cultures that prioritise ethical decision-making and integrating ethical reasoning into individual professional identities [26]. Organisational structures should redefine project success metrics to include societal and ethical considerations, ensuring alignment between organisational missions and individual practices [95]. Additionally, tools should be developed to help practitioners anticipate risks and measure AI systems' impact, shifting from reactive to anticipatory approaches [95]. To mitigate fragmented accountability within organisations, developers should communicate across project units, and critically reflect on their roles and responsibilities within the project supply chain [125]. We also suggest practitioners make transparent any risks of the system to their clients. Such transparency can help to shift power dynamics, directly involving teachers and schools in the accountability process [68, 85]. Finally, a more integrated approach to ethics in AI is needed within the organisation to influence practitioners' ethical behaviour, self-reflection and decision-making. This approach should include formal education, industry standards, and professional codes [41]. Ethical decision-making of practitioners involves a dynamic interplay of personal, organisational, and applied ethical influences [41], and these domains must be effectively integrated to support practitioners in making sound ethical decisions.

Empower Parents for Active Oversight: While authority dynamics existing between schools, parents, and students are partly unavoidable, steps can be taken to mitigate ethical gaps arising from this power imbalance. As detailed in section 4.3.1, representatives from two schools (S1, S10) recognised the empowerment of parents through complaint mechanisms, even though this is limited by their inadequate AI literacy and a tendency to uncritically accept schools' decisions. These school representatives underscored the critical role of obtaining parental consent to pre-empt complaints. One mitigation strategy, then, is to transform parents' passive trust into more active oversight (seen in red in Fig.2). This requires schools to engage in open and clear communication about potential AI implications, which is of course contingent upon the

schools’ own comprehensive understanding of these consequences. Such knowledge and transparency in the consent process not only informs parents about potential risks but also involves them directly in accountability processes.

Empower Students’ Voices: It is widely recognised in HCI that best practices involve the participation of all stakeholders which, in this case, includes students. However, the structural power imbalance between students, teachers, and school decision-makers, coupled with vendors’ pragmatism under organisational pressures, often makes meaningful student involvement in design processes difficult. We suggest involving student representatives in the adoption communication process (see Fig. 2). Students and their representatives should be granted a genuine say or the power to veto their school’s decisions regarding AIEd adoption. As previous studies have argued, maybe one way to ensure that participation in design is meaningful and that power dynamics are actively addressed is by granting stakeholders – especially marginalised ones – the ability to veto projects that fail to align with their needs and values [71, 95].

In addition to efforts aimed at assisting practitioners, equipping schools and parents, and granting properly informed students veto power, external processes should play a role in addressing ethical challenges. External groups, such as advocacy organisations or independent oversight bodies can often apply pressure more freely than internal stakeholders, who may hesitate to speak up due to concerns about potential repercussions [95]. Building coalitions and creating communication channels between all relevant parties – practitioners, companies, teachers, students, school decision-makers, parents, as well as external groups – can help increase momentum to address ethical concerns. This collaborative approach helps ensure that ethical issues are not overlooked and that a broader range of voices is involved in making decisions about AIEd adoption and use.

While we have focused on AIEd, the same principles can be employed by HCI scholars and practitioners to facilitate ethical AI implementation in other sensitive contexts. We will discuss this around our findings in the next subsection.

5.3 Using an HCI Lens to Unpack Ethical Complexities for Multi-User AI Adoption

The HCI community, with its commitment to equity, inclusivity, and diversity, has laid the groundwork for promoting ethical literacy [65, 66] through initiatives such as CHIiversity [108] and CHIInclusion [109]. There is a pressing need for HCI to assist designers in recognising and navigating the various ethical tensions and trade-offs that impact less-powerful groups within multi-user systems. Designers’ commitment to reflexivity is crucial for effectively embedding ethical considerations into systems design [12, 32, 35, 36]. This section explores how HCI could tackle value tensions and power dynamics (Section 5.3.1), scrutinise the authenticity of technology acceptance in hierarchical contexts (Section 5.3.2), and prioritise vulnerable users in multi-user AI environments (Section 5.3.3).

5.3.1 Navigating Value Tensions and Power Dynamics

HCI research constantly highlights ethical complexities arising from value tensions and differing ethical perceptions among stakeholders. A study on home-entryway surveillance [120], for example, reveals conflicting privacy perceptions between parents and teens [27]. Such a tension is also seen in our study: contrary to the school’s and parents’ perception of classrooms as public spaces, our prior study [43] and other studies (e.g., [99]) show that students find such monitoring intrusive and disruptive to their privacy and learning, and are unwilling to share learning-related emotions with teachers or AIEd. In contrast, parents in our study prefer AIEd to report detailed information about their children’s psychological well-being at school, citing a need for their care. These disparities are often observed in child-parent relationships [39, 61] alongside the tension between protection and privacy invasion. Notably, notions of ‘care’, while ostensibly positive, are often seen in technology-facilitated misuses or abuses (e.g., [61, 73]) that harm vulnerable groups.

Schools and parents generally agree to compromise some student privacy in exchange for perceived educational gains. This privacy trade-off mirrors findings in elder care where safety is prioritised and less-disruptive technologies are more readily accepted [128]. However, key differences exist: Unlike competent older adults who can (ideally) choose to accept or reject privacy trade-offs, students in unequal power relations often lack effective agency. Additionally, younger students may not fully comprehend the ramifications of sharing personal information, nor possess the capacity to provide informed consent that parents give on their behalf. These trade-offs call for nuanced ethical consideration.

We lack research dedicated to addressing the tensions among stakeholders – both within AI in education, and, as has been noted by HCI scholars, in AI in other high-stakes sectors, e.g., [25, 70, 88]. Our study contributes to this ongoing discussion by illuminating these complexities in the educational context and providing insights to navigate ethical challenges. It shows how crucial it can be to understand the intricate networks of relationships, power differentials, and value tensions in technology adoption that can generate ethical risks.

5.3.2 Questioning the Authenticity of Acceptance for Disempowered Groups

Genuine acceptance, rather than mere acquiescence, is considered vital in HCI. Authentic acceptance is closely tied to user autonomy and informed consent [99]. While the concept of authentic acceptance has been discussed across domains like online exam proctoring [110] and pervasive healthcare interventions [103, 133], its application in AIEd reveals significant deviations. For instance, the Theoretical Framework for Acceptance, derived from the health technology community [103], identifies the affective attitude – the users’ feelings towards and impacts from the technology – as a primary attribute of acceptance. However, in AIEd, students’ negative emotions and behavioural changes can be, as we illustrated, dismissed as a natural part of a beneficial technological transition. Our three stakeholder groups expected students to ultimately accept AIEd and they aimed to alleviate student concerns through persuasion about its benefits. This gradual acceptance of technological intrusiveness, while possibly justified in some cases, may sometimes involve introducing marginally acceptable yet beneficial technologies via a ‘foot in the door’ approach [91] – an approach well recognised in HCI. In our study on AIEd, all three groups endorsed the established authority power of schools and parents over students. This power dynamic raises thus critical questions about the authenticity of any acceptance among vulnerable student users.

5.3.3 Prioritising Vulnerable Users in Designing Multi-user AI Systems

While HCI communities advocate for human-centred design and equal stakeholder engagement in technology design (e.g., [122]), existing studies show a design imbalance in multi-user AI systems particularly where power differentials exist. For instance, AI-enabled HR tools [88] often reflect the policy designers’ values, overlooking the employees directly affected by the technology. Similarly, our AIEd findings reveal teacher-centric design (section 2.1) aimed at enhancing teachers’ roles (e.g., [9, 50]). While this approach can sometimes benefit students, it could also relegate them to marginalised and indirect secondary users.

We observe a pattern behind this imbalanced design focus in multi-user contexts. Previous research introduced the Computer-Human-Human-Interaction (CHHI) model for call centre systems [106, 107], delineating both direct and indirect user interactions among customers, agents, and computers. In the AIEd context, this model manifests as AI-Teacher-Student-Interaction. Teachers, who receive AI insights to support students, are considered primary users, while the students, who do not directly interact with AI, are seen as ultimate-end-users – mere passive recipients of AIEd’s benefits. However, this overlooks the implicit AI-student interaction that occurs in the AI monitoring and evaluation process, directly impacting students by potentially changing their behaviours [43] and affecting their wellbeing.

HCI has consistently emphasised balanced design partnerships, particularly in existing power structures (e.g., [75, 131]), advocating for including and representing all potential users (e.g., [77]), especially those most impacted and less powerful (e.g., [88, 98]). Novel approaches could take cues from the healthcare sector—another high-stakes, multi-user context with its own power dynamics. Healthcare has found a way to emphasise patient-centric rather than doctor-centred design of services and technology, with patients acknowledged as key stakeholders and beneficiaries [121]. Somewhat analogously, our AIED case highlights a need to shift from teacher-centric to student-centric design, prioritising students’ learning journey and values and thereby offering insights applicable across other multi-user contexts. In parallel, co-design studies underscore emerging adult’s and children’s demand for self-autonomy and for being treated as equals in their digital experiences [122]. These lessons suggest that in contexts like AIED, system designers must go beyond mere data regulation compliance (Gap 2) to genuinely empower users, particularly the vulnerable, to proactively engage in design practices (e.g., [101, 122]).

6 LIMITATIONS AND FUTURE WORK

Although our previous work on AIED has included student participants, this study focused on existing decision-makers. Yet, as we have stressed, student voices are too easily neglected. Accordingly, future research should explore methods for increasing the effective inclusion of less empowered groups like students in technology development and deployment.

Given the power dynamics present in AIED design and adoption, we have proposed approaches to potentially mitigate these dynamics, such as by improving the AI literacy of responsible decision-makers (e.g., parents and school governors), and granting students power to veto adoption decisions. Future research should explore these strategies for balancing power relationships and assess whether these actions change the dynamics to foster ethically sound practices.

While our use of interviews provided valuable insights into stakeholders’ concerns and tensions, it may not fully capture the potential negotiations and dynamics that emerge in direct interaction. To address this, future research could incorporate participatory methodologies such as focus groups or stakeholder workshops, bringing together students, teachers, developers, and school governors into dialogue. This approach would help assess whether conflicting considerations can be resolved through dialogue and whether such engagement influences each group’s ethical stance on AIED adoption.

7 CONCLUSION

The HCI community is committed to addressing social and ethical factors in technology design and deployment that generate harm and injustice. Our study contributes to this larger endeavour, by highlighting the complex ethical challenges arising from power dynamics in the design, adoption, and governance of AI monitoring and of students in education (AIED). We identify three critical ethical risks: the exclusion of vulnerable groups (students), the pragmatism of developers working under various pressures, and the limited technological literacy of customers (schools) which causes them to fail to diligently address various risks. Those issues are tied to two sources of power — knowledge and authority — that are found among stakeholders, creating a vacuum of accountability and leaving vulnerable users — students — to ethical harms.

This study provides empirical evidence that expands understanding of the ethical challenges in AI design, deployment, adoption, governance and use. It argues that ethical considerations must extend beyond technical aspects of AI systems to address broader socio-technical or systemic challenges where power dynamics are at play. By advocating for a more accountable approach to ethical decision-making within the AIED design-making ecosystem, we stress the importance of empowering all stakeholders in AIED adoption: schools for effective stewardship, practitioners for ethical actions,

parents for active oversight, and students with a say and veto power. This includes fostering AI literacy, enabling transparency, and creating channels for communication and active participation, particularly for vulnerable groups like students, to mitigate power imbalances and strengthen accountability mechanisms across all levels of AIED deployment.

These results and discussions plausibly extend beyond AI in education to many other sensitive sectors adopting such technologies, such as digital health and online public services. In closing, we not only call for reassessing current design and adoption practices in education and other contexts, but also advocate for confronting power imbalances that lead to negative ethical consequences. Our aim is to catalyse changes that empower vulnerable groups, promote social justice, and establish comprehensive ethical protections in AI development and deployment.

ACKNOWLEDGEMENT

This research was funded through the Australian Government's Research and Training program.

APPENDIX A

Appendix A: Participant Information

School	School Type	School Size	School Located City Size	AIEd Experience	Gender
S1	Primary and High	300k	Large	Yes	Male
S2	Primary	800	Small	No	Male
S3	Kindergarten, Primary	2.8k	Small	Yes	Female
S4	Junior High	15k	Medium	Yes	Male
S5	High	3k	Small	No	Male
S6	Junior High	3k	Medium	Yes	Male
S7	High	800	Large	Yes	Female
S8	Primary	1.3k	Rural Area	No	Male
S9	High	2k	Large	Yes	Male
S10	Primary	1.7k	Medium	No	Male

Parent	School Type	City Size	AIEd Experience	Gender
P1	Primary	Large	Yes	Female
P2	Kindergarten	Small	No	Male
P3	Primary	Small	Yes	Male
P4	Primary	Rural Area	Yes	Female
P5	Junior High	Large	Yes	Female
P6	Primary	Medium	Yes	Female
P7	High	Small	Yes	Female
P8	Junior High	Large	Yes	Male
P9	High	Small	No	Female
P10	High	Large	Yes	Female

Developer	Roles	AIEd Experience	Gender
D1	IT Solutionist	No	Female
D2	Product Manager	Yes	Female
D3	Database Developer	No	Male
D4	Project Manager	No	Female
D5	IT Solutionist	No	Female
D6	Service Delivery Manager	Yes	Male
D7	Business Analyst, Client Manager	Yes	Male
D8	AI Algorithm Researcher & Developer	Yes	Male
D9	Product Manager – EdTech Specialist	Yes	Female
D10	AI Algorithm Developer	Yes	Male

APPENDIX B

Examples of AI-enabled classroom performance management systems provided during interviews:

These systems employ advanced facial and emotion recognition technologies to enhance classroom interactions. Key functionalities include:

Attendance and Engagement Observation: The system tracks student attendance and gauges attention levels during lessons, using facial recognition to identify instances where students may be distracted or miss crucial learning content.

Real-Time Support Identification: In instances where a student's facial expressions indicate confusion or struggle with the material, the system can alert teachers. This enables timely intervention during the lesson.

Post-Class Analysis and Reporting: After classes, the system generates nuanced feedback, pinpointing specific chapters or concepts where a student may need more practice. This personalised approach helps in identifying areas of weakness for each student.

Academic Risk Prediction and Alerts: By analysing a student's performance in real-time, the system can predict potential academic challenges in various subjects. It proactively notifies teachers and parents, allowing them to intervene before a student's performance declines significantly.

Overall, these AI-enabled systems are designed to not only observe classroom dynamics but also to provide actionable insights, supporting personalised learning experiences and enhancing academic outcomes.

REFERENCES

- [1] Karan Ahuja, Dohyun Kim, Franceska Xhakaj, Virag Varga, Anne Xie, Stanley Zhang, Jay Eric Townsend, Chris Harrison, Amy Ogan and Yuvraj Agarwal. 2019. EduSense: Practical classroom sensing at Scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3, 1-26.
- [2] Karan Ahuja, Deval Shah, Sujeeth Paredy, Franceska Xhakaj, Amy Ogan, Yuvraj Agarwal and Chris Harrison. 2021. Classroom Digital Twins with Instrumentation-Free Gaze Tracking. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama, Japan, Article 484. <http://dx.doi.org/10.1145/3411764.3445711>
- [3] Daniela Altun, Christopher Krauss, Alexander Streicher, Christoph Mueller, Daniel Atorf, Lisa Rerhaye and Dietmar Kunde. 2022. Lessons Learned from Creating, Implementing and Evaluating Assisted E-Learning Incorporating Adaptivity, Recommendations and Learning Analytics. In *International Conference on Human-Computer Interaction*. Springer, 257-270.
- [4] Nese Alyuz, Eda Okur, Utku Genc, Sinem Aslan, Cagri Tanriover and Asli Arslan Esme. 2017. An unobtrusive and multimodal approach for behavioral engagement detection of students. In *Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education*. Association for Computing Machinery, Glasgow, UK, 26-32. <http://dx.doi.org/10.1145/3139513.3139521>
- [5] Asma Shannan Alzahrani, Yi-Shan Tsai, Sehrish Iqbal, Pedro Manuel Moreno Marcos, Maren Scheffel, Hendrik Drachsler, Carlos Delgado Kloos, Naif Aljohani and Dragan Gasevic. 2023. Untangling connections between challenges in the adoption of learning analytics in higher education. *Education and Information Technologies* 28, 4, 4563-4595.
- [6] Dana Alzoubi, Jameel Kelley, Evrim Baran, Stephen B. Gilbert, Aliye Karabulut Ilgu and Shan Jiang. 2021. TeachActive Feedback Dashboard: Using Automated Classroom Analytics to Visualize Pedagogical Strategies at a Glance. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama, Japan, Article 312. <http://dx.doi.org/10.1145/3411763.3451709>
- [7] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett and Kori Inkpen. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1-13.
- [8] Mona Ashok, Rohit Madan, Anton Joha and Uthayasankar Sivarajah. 2022. Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management* 62, 102433. <http://dx.doi.org/https://doi.org/10.1016/j.ijinfomgt.2021.102433>
- [9] Sinem Aslan, Nese Alyuz, Cagri Tanriover, Sinem E Mete, Eda Okur, Sidney K D'Mello and Asli Arslan Esme. 2019. Investigating the impact of a real-time, multimodal student engagement analytics technology in authentic classrooms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1-12.
- [10] Australian-Government. Australia's AI Ethics Principles. Retrieved from <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles>.
- [11] Ryan S Baker and Aaron Hawn. 2022. Algorithmic bias in education. *International Journal of Artificial Intelligence in Education* 32, 4, 1052-1092.
- [12] Shaowen Bardzell, Jeffrey Bardzell, Jodi Forlizzi, John Zimmerman and John Antanitis. 2012. Critical design and critical theory: the challenge of designing for provocation. In *Proceedings of the Designing Interactive Systems Conference*. 288-297.
- [13] BBC. 2018. Amazon scrapped 'sexist AI' tool. Retrieved from <https://www.bbc.com/news/technology-45809919>.
- [14] Haydn Belfield. 2020. Activism by the AI community: Analysing recent achievements and future prospects. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. 15-21.
- [15] Aditi Bhutoria. 2022. Personalized education and artificial intelligence in United States, China, and India: A systematic Review using a Human-In-The-Loop model. *Computers and Education: Artificial Intelligence*, 100068.

- [16] Ann Blandford, Dominic Furniss and Stephann Makri. 2016. Qualitative HCI research: Going behind the scenes. *Synthesis lectures on human-centered informatics* 9, 1, 1-115.
- [17] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2, 77-101.
- [18] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4, 589-597. <http://dx.doi.org/10.1080/2159676x.2019.1628806>
- [19] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology* 18, 3, 328-352. <http://dx.doi.org/10.1080/14780887.2020.1769238>
- [20] Virginia Braun and Victoria Clarke. 2021. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health* 13, 2, 201-216. <http://dx.doi.org/10.1080/2159676x.2019.1704846>
- [21] David Byrne. 2022. A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & quantity* 56, 3, 1391-1412.
- [22] Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker and Kate Crawford. 2017. AI now 2017 report. 2017.
- [23] Jake Chanenson, Brandon Sloane, Navaneeth Rajan, Amy Morril, Jason Chee, Danny Yuxing Huang and Marshini Chetty. 2023. Uncovering Privacy and Security Challenges In K-12 Schools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-28.
- [24] Zhilong Chen, Jinghua Piao, Xiaochong Lan, Hancheng Cao, Chen Gao, Zhicong Lu and Yong Li. 2022. Practitioners Versus Users: A Value-Sensitive Evaluation of Current Industrial Recommender System Design. In *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 533 (November 2022), 32 pages. .
- [25] Hao-Fei Cheng, Logan Stapleton, Ruiqi Wang, Paige Bullock, Alexandra Chouldechova, Zhiwei Steven Wu and Haiyi Zhu. 2021. Soliciting Stakeholders' Fairness Notions in Child Maltreatment Predictive Systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama, Japan, Article 390. <http://dx.doi.org/10.1145/3411764.3445308>
- [26] Shruthi Sai Chivukula, Aiza Hasib, Ziqing Li, Jingle Chen and Colin M Gray. 2021. Identity claims that underlie ethical awareness and action. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1-13.
- [27] Lorrie Faith Cranor, Adam L. Durity, Abigail Marsh and Blase Ur. 2014. Parents' and Teens' Perspectives on Privacy In a Technology-Filled World. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. 19-35.
- [28] Yngve Dahl, Jarl Kåre Reitan and Anita Das. 2018. Value tensions in telecare: an explorative case study. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*. 559-570.
- [29] Yngve Dahl and Dag Svanæs. 2020. Facilitating Democracy: Concerns from Participatory Design with Asymmetric Stakeholder Relations in Health Care. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Honolulu, HI, USA, 1-13. <http://dx.doi.org/10.1145/3313831.3376805>
- [30] Hendrik Drachler and Wolfgang Grellner. 2016. Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*. Association for Computing Machinery, Edinburgh, United Kingdom, 89-98. <http://dx.doi.org/10.1145/2883851.2883893>
- [31] Mohamed Ez-Zaouia, Aurélien Tabard and Elise Lavoué. 2020. EMODASH: A dashboard supporting retrospective awareness of emotions in online learning. *International Journal of Human-Computer Studies* 139, 102411.
- [32] Mary Flanagan, Daniel C Howe and Helen Nissenbaum. 2005. Values at play: Design tradeoffs in socially-oriented game design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 751-760.
- [33] Luciano Floridi and Josh Cowls. 2022. A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design*, 535-545.
- [34] Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo and Francesca Rossi. 2021. An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Ethics, governance, and policies in artificial intelligence*, 19-39.
- [35] Sarah Fox, Jill Dimond, Lilly Irani, Tad Hirsch, Michael Muller and Shaowen Bardzell. 2017. Social Justice and Design: Power and oppression in collaborative systems. In *Companion of the 2017 acm conference on computer supported cooperative work and social computing*. 117-122.
- [36] Sarah Fox, Catherine Lim, Tad Hirsch and Daniela K Rosner. 2020. Accounting for design activism: On the positionality and politics of designerly intervention. *Design Issues* 36, 1, 5-18.
- [37] Rachael Garrett, Kristina Popova, Claudia Núñez-Pacheco, Thórhildur Ásgeirsdóttir, Airi Lampinen and Kristina Höök. 2023. Felt Ethics: Cultivating Ethical Sensibility in Design Practice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-15.
- [38] Dragan Gašević, Shane Dawson, Tim Rogers and Danijela Gasevic. 2016. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education* 28, 68-84.
- [39] Arup Kumar Ghosh, Karla Badillo-Urquiola, Mary Beth Rosson, Heng Xu, John M Carroll and Pamela J Wisniewski. 2018. A matter of control or safety? Examining parental use of technical monitoring apps on teens' mobile devices. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [40] Colin M Gray. 2023. Instructional designers as ethical mediators. *Learning and instructional design technology in the mirror: Integrating ethics and social responsibility into research and practice*, 29-40.

- [41] Colin M Gray and Shruthi Sai Chivukula. 2019. Ethical mediation in UX practice. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1-11.
- [42] Wolfgang Greller and Hendrik Drachsler. 2012. Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society* 15, 3, 42-57.
- [43] Bingyi Han, George Buchanan and Dana McKay. 2023. Learning in the Panopticon: Examining the Potential Impacts of AI Monitoring on Students. In *Proceedings of the 34th Australian Conference on Human-Computer Interaction*. Association for Computing Machinery, Canberra, ACT, Australia, 9–21. <http://dx.doi.org/10.1145/3572921.3572937>
- [44] Bingyi Han, Simon Coghlan, George Buchanan and Dana McKay. 2024. Who is Helping Whom? Student Concerns about AI-Teacher Collaboration in Higher Education Classrooms. *arXiv preprint arXiv:2412.14469*.
- [45] Bingyi Han, Sadia Nawaz, George Buchanan and Dana McKay. 2023. Ethical and Pedagogical Impacts of AI in Education. In *International Conference on Artificial Intelligence in Education*. Springer, 667-673.
- [46] Bingyi Han, Sadia Nawaz, George Buchanan and Dana McKay. 2023. Ethical and Pedagogical Impacts of AI in Education. In *Springer Nature Switzerland*, 667-673.
- [47] Bingyi Han, Sadia Nawaz, George Buchanan and Dana McKay. 2025. Students' Perceptions: Exploring the Interplay of Ethical and Pedagogical Impacts for Adopting AI in Higher Education. *International Journal of Artificial Intelligence in Education*. <http://dx.doi.org/10.1007/s40593-024-00456-4>
- [48] Wayne Holmes, Kaska Porayska-Pomsta, Ken Holstein, Emma Sutherland, Toby Baker, Simon Buckingham Shum, Olga C Santos, Mercedes T Rodrigo, Mutlu Cukurova and Ig Ibert Bittencourt. 2021. Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, 1-23.
- [49] Kenneth Holstein, Bruce M McLaren and Vincent Aleven. 2017. SPACLE: Investigating learning across virtual and physical spaces using spatial replays. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. 358-367.
- [50] Kenneth Holstein, Bruce M. McLaren and Vincent Aleven. 2018. Student Learning Benefits of a Mixed-Reality Teacher Awareness Tool in AI-Enhanced Classrooms. In *Springer International Publishing*, 154-168.
- [51] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé III, Miro Dudik and Hanna Wallach. 2019. Improving fairness in machine learning systems: What do industry practitioners need? In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1-16.
- [52] Silas Hsu, Tiffany Wenting Li, Zhilin Zhang, Max Fowler, Craig Zilles and Karrie Karahalios. 2021. Attitudes Surrounding an Imperfect AI Autograder. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama, Japan, Article 681. <http://dx.doi.org/10.1145/3411764.3445424>
- [53] Stephen Hutt, Kristina Krasich, James R. Brockmole and Sidney K. D'Mello. 2021. Breaking out of the lab: Mitigating mind wandering with gaze-based attention-aware technology in classrooms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [54] Charlotte Jee. 2019. A biased medical algorithm favored white people for health-care programs. Retrieved from <https://www.technologyreview.com/2019/10/25/132184/a-biased-medical-algorithm-favored-white-people-for-healthcare-programs/>.
- [55] Eunkyoung Jo, Daniel A Epstein, Hyunhoon Jung and Young-Ho Kim. 2023. Understanding the benefits and challenges of deploying conversational AI leveraging large language models for public health intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-16.
- [56] Anna Jobin, Marcello Ienca and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature machine intelligence* 1, 9, 389-399.
- [57] Kyle ML Jones, Andrew Asher, Abigail Gobin, Michael R Perry, Dorothea Salo, Kristin A Briney and M Brooke Robertshaw. 2020. "We're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology* 71, 9, 1044-1059.
- [58] Priya C Kumar, Marshini Chetty, Tamara L Clegg and Jessica Vitak. 2019. Privacy and security considerations for digital technology use in elementary schools. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1-13.
- [59] Tzu-Sheng Kuo, Hong Shen, Jisoo Geum, Nev Jones, Jason I Hong, Haiyi Zhu and Kenneth Holstein. 2023. Understanding Frontline Workers' and Unhoused Individuals' Perspectives on AI Used in Homeless Services. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*.
- [60] Celeste Lawson, Colin Beer, Dolene Rossi, Teresa Moore and Julie Fleming. 2016. Identification of 'at risk' students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology Research and Development* 64, 5, 957-968.
- [61] Karen Levy and Bruce Schneier. 2020. Privacy threats in intimate relationships. *Journal of Cybersecurity* 6, 1, tyaa006.
- [62] Chenglu Li, Wanli Xing and Walter Leite. 2022. Using fair AI to predict students' math learning outcomes in an online platform. *Interactive Learning Environments*, 1-20. <http://dx.doi.org/10.1080/10494820.2022.2115076>
- [63] Xiaotian Vivian Li, Mary Beth Rosson and Jenay Robert. 2022. A Scenario-based Exploration of Expected Usefulness, Privacy Concerns, and Adoption Likelihood of Learning Analytics. In *Proceedings of the Ninth ACM Conference on Learning@ Scale*. 48-59.
- [64] Adam Linson, Yucheng Xu, Andrea R English and Robert B Fisher. 2022. Identifying Student Struggle by Analyzing Facial Movement During Asynchronous Video Lecture Viewing: Towards an Automated Tool to Support Instructors. In *International Conference on Artificial Intelligence in Education*. Springer, 53-65.
- [65] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1-16.

- [66] Duri Long, Jessica Roberts, Brian Magerko, Kenneth Holstein, Daniella DiPaola and Fred Martin. 2023. AI Literacy: Finding Common Threads between Education, Design, Policy, and Explainability. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Hamburg, Germany, Article 329. <http://dx.doi.org/10.1145/3544549.3573808>
- [67] Alex Jiahong Lu, Tawanna R Dillahunt, Gabriela Marcu and Mark S Ackerman. 2021. Data Work in Education: Enacting and Negotiating Care and Control in Teachers' Use of Data-Driven Classroom Surveillance Technology. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2, 1-26.
- [68] Alex Jiahong Lu, Gabriela Marcu, Mark S. Ackerman and Tawanna R Dillahunt. 2021. Coding Bias in the Use of Behavior Management Technologies: Uncovering Socio-technical Consequences of Data-driven Surveillance in Classrooms. In *Designing Interactive Systems Conference 2021*. Association for Computing Machinery, Virtual Event, USA, 508-522. <http://dx.doi.org/10.1145/3461778.3462084>
- [69] David Lundie. 2016. Authority, autonomy and automation: The irreducibility of pedagogy to information transactions. *Studies in philosophy and education* 35, 3, 279-291.
- [70] Michael A Madaio, Luke Stark, Jennifer Wortman Vaughan and Hanna Wallach. 2020. Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [71] Michael Madaio, Su Lin Blodgett, Elijah Mayfield and Ezekiel Dixon-Román. 2022. Beyond "fairness": Structural (in) justice lenses on ai for education. In *The ethics of artificial intelligence in education*, Routledge, 203-239.
- [72] Kirsti Malterud, Volkert Dirk Siersma and Ann Dorrit Guassora. 2016. Sample size in qualitative interview studies: guided by information power. *Qualitative health research* 26, 13, 1753-1760.
- [73] Dana McKay and Charlynn Miller. 2021. Standing in the way of control: A call to action to prevent abuse through better design of smart technologies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [74] H James McLaughlin. 1991. Reconciling care and control: Authority in classroom relationships. *Journal of teacher education* 42, 3, 182-195.
- [75] Brenna McNally, Mona Leigh Guha, Matthew Louis Mauriello and Allison Druin. 2016. Children's perspectives on ethical issues surrounding their past involvement on a participatory design team. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3595-3606.
- [76] Danielle S McNamara. 2023. AIED: From Cognitive Simulations to Learning Engineering, with Humans in the Middle. *International Journal of Artificial Intelligence in Education*, 1-13.
- [77] Danaë Metaxa-Kakavouli, Kelly Wang, James A Landay and Jeff Hancock. 2018. Gender-inclusive design: Sense of belonging and bias in web interfaces. In *Proceedings of the 2018 CHI Conference on human factors in computing systems*. 1-6.
- [78] Debra Meyerson. 2004. The tempered radicals: How employees push their companies—little by little—to be more socially responsible. *Stanford Social Innovation Review*, 14-23.
- [79] Fengchun Miao, Wayne Holmes, Ronghuai Huang and Hui Zhang. 2021. *AI and education: A guidance for policymakers*. UNESCO Publishing.
- [80] Ministry-of-Education-of-the-People's-Republic-of-China. 2018. National Artificial Intelligence in Education Action Plan (2019-2022). Retrieved from http://www.moe.gov.cn/srcsite/A16/s3342/201804/t20180425_334188.html.
- [81] Jakub Mlynar, Farzaneh Bahrami, André Ourednik, Nico Mutzner, Himanshu Verma and Hamed Alavi. 2022. AI beyond Deus ex Machina—Reimagining Intelligence in Future Cities with Urban Experts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1-13.
- [82] Sungjin Nam and Perry Samson. 2019. Integrating Students' Behavioral Signals and Academic Profiles in Early Warning System Springer International Publishing, 345-357. http://dx.doi.org/10.1007/978-3-030-23204-7_29
- [83] Sadia Nawaz, Gregor Kennedy, James Bailey, Chris Mead and Lev Horodyskyj. 2018. Struggle town? Developing profiles of student confusion in simulation-based learning environments. In *35th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education, ASCILITE*. 224-233.
- [84] Andy Nguyen, Ha Ngan Ngo, Yvonne Hong, Belle Dang and Bich-Phuong Thi Nguyen. 2023. Ethical principles for artificial intelligence in education. *Education and Information Technologies* 28, 4, 4221-4241.
- [85] Amy Ogan. 2019. Reframing classroom sensing: Promise and peril. *Interactions* 26, 6, 26-32.
- [86] Organization-for-Economic-Co-operation-and-Development. 2021. OECD Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. Retrieved from <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.
- [87] Antti Oulasvirta, Aurora Pihlajamaa, Jukka Perkiö, Debarshi Ray, Taneli Vähäkangas, Tero Hasu, Niklas Vainio and Petri Myllymäki. 2012. Long-term effects of ubiquitous surveillance in the home. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. Association for Computing Machinery, Pittsburgh, Pennsylvania, 41-50. <http://dx.doi.org/10.1145/2370216.2370224>
- [88] Hyanghee Park, Daehwan Ahn, Kartik Hosanagar and Joonhwan Lee. 2022. Designing fair AI in human resource management: Understanding tensions surrounding algorithmic evaluation and envisioning stakeholder-centered solutions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1-22.
- [89] Yanyi Peng, Masato Kikuchi and Tadachika Ozono. 2023. Development and Experiment of Classroom Engagement Evaluation Mechanism During Real-Time Online Courses. In *International Conference on Artificial Intelligence in Education*. Springer, 590-601.
- [90] Catherine Petrozzino. 2021. Who pays for ethical debt in AI? *AI and Ethics* 1, 3, 205-208. <http://dx.doi.org/10.1007/s43681-020-00030-3>

- [91] James Pierce. 2019. Smart home security cameras and shifting lines of creepiness: A design-led inquiry. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [92] Sara Polak, Gianluca Schiavo and Massimo Zancanaro. 2022. Teachers' Perspective on Artificial Intelligence Education: an Initial Investigation. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New Orleans, LA, USA, Article 431. <http://dx.doi.org/10.1145/3491101.3519866>
- [93] Stefan AD Popenici and Sharon Kerr. 2017. Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning* 12, 1, 1-13.
- [94] Erich Prem. 2023. From ethical AI frameworks to tools: a review of approaches. *AI and Ethics*, 1-18.
- [95] Bogdana Rakova, Jingying Yang, Henriette Cramer and Rumman Chowdhury. 2021. Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1, 1-23.
- [96] James Reason. 1990. The contribution of latent human failures to the breakdown of complex systems. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 327, 1241, 475-484.
- [97] Samantha Robertson, Tonya Nguyen and Niloufar Salehi. 2021. Modeling assumptions clash with the real world: Transparency, equity, and community challenges for student assignment algorithms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [98] Samantha Robertson and Niloufar Salehi. 2020. What If I Don't Like Any Of The Choices? The Limits of Preference Elicitation for Participatory Algorithm Design. *arXiv preprint arXiv:2007.06718*.
- [99] John Rooksby, Alistair Morrison and Dave Murray-Rust. 2019. Student perspectives on digital phenotyping: The acceptability of using smartphone data to assess mental health. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [100] Xingran Ruan, Charaka Palansuriya and Aurora Constantin. 2023. Affective Dynamic Based Technique for Facial Emotion Recognition (FER) to Support Intelligent Tutors in Education. In *International Conference on Artificial Intelligence in Education*. Springer, 774-779.
- [101] Laura Scheepmaker, Kay Kender, Christopher Frauenberger and Geraldine Fitzpatrick. 2021. Leaving the field: Designing a socio-material toolkit for teachers to continue to design technology with children. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [102] Daniel Schiff, Justin Biddle, Jason Borenstein and Kelly Laas. 2020. What's next for ai ethics, policy, and governance? a global overview. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. 153-158.
- [103] Mandeep Sekhon, Martin Cartwright and Jill J Francis. 2017. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC health services research* 17, 1, 1-13.
- [104] Max Smolaks. 2019. Chinese university deploys facial recognition tech to monitor student behavior. Retrieved 13th April, 2023 from <https://aibusiness.com/verticals/chinese-university-deploys-facial-recognition-tech-to-monitor-student-behavior>
- [105] Erica Southgate, Karen Blackmore, Stephanie Pieschl, Susan Grimes, Jessey McGuire and Kate Smithers. 2019. Artificial intelligence and emerging technologies in schools.
- [106] Anette Steel. 2003. Understanding and enhancing call centre computer. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, Ft. Lauderdale, Florida, USA, 690-691. <http://dx.doi.org/10.1145/765891.765931>
- [107] Anette Steel, Matt Jones, Mark Apperley and Tristan Jehan. 2002. The use of auditory feedback in call centre CHHI. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, Minneapolis, Minnesota, USA, 768-769. <http://dx.doi.org/10.1145/506443.506588>
- [108] Angelika Strohmayer, Rosanna Bellini, Janis Meissner, Samantha Mitchell Finnigan, Ebtisam Alabdulqader, Austin Toombs and Madeline Balaam. 2018. #CHIiversity: Implications for Equality, Diversity, and Inclusion Campaigns. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Montreal QC, Canada, Paper alt03. <http://dx.doi.org/10.1145/3170427.3188396>
- [109] Angelika Strohmayer, Cayley MacArthur, Velvet Spors, Michael Muller, Morgan Vigil-Hayes and Ebtisam Alabdulqader. 2019. CHIInclusion: Working Toward a More Inclusive HCI Community. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Glasgow, Scotland UK, Paper W27. <http://dx.doi.org/10.1145/3290607.3299012>
- [110] Arnout Terpstra, Alwin De Rooij and Alexander Schouten. 2023. Online Proctoring: Privacy Invasion or Study Alleviation? Discovering Acceptability Using Contextual Integrity. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-20.
- [111] Gareth Terry and Nikki Hayfield. 2020. Reflexive thematic analysis. *Handbook of qualitative research in Education*. Edward Elgar Publishing Limited, 430-441.
- [112] The-Institute-for-Ethical-AI-in-Education. 2021. The Ethical Framework for AI in Education Retrieved from <https://www.buckingham.ac.uk/research/research-in-applied-computing/the-institute-for-ethical-ai-in-education/>.
- [113] Lisa R Trainor and Andrea Bundon. 2021. Developing the craft: Reflexive accounts of doing reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 13, 5, 705-726.
- [114] Yi-Shan Tsai and Dragan Gasevic. 2017. Learning analytics in higher education---challenges and policies: a review of eight learning analytics policies. In *Proceedings of the seventh international learning analytics & knowledge conference*. 233-242.

- [115] Yi-Shan Tsai, Pedro Manuel Moreno-Marcos, Kairit Tammets, Kaire Kollom and Dragan Gašević. 2018. SHEILA policy framework: informing institutional strategies and policy processes of learning analytics. In *Proceedings of the 8th international conference on learning analytics and knowledge*. 320-329.
- [116]
- [117] UNESCO. 2021. Recommendation on the Ethics of Artificial Intelligence. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14>.
- [118] UNESCO. 2023. Generative Artificial Intelligence in education: What are the opportunities and challenges? Retrieved 1st Aug, 2023 from <https://www.unesco.org/en/articles/generative-artificial-intelligence-education-what-are-opportunities-and-challenges>.
- [119] UNICEF. 2021. AI and education: Guidance for policy makers. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000376709>.
- [120] Blase Ur, Jaeyeon Jung and Stuart Schechter. 2014. Intruders versus intrusiveness: teens' and parents' perspectives on home-entryway surveillance. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 129-139.
- [121] Himanshu Verma, Jakub Mlynar, Roger Schaefer, Julien Reichenbach, Mario Jreige, John Prior, Florian Evéquo and Adrien Depeursinge. 2023. Rethinking the role of AI with physicians in oncology: revealing perspectives from clinical and research workflows. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-19.
- [122] Ge Wang, Jun Zhao, Max Van Kleek and Nigel Shadbolt. 2023. 'Treat me as your friend, not a number in your database': Co-designing with Children to Cope with Datafication Online. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-21.
- [123] Alexander Whitelock-Wainwright, Yi-Shan Tsai, Hendrik Drachslers, Maren Scheffel and Dragan Gašević. 2021. An exploratory latent class analysis of student expectations towards learning analytics services. *The Internet and Higher Education*, 100818.
- [124] Alexander Whitelock-Wainwright, Dragan Gašević, Yi-Shan Tsai, Hendrik Drachslers, Maren Scheffel, Pedro J Muñoz-Merino, Kairit Tammets and Carlos Delgado Kloos. 2020. Assessing the validity of a learning analytics expectation instrument: a multinational study. *Journal of Computer Assisted Learning* 36, 2, 209-240.
- [125] David Gray Widder and Dawn Nafus. 2023. Dislocated accountabilities in the "AI supply chain": Modularity and developers' notions of responsibility. *Big Data & Society* 10, 1, 20539517231177620.
- [126] Ben Williamson. 2019. Datafication of education: a critical approach to emerging analytics technologies and practices. *Rethinking pedagogy for a digital age*, 212-226.
- [127] World-Economic-Forum. 2019. Generation AI: Establishing global Standards for children and AI. Retrieved from <https://www.weforum.org/publications/generation-ai-establishing-global-standards-for-children-and-ai/>.
- [128] Alan Yusheng Wu and Cosmin Munteanu. 2018. Understanding older users' acceptance of wearable interfaces for sensor-based fall risk assessment. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1-13.
- [129] Wanli Xing, Xin Chen, Jared Stein and Michael Marcinkowski. 2016. Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in human behavior* 58, 119-129.
- [130] Kexin Bella Yang, Vanessa Echeverria, Zijing Lu, Hongyu Mao, Kenneth Holstein, Nikol Rummel and Vincent Aleven. 2023. Pair-Up: Prototyping Human-AI Co-orchestration of Dynamic Transitions between Individual and Collaborative Learning in the Classroom. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Hamburg, Germany, Article 453. <http://dx.doi.org/10.1145/3544548.3581398>
- [131] Jason C Yip, Kiley Sobel, Caroline Pitt, Kung Jin Lee, Sijin Chen, Kari Nasu and Laura R Pina. 2017. Examining adult-child interactions in intergenerational participatory design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5742-5754.
- [132] Mireia Yurrita, Tim Draws, Agathe Balayn, Dave Murray-Rust, Nava Tintarev and Alessandro Bozzon. 2023. Disentangling Fairness Perceptions in Algorithmic Decision-Making: the Effects of Explanations, Human Oversight, and Contestability. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-21.
- [133] Martina Ziefle and Carsten Röcker. 2010. Acceptance of pervasive healthcare systems: A comparison of different implementation concepts. In *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*. IEEE, 1-6.

5.4 ADDITIONAL REFLECTIONS

The presentation of Study 2 evolved through multiple iterations and critical reflection, shaped by disciplinary expectations and publication processes. A recurring tension in reviewer feedback highlighted a common critique within HCI – while stakeholder inclusion is acknowledged as best practice, industry limitations (e.g., time and cost pressures) in real-world practice often hinder its implementation. This raised important questions about the novelty and practical values of my study, exposing these barriers. The peer comments have urged me to “move beyond shallow critiques to operationalise the calls to action – from ‘what’ to ‘how.’”

Yet, I argue that before asking *how* to fix such issues, we must understand *why* they persist. Study 2 contributes precisely by exposing the structural dynamics – such as power imbalances – that inhibit ethical and inclusive AIEd design in practice. These insights offer a necessary foundation for future interventions, policy development, and more context-sensitive ethical frameworks. Without empirical insights to reveal these dynamics, there would be little basis for raising awareness, fostering discussions, or building consensus for ethical actions like regulation.

As part of this work, I also began developing, or more precisely, translating, the concept of “Ethical Debt” into HCI. This concept describes how flawed initial decision-making processes about technology can lead to harmful or unethical downstream consequences – consequences that decision-makers, especially those who must subsequently address them, come to regret. Together with my supervisory team, *Ethical Debt* is conceptualised as a process comprising three stages: causes, effects, and after-implementation remediations. Study 2 findings are framed as key gaps in ethical decision-making, presenting the causal factors driving ethical debt, with power structures emerging as a critical lens in this process. While *Ethical Debt* is not fully theorised within this thesis, the idea emerged from Study 2 and is currently being developed further in a separate conceptual work.

Chapter 6: Students' Responses: Affective Responses, Learning Beliefs, and Ethical Priorities to AIED Design

This chapter presents the findings from Study 3, which offers a cohesive conclusion to the storyline of this thesis by validating findings from Studies 1 and 2. The previous chapters show a disconnect between stakeholder considerations: students' concerns about AIED (Chapter 4, Study 1) are largely overlooked by more powerful stakeholders, who expect students to accept trade-offs such as compromised privacy and emotional discomfort in exchange for potential educational benefits (Chapter 5, Study 2). Building on these insights, Study 3 serves two aims. First, it examines whether students agree with the value trade-offs imposed by powerful stakeholders. Second, while Study 1 broadly explored students' perceived impacts of AIED and highlighted the role of emotional responses in shaping their learning experiences, Study 3 seeks to test and validate these qualitative insights through a more targeted approach. Specifically, it places students in fictitious classroom scenarios involving specific AIED features and systematically measures their affective and cognitive responses – focusing on their emotional reactions, beliefs about learning, and ethical concerns. This allows for a more direct and structured examination of how students perceive AIED technologies in use, and whether their values and preferences align with institutional assumptions.

Study 3 addresses RQ3: *“What are students' affective responses, beliefs about learning, and prioritised ethical values respecting AIED design and implementation?”* In Study 3, 132 Australian university students participated in an online experiment, with their responses collected through a survey and analysed both quantitatively and qualitatively. The findings build on and integrate insights from Study 1 and Study 2, offering a deeper understanding of the challenges for AIED adoption, regarding the misaligned value priorities. The results suggest that students are critical of the value trade-offs assumed by more powerful stakeholders and desire greater control over how they engage with the AIED system. Importantly, Study 3 reaffirms the significance of emotional well-being in AIED design, an aspect that strongly shapes students' acceptance or rejection of the technology but remains largely overlooked by other stakeholders.

This chapter begins with a brief recap of the study's approach (Section 6.1) and its contribution to the thesis (Section 6.2). It then presents the manuscript arising from Study 3 (Section 6.3), followed by additional reflections (Section 6.4).

6.1 APPROACH FOR STUDY 3

As detailed in Chapter 3 Section 3.6, a mixed-method approach was adopted to collect quantitative and qualitative data. 132 Australian university students participated in the online experiment through a survey. The experiment was structured in two parts.

Part 1. Participants were shown six video prototypes illustrating various AI sensing and intervention features in classroom settings. These features included *Presence* (whether the scenario involved sensing technology), *Detection Method* (whether gaze-tracking or emotion-detection was used for monitoring), and *Teaching Intervention* (whether assistance was provided via a system-generated hint displayed on the screen or a screen alert indicating that a human tutor would soon assist the student). Participants were asked to rate their affective responses (e.g., whether they found the system enjoyable or uncomfortable to learn with) and beliefs about learning (e.g., whether they perceived the system as enhancing personalisation, learning efficiency, etc.) regarding these features using a Likert-scale survey. They also provided detailed perceptions through open-ended questions about the technology presented.

The video scenarios in Study 3 align with the story stems presented in Study 1 and AIED introductory materials presented before interviews in Study 2. This approach ensures consistent insights across the thesis. Additionally, the scale items for affective responses and beliefs about learning are grounded in Study 1's qualitative themes - emotional responses (individual dimension), and the pedagogical impacts, to validate and extend those findings.

Part 2. Participants rated and ranked a series of ethical concern statements, such as issues the technology might create for student autonomy, privacy, transparency, accuracy, and fairness, based on their perceived importance. To assess the insights from Study 2 - whether students agree with the powerful stakeholders' value trade-offs for perceived learning benefits, I included "Learning" as a comparison term alongside these ethical concerns.

Students' ratings of affective responses, learning beliefs, and rankings on ethical concerns were analysed quantitatively using descriptive and comparative statistical

methods. Open-ended responses were analysed qualitatively using inductive Thematic Analysis.

6.2 FINDINGS AND CONTRIBUTION TO THE THESIS

Study 3 contributes to this thesis's overarching research question – “What are the ethical challenges in adopting artificial intelligence in education (AIEd) from the perspectives of different stakeholder groups?” – by examining how students evaluate specific features of AIEd technologies. The findings validate and deepen the insights from Study 1 and contextualise them in light of the value trade-offs highlighted in Study 2, bringing to the fore critical tensions between students and more powerful stakeholders.

The findings reinforce that students' emotional responses are a key dimension in how AIEd is perceived and experienced. Students reported consistently negative affective reactions toward monitoring-oriented sensing features, highlighting discomfort and distraction in learning, with no significant differences between gaze tracking and emotion detection. They also preferred system-generated hints over tutor assistance, citing social concerns about embarrassment from unwanted exposure of their learning struggles to peers. While some students acknowledged potential benefits such as improved personalisation and learning efficiency (in Study 1), these beliefs did not outweigh their discomfort with being constantly monitored, as shown in Study 3.

Importantly, students ranked autonomy and privacy as their top ethical concerns, placing them above learning benefits. They did not readily accept the value trade-offs assumed by more powerful stakeholders, as identified in Study 2. Instead, they expressed a clear preference for having more control over how AIEd systems are implemented. These findings point to a strong value misalignment between students and decision-makers, validating the gaps observed in the previous two studies.

Study 3 makes three contributions. First, it creates and validates a scale for assessing AIEd technologies across learning beliefs and affective responses, building on the qualitative insights from Study 1. The adapted version of this scale is shown in Table 3 (originally Table 2 in the publication manuscript, presented in Section 6.3 of this thesis). Second, it identifies the priorities of students' ethical concerns in AIEd, presented in Table 4, guiding future AIEd design that centres on student value. Third, it offers five actionable, student-centred design implications for AI-enabled teaching and learning interventions, summarised in Table 5.

Table 3. Validated scale for assessing AIEd across beliefs on learning and affective responses.

Construct [Reliability]	Items and Example Statement
Belief about Learning [Cronbach's Alpha = 0.948, high reliability]	Personalisation. 'This system would be useful in meeting my individual learning needs.'
	Efficiency. 'Using this system could make my learning process easier.'
	Control of pace. 'I would have complete control of my own learning at the pace I want when using this system.'
	Distraction. 'It would distract me to think that my learning progress was observed.'
	Grade performance. 'Using this system could help me gain better grades on the learning tasks.'
	Over-helping. 'Using this system would give me so much help that I don't really learn.'
Affective Response [Cronbach's Alpha = 0.957, high reliability]	Enjoyable. 'My learning experience with this new feature is more enjoyable.'
	Confident. 'Using this system helps me feel more confident in my learning.'
	Anxious. 'I feel anxious about using the system to learn.'
	Uncomfortable. 'It is uncomfortable to think that my learning progress is observed.'
	Worried about flagging out. 'I am worried about the system flagging that I am struggling.'
	Fearing mistakes. 'I fear making mistakes, as it is observing my learning progress.'

Table 4. Synthesised rankings and ratings of students' ethical concerns in AIEd.

Note: This Table synthesises findings from Study 3, integrating results from two figures in the publication manuscript. Pairwise prioritisation of ethical constructs using the Bradley–Terry model (Figure 3) is reported in the second column, while the third column presents the corresponding ethical items from Likert ratings (Figure 2) with their mean concern scores (scale: 1 = Not at all concerned; 5 = Extremely concerned). Constructs are ordered by their probability of being prioritised (“wins”).

Ranked Ethical Construct (Probability of ‘wins’)	Ethical Items Highlighted (Mean Concern Level)
1. Autonomy (0.39)	Not being given the option to opt out of using the system. (4.39) Not being asked for my consent before using the system. (4.21) Having the fact that I am interacting with AI hidden from me. (3.70)
2. Privacy (0.27)	Protecting my personal data from breaches. (4.10) Sharing information about my activities with others, including teachers. (4.09)

	Sharing information about my emotional states with others, including teachers. (4.23)
3. Transparency (0.09)	Not being informed about how all the data collected about me is being used. (4.17) Not being informed about the potential risks and flaws of the system. (4.10) Not really understanding how the system's AI algorithm works. (3.32)
4. Learning Beneficence (0.09)	The system not respecting my individual learning pace and preferences. (4.05) Not having access to this system, and therefore not getting enough attention from teaching staff. (2.30) Not having access to this system and therefore waiting for help for too long. (2.16)
5. Accuracy (0.08)	The system inaccurately interpreting my behaviour and emotions. (4.25) The system offering too much help when I don't need it. (3.82) The system failing to recognise when I need help. (3.20)
6. Fairness (0.08)	Not getting the same amount of support as other students get. (2.39) The system discriminating against students based on, e.g., their race and gender. (2.92) The system only being available to privileged students. (2.61)

Table 5. Five Student-Centred Design Implications for AI Learning Intervention

Note: These implications synthesise design preferences expressed by students across quantitative and qualitative data in Study 3. They provide actionable guidance for developers and educators seeking to align AIED with student agency, privacy, and well-being.

No	Design Implication	Description
1	Minimise intrusive sensing	Prioritise less invasive approaches (e.g., task data, individual sensing with opt-out options) instead of gaze tracking or emotion recognition.
2	Enable customisable, student-triggered interventions	Hints should be lightweight, optional, and adjustable in frequency and style; interventions must be suggestions, not impositions.
3	Redesign teacher-mediated interventions	Preserve agency by giving students control over when teachers are notified; use private or aggregated channels to reduce embarrassment.
4	Implement dynamic consent and transparency	Move beyond one-off consent; provide layered, situational opt-in prompts and explain clearly why and how the system acts.
5	Provide private feedback channels	Allow students to receive hints or warnings privately before teacher involvement, creating a buffer for self-correction.

Study 3 directly addresses *Problem 3: Neglecting Students' Perspectives in AIED Design*, by offering concrete empirical insights into how students respond to specific AIED features and what they value most in technology-supported learning environments. It reinforces the importance of designing AIED systems that not only meet performance goals but also uphold students' preferences, agency, and well-being.

Study 3 also addresses *Problem 1: Lack of Understanding of Unintended Consequences of AIEd*, by revealing how discomfort, distraction, anxiety, and behavioural impacts emerge as unintended side effects of well-intentioned AIEd features. These responses are not abstract concerns but tangible student reactions that, if ignored, risk undermining acceptance and learning engagement. The findings offer design implications (detailed in Section 6.3) that can help make AIEd systems more acceptable and ethically aligned with student needs.

Study 3 further addresses *Problem 2: Limited Understanding of Stakeholder Dynamics in AIEd Adoption*. When viewed alongside Studies 1 and 2, it becomes clear that students' ethical priorities - particularly autonomy and privacy - are at odds with the assumptions held by more powerful stakeholders. They often justify intrusive technologies with perceived educational benefits and expect students to comply. This tension amplifies the ethical gaps and accountability concerns discussed in Chapter 5 and calls for more inclusive governance approaches that meaningfully integrate student voices.

Study 3 also addresses *Problem 4: Limited Understanding of Ethics in AIEd*. It reinforces that ethics in AIEd cannot be reduced to technical concerns such as data protection or algorithmic fairness. Rather, it highlights that values like autonomy, privacy, and well-being are central to how students perceive and accept AIEd technologies. Yet, some of these dimensions are often marginalised in institutional adoption processes. A particularly important contribution of Study 3 is its illumination of students' *emotional well-being* as a critical, yet underrepresented, ethical concern. The findings reveal that students' negative responses to AIEd features are closely tied to discomfort, anxiety, and pressure - emotional experiences that shape how students relate to and accept AIEd systems. However, these affective responses remain largely absent from existing AIEd ethics discussions. This insight points to a concrete direction for extending current AIEd frameworks: embedding emotional well-being as a core component of human-centred ethical design. I return to this proposition in Chapter 7 (Section 7.2), where I argue for incorporating emotional ethics into future frameworks as a central pillar of students' well-being in technology-mediated learning environments.

Ultimately, Study 3 findings synthesise and bring closure to the thesis narrative. It solidifies the thesis's central argument: that ethical, context-sensitive AIEd design and adoption must be grounded in students' lived experiences and value priorities. Ethical

adoption is not only about technical safeguards but also about recognising students as active agents whose well-being and engagement should guide how AI technologies are designed and implemented in education.

6.3 STUDENTS' AFFECTIVE RESPONSES, BELIEFS ABOUT LEARNING, AND PRIORITISED ETHICAL VALUES TO AIED FEATURES

Study 3 findings have resulted in a manuscript that was submitted for publication to the CHI 2026 conference in September 2025.

The paper manuscript begins on the next page.

AI Sensing and Intervention in Higher Education: Student Perceptions of Learning Impacts, Affective Responses, and Ethical Priorities

AI technologies that sense student attention and emotions to enable more personalised teaching interventions are increasingly promoted, but raise pressing questions about student learning, wellbeing, and ethics. In particular, students' perspectives about AI sensing-intervention in learning are often overlooked. We conducted an online mixed-method experiment with Australian university students (N=132), presenting video scenarios varying by whether sensing was used (in-use vs. not-in-use), sensing modality (gaze-based attention detection vs. facial-based emotion detection), and intervention (by digital device vs. teacher). Participants also completed pairwise ranking tasks to prioritise six core ethical concerns. Findings revealed that students valued targeted intervention but responded negatively to AI monitoring, regardless of sensing methods. Students preferred system-generated hints over teacher-initiated assistance, citing learning agency and social embarrassment concerns. Students' ethical considerations prioritised autonomy and privacy, followed by transparency, accuracy, fairness, and learning beneficence. We advocate designing customisable, social-sensitive, non-intrusive systems that preserve student control, agency, and well-being.

CCS CONCEPTS • Human-centred computing → User studies; *Empirical studies in HCI*.

Additional Keywords and Phrases: AI in Education, Ethics, Privacy, Student Agency

1 INTRODUCTION

AI-enabled sensing and intervention systems are increasingly promoted in higher education [104] to support data-informed teaching [63, 84] and personalised learning [4]. These technologies often use camera-based inputs, such as eye movement and facial cues, to monitor students' attention and emotional engagement [13, 77]. Such systems can then either generate automated intervention suggestions for teachers (e.g., [7, 77, 94]), or provide real-time tailored feedback to students (e.g., through Intelligent Tutoring Systems [79]). These systems aim to enhance teachers' situational awareness [65, 93, 141], and deliver timely, individualised learning support to students.

Despite rapid technical advances, research on AI in Education (AIEd) often overlooks deeper student-centred and learning factors [18, 24]. Prior work has largely taken a technology-centred perspective to understand which design features can improve system effectiveness, efficiency, and usability (e.g., [86]). Yet technical performance alone does not guarantee positive or meaningful educational experiences (e.g., [57]). Particularly underexplored are students' perspectives on how various system features might influence their affective experiences [80] and learning [29]. These aspects can significantly shape students' engagement with AIEd [17, 50, 56]. For instance, discomfort from being observed by AIEd may trigger behavioural change [56] and undermine the technology's effectiveness [57]. Misalignment with student needs and expectations could precipitate disengagement or resistance, even if the tools are technically effective.

Crucial AI ethics considerations are equally underexamined in AIEd [62, 104]. For instance, improving personalisation and accuracy requires extensive data, threatening privacy [130]. Timely intervention may aid learning efficiency but override self-directed learning status [58, 91], diminishing autonomy. Improved AI accuracy may mean increased model opacity, while greater transparency can reduce system performance [10]. Such ethical trade-offs present design

dilemmas. Furthermore, designing AI without appreciating how students perceive value tensions risks building technically sound but socially unacceptable systems. Student-centric AIED thus requires consideration of relevant ethical concerns from the outset [46].

To address these gaps, we conducted a two-stage experimental study with 132 Australian university students via an online survey. Stage 1 examined student preferences for AI sensing-intervention features. Participants viewed six video scenarios featuring screen mockups of AI systems, which varied along three dimensions: 1) whether sensing was used; 2) sensing approach (gaze-tracking attention detection vs. facial-based emotion detection); 3) form of intervention (system-generated instruction vs. teacher-mediated assistance). Participants rated their responses using five-point Likert items after each scenario and provided open-ended feedback. Stage 2 explored how students weigh ethical value trade-offs, asking participants to rank concerns through pairwise comparison tasks. Our research answers:

RQ1 (Stage 1): *How do students perceive AI sensing and intervention systems with various features?*

- RQ1 (a): *How do students evaluate the likely learning impacts of these systems?*
- RQ1 (b): *How do students affectively respond to these systems?*

RQ2 (Stage 2): *How do students prioritise ethical concerns regarding AI sensing and intervention systems?*

Our findings reveal that students' affective responses and beliefs about learning and ethics indicate a consistently negative reception to the monitoring element of such technology (regardless of whether gaze-based or facial-based detection was involved) in some of its current forms. In contrast, students showed a clear preference for system-generated direct support over teacher-initiated assistance, citing desires for agency and concerns about embarrassment before peers. Regarding ethical views, students were most concerned about autonomy and privacy, suggesting they highly value control over learning processes and data in interactions with educational technology. Our study makes three contributions to HCI and AIED design. It:

1. Foregrounds human-centred dimensions in student-AI interaction, examining crucial affective and pedagogical responses that are rarely evaluated. Our findings challenge prevailing notions of effectiveness metrics, highlighting the importance of incorporating emotional well-being, social dynamics, and ethical values into AIED design and assessment.
2. Presents new empirical evidence for how students prioritise ethical values when engaging with AIED. This offers a foundation for aligning AI tools with student values and provides a transferable method for ethical design in other sensitive domains where value tension arises.
3. Proposes five student-centred design implications for AIED: minimise intrusive sensing; enable student-triggered, customisable, and lightweight system interventions; redesign teacher-mediated interventions to preserve student agency and control; implement dynamic consent mechanisms and transparency to reinforce trust; and provide private feedback channels before teacher involvement.

2 RELATED WORK

This section reflects the two-stage structure of this study. Section 2.1 reviews previous work for Stage 1, justifying the hypotheses to answer RQ1. Section 2.2 reviews ethical literature for Stage 2, introducing ethical principles examined to address RQ2.

2.1 Student experiences with AI sensing-intervention

2.1.1 AI sensing-intervention

Recent advances in educational AI increasingly emphasise real-time sensing of learner states to enable adaptive feedback and enhance instructors' situational awareness [2, 35, 143]. Timely data-informed learning support has been recognised as an effective mechanism for education [1, 23]. While sensor-equipped or instrumented classrooms have explored embedding computing into physical infrastructures (e.g., pressure-sensitive chairs or mice [6], wearable devices with biometric sensors that physiologically identify arousal states [11, 54, 140]), these approaches pose logistical or financial challenges [1]. These practical limitations have generated significant interest in visual approaches in AI sensing [1], using computer vision and machine learning to capture students' visual and behavioural cues (e.g., gaze direction [2, 66, 87], facial expressions [13, 44], body gestures [76, 92], head orientation [1, 105, 106]) and to infer cognitive attention, emotional engagement, and interactions [35, 132, 143].

AI Sensing: Among camera-based sensing modalities, gaze-based attention detection and facial emotion detection are predominant [1, 3]. Using existing cameras, such as those in student laptops, makes AI sensing practical for widespread adoption.

- *Gaze-based Attention Detection* leverages eye-tracking technology to capture temporal changes in gaze and infer attentional state in real time [66]. This approach offers insights into crucial cognitive processes such as focus and distraction [3, 71]. It has been applied to assess student performance in complex tasks like surgical simulations [77, 113] and to mitigate mind-wandering with timely interventions (e.g., GazeTutor [36], Guru ITS [66]). Such systems may successfully redirect learners' attention and improve learning outcomes [36, 66].
- *Facial Emotion Detection* potentially offers rich situational learning information on student engagement [134] and states like excitement [76], confusion, and frustration [20, 70]. Computer vision detects facial landmarks, which are fed into machine learning models trained to classify emotional expressions [13, 44, 117]. This may give valuable insight into students' motivation and real-time learning experiences. A well-known example is EmoDash [44], a teacher-facing dashboard that visualises learners' emotional states in real time.

AI Intervention: Insights from AI sensing of attention and emotion can be directly used to trigger instructional interventions aimed at providing timely, personalised support. Two main forms exist:

- *Teacher-Mediated Assistance:* Teacher-facing systems [63, 84] can provide insight about individual students or class-wide progress, such as who is struggling or disengaged [124], enabling teachers to adjust instruction [7, 8, 44, 51], target their attention [64] and offer timely support [33, 64, 66, 138]. This positions teachers as “orchestrators of when and how to use AIED tools” [90].
- *System-Generated Instruction:* Student-facing systems [15] include Intelligent Tutoring Systems (ITS) [117], which automatically deliver hints, feedback, or re-engagement messages [79] directly to students. For instance, attention-aware ITS such as GazeTutor [36] redirect attention when students become distracted, and Guru [66] and ATS [86] adapt feedback and provide hints based on emotional cues.

While attracting much interest, these promising technologies still lack strong empirical evidence and robust pedagogical grounding [129]. Furthermore, little is known about how users, especially students, respond to these approaches and their combinations. Addressing this gap is essential for student-centred AIED design.

2.1.2 Gaps in understanding students' experiences with AIED

Understanding students' perceptions of AIED is critical for aligning technology with users' values and needs [9, 19], and ensuring the intended pedagogical outcomes are achieved. Recent HCI research shows that when students feel uncomfortable with AI systems that monitor their behaviour, they may consciously alter how they act, pandering to or even deceiving the system's sensing mechanisms [56]. Such responses not only reduce AI accuracy and effectiveness but may also disrupt students' sense of agency and trust in the learning process [57, 59]. Thus, we must examine not just how well AIED systems function technically, but also how students experience, interpret, and respond to them.

However, most studies exploring students' attitudes toward AIED remain general. They often report overall positive views (e.g., [5, 72, 108]) and advocate for the technology's implementation [5, 108], citing benefits such as timely feedback [53, 118] and improved teaching quality [45, 53]. A few studies indicate student concerns [135], including data security and privacy [69, 130], transparency and user consent [135, 136], misinterpretation of learning performance or needs [53, 118], and feelings of excessive monitoring and control [118]. Yet deeper exploration of students' responses to specific AI applications or features is lacking [49].

Furthermore, while some empirical studies have evaluated students' experiences with specific AIED technologies such as ITS and Affective ITS, these evaluations largely focus on student performance metrics like pre-/post-learning gains according to scores [82, 86], faster task-completion times [99], usability ratings [82, 86, 99], and technology acceptance [29]. While such findings demonstrate statistically significant learning improvements and system effectiveness [49], they risk narrowing student attitude and acceptance to system efficiency alone. Deeper human aspects of learning experiences, such as students' learning status and autonomy, motivation, high-order thinking, and self-efficacy, are rarely examined [29]. Learning is more than task completion; it is a complex phenomenon that involves developing curiosity, agency, and personal growth. Without deeper investigation into how students experience or perceive AIED, we risk failing to evaluate such technologies from a pedagogically meaningful, user-centred standpoint [9, 19].

Two recent speculative studies have begun to explore students' nuanced reactions to classroom observational AI [56] and personalised intervention [59], together painting a relatively negative picture about students' perceptions. They offer early insights into the emotional and pedagogical concerns that AIED may raise, including potential heightened anxiety, disrupted learner status [58, 91] and pedagogical interactions [56, 88, 137]. While these studies also noted positive expectations such as receiving timely support and gaining data-informed insights, they did not identify which specific features of AIED trigger negative emotional or learning-related responses in students. Nor did they address how students interpret trade-offs when sensing data is used to deliver the benefits they anticipate. This points to a critical gap in understanding how combinations of sensing and intervention features affect students' attitudes.

Hence, this study addresses the lack of research on students' learning-related beliefs and affective responses to AI sensing-intervention through RQ1. RQ1(a) – *How do students evaluate the likely learning impacts of these systems?* – builds on prior work highlighting students' expectations for personalisation, efficiency [53, 118], academic improvement [82, 86], and learning autonomy [56, 59]. RQ1(b) – *How do students affectively respond to these systems?* – draws on evidence that students' emotional responses significantly shape their attitudes toward AI [72] and learning engagement [57], including enjoyment [56], confidence [83, 86, 99], anxiety [56], discomfort [56], and fear of mistakes [56].

2.1.3 Hypotheses in Stage 1

To address RQ1, we generated a set of hypotheses on students' perceptions of different AI sensing-intervention approaches. Given that such systems typically involve a sequential process that first collects students' behavioural data

through sensing and delivers adaptive support through intervention, we break down the evaluation to examine perceptions of each component individually.

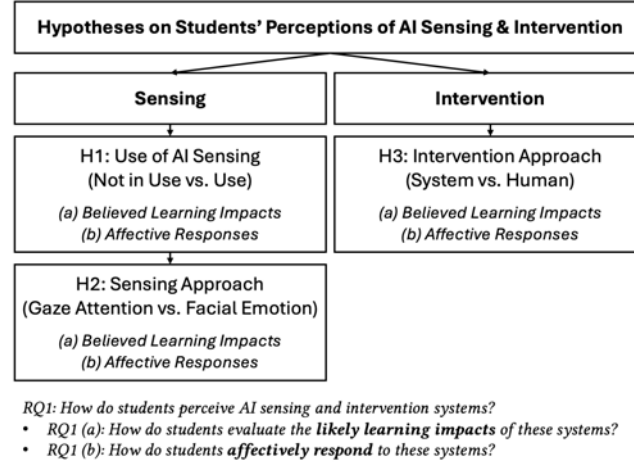


Figure 1. Overview of study hypotheses.

Prior research on ITS has shown statistically significant improvement in learning outcomes (e.g., pre/post-test gains in scores and reduced task-completion time [82, 86, 99]), while other studies often highlight students' positive expectations of receiving learning benefits such as personalisation and improved efficiency [53, 118]. However, studies have also reported negative perceptions when students are aware of being observed by AI, including emotional discomfort and a sense that monitoring may disrupt learning [56, 59]. Building on these contrasting findings, we propose:

H1 - Students perceive AI systems that rely on sensing more negatively than systems that do not use sensing – both in terms of (a) potential learning impacts and (b) affective responses.

It also seems reasonable to expect that students may perceive gaze-based attention detection more positively than facial-based emotion detection, due to differences in how personal the collected data feels and perhaps also the perceived pedagogical relevance. Facial emotion recognition, which relies on sensitive biometric data (e.g., facial expressions), is legally classified as high-risk in school settings under the GDPR and EU AI Act [40]. In contrast, gaze-based attention detection typically involves observable behavioural cues (where the student is looking) and may feel less intrusive. Students may therefore feel more comfortable with gaze-based systems. Pedagogically, attention is a well-established indicator of learning outcomes [109], whereas the connection between AI-detected emotions and learning is perhaps more ambiguous and context-dependent. Accordingly, we propose:

H2 - Students perceive gaze-based attention detection more positively than facial-based emotion detection – both in terms of (a) potential learning impacts and (b) affective responses.

Previous research suggests some students prefer automated assistance over human support to avoid feeling judged [57, 115]. In contrast, other studies report a preference for human-delivered help, citing perceived empathy [41, 42]. This tension becomes particularly relevant in the context of AI sensing-intervention systems, where automated support might feel more private but less human. Meanwhile, ITS's offering personalised automated hints and explanations has been

shown to improve learning outcomes and receive positive evaluation from students [78, 131]. To explore this further, we specifically examine which form of assistance students feel more comfortable receiving. We propose:

H3 - Students perceive system-generated assistance more positively than teacher-mediated assistance – both in terms of (a) potential learning impacts and (b) affective responses.

In summary (see Fig.1), H1 relates to whether AI sensing should be used at all. H2 involves considering whether the type of sensing matters (gaze-based vs. facial-based). H3 relates to asking how, regardless of sensing, the intervention should be delivered (system-generated vs. teacher-mediated assistance). All hypotheses align with our overarching research aim of understanding how students evaluate AI sensing-intervention tools.

2.2 Students' ethical priorities for AI sensing-intervention

Ethical values are central to human-centred computing [102]. Recent review studies [125], academic critics [61, 62], and policy frameworks (e.g., UNESCO [97], World Economic Forum [139]) highlight key ethical concerns for AIEd. These concerns primarily involve data privacy, bias and fairness, transparency, autonomy, accuracy, and impacts on pedagogy [62]. However, these values may conflict in real-world implementation. For instance, improving model accuracy may require big data, thereby compromising privacy [130]. Enhancing algorithm performance may reduce transparency [10], as it can involve using complex 'black box' models that are difficult to interpret or explain [107]. Moreover, opting out of AI-enabled education to preserve autonomy may lead students to forgo potential learning benefits, raising fairness concerns. Students themselves report experiencing such trade-offs, weighing the benefits of participating in AI-enabled systems against perceived privacy concerns [111]. Such trade-offs mean that fully meeting all ethical principles simultaneously can be costly or impossible. In practice, developers often make decisions about which values to prioritise or compromise [74]. However, it is important to recognise that compromising any ethical principle, even with good intent, may harm students.

No existing study has examined how students themselves evaluate these trade-offs. While prior research has documented students' general attitude to AI, often highlighting privacy and fairness as the highest-rated concerns [67], these insights remain high-level. Yet users' attitudes toward ethical risk are highly context-dependent [122]. For instance, people may accept algorithmic decision-making in social media recommendations and reject it in financial services [67]. Explainability may be deprioritised compared to accountability, fairness, security, privacy, and accuracy in tax fraud detection [71]. In education, overlooking how students navigate value tensions could inadvertently undermine agency and trust. Hence, we pose *RQ2: How do students prioritise ethical concerns regarding AI sensing and intervention systems?* Below, we explain key ethical concepts examined in Stage 2.

2.2.1 Ethical constructs examined in Stage 2

To ground this study's ethical prioritisation task in Stage 2, we draw on widely recognised frameworks. Jobin et al. [68] identified five core AI ethics principles: non-maleficence, privacy, justice and fairness, transparency, and responsibility. Influenced by bioethics, Floridi and Cowls [47] stressed beneficence, non-maleficence, justice, autonomy, and explicability. Such broad principles, however, need contextualisation within educational settings [61, 62]. Holmes et al.'s [62] influential AIEd-specific framework highlighted privacy, data security, fairness, bias, autonomy and agency, transparency, accountability, and inclusion. Six recurring ethical values from these frameworks form the basis of our Stage 2 ranking task. Here, we explain these ethical concepts and their significance for AI sensing-intervention.

(Learning) Beneficence: Beneficence refers to assisting and benefitting others [47]. In AI ethics, beneficence demands that AI systems be capable of achieving benefit [98]. In the AIEd context, beneficence requires that technologies meaningfully support students' learning, engagement, and educational development [104]. Importantly, beneficence in AIEd must be grounded in evidence of actual educational value [129].

Privacy: Privacy is a core expression of the non-maleficence principle [48] of "do no harm" [98] and requires safeguarding individuals' personal information. AI systems may collect, store, and process large volumes of sensitive personal data [97]. Privacy concerns include preventing data misuse, hacking, or leakage [83] and respecting individuals' control over how their information is gathered and shared. Persistent privacy worries have been reported by students [60, 69, 81, 100, 126, 130], educators [53, 85], and school officials [26], particularly regarding excessive data collection [97] and breaches of personal information [83].

Accuracy: Accuracy, another aspect of non-maleficence, requires that AI systems avoid harm [48] through correct interpretations and decisions [98]. In AIEd, inaccurate interpretation and predictions concerning behaviour can mislabel students and trigger disruptive interventions [28, 52, 57]. For instance, remote exam AI proctoring tools may falsely flag innocent behaviours, causing distress and unfair allegations of cheating [30]. Accuracy failures in AIEd have consequences for fairness, well-being, and trust (e.g., [30]).

Autonomy: Autonomy refers to students' ability to pursue meaningful goals [62] and make informed decisions [48]. Respecting autonomy may involve the right to consent to AI use, to opt out, and to exercise control over data [60]. It may entail preserving some level of student agency in learning [59]. The degree of choice that students have concerning AI in education may depend on circumstances, but a perceived lack of choice may negatively affect student attitudes to AIEd [59].

Fairness: Fairness is very frequently mentioned in discussions of AIEd [125]. Fairness requires that AI models avoid or mitigate discriminatory or unjust consequences [55, 122] to ensure equitable treatment across diverse backgrounds [95]. This is particularly pressing in AIEd applications such as personalised content delivery, performance evaluation and prediction, and facial emotional recognition (e.g., [39]). Here, biased training data or model assumptions [89, 96] can cause unfair student outcomes [14, 128, 142] regarding race, ethnicity, nationality, and gender [62]. For instance, automated essay-grading systems may assign lower scores to essays by Black over White students for equivalent content [28, 52]. Notably, fairness has been shown to most strongly influence users' satisfaction with algorithms, followed by transparency [120].

Transparency: Transparency in AI systems can build trust [119] and empower users [27]. It can involve explanations [114] to stakeholders about data collection, algorithmic function [101], and system limitations, and is linked to informed consent [95] and respect for autonomy [30]. Low transparency in AIEd can leave students uncertain about what is being measured or how interventions are triggered [56], generating feelings of surveillance and anxiety, and self-censorship [56]. Transparency may be a requirement [30] for student understanding, trust, and meaningful participation [56].

3 STUDY METHOD

This study employed a two-stage design. Stage 1 (RQ1) examined students' preferences for AI sensing-intervention features through an online experiment with video scenarios, followed by Likert-scale responses and open-ended reports. Stage 2 (RQ2) examined ethical priorities where participants rated and ranked ethical principles. Ethical approval for the study was granted by [anonymised].

3.1 Stage 1: Scenario-Based Experiment

3.1.1 Materials and variable design

We employed a within-subjects experimental design with three independent variables: whether sensing was used (sensing-in-use vs not-in-use), sensing modality (gaze-based attention detection vs facial-based emotion-detection, only when sensing was in-use), and intervention form (system-generated vs. teacher-mediated intervention).

- Use of AI sensing: whether the scenario depicted a system that relied on AI sensing (*sensing-in-use*) or a system that did not use sensing (*not-in-use*). In the experiment, this was conveyed by informing participants that AI sensing was operating (*sensing-in-use*) or by omitting such information (*not-in-use*). In practice, students' learning status could also be inferred through non-sensing approaches (e.g., time spent on tasks or interaction patterns within the software), though we did not detail these alternative approaches in the scenarios.
- Sensing approach: whether the sensing involved *gaze-based attention detection* or *facial-based emotion-detection*.
- Intervention form: whether the learning intervention was offered via a *system-generated hint* (displayed on the screen) or via a *human teacher* (meaning the issuing of a screen alert that a human teacher would soon visit the student to assist).

Experimental Scenarios. To represent these combinations (see Table.1), we created six video scenarios using simple screen-based prototypes (Appendix A) and presented them within a Qualtrics survey. Each video simulated a student's view of a learning task within a typical learning management system, overlaid with animated features representing AI sensing-intervention functionalities. These visual materials were accompanied by text instructions and voice-overs within the videos to provide participants with context for the prototypes they were viewing. This commentary emphasised that our goal was not to criticise the technology but rather to explore implications to inform better design. All instructions were carefully worded with neutral terms such as "... the system observes you", avoiding expressions that might bias participants' perceptions.

3.1.2 Procedure

Participants entered the study via the Qualtrics survey platform. After reading an overview introducing that the research focused on exploring their attitudes to intelligent technologies in education, they provided consent for participation. We informed them that two sequential stages of data collection followed.

Table 1: Scenario presentation sequence in Stage 1 ^a

Sensing not-in-use (S1, S2 randomised)	
S1: System-generated hint (System hint)	S2: Human-mediated assistance (Human assistance)
Sensing in-use (S3-S6 randomised)	
S3: Gaze-based attention detection + System hint	S4: Gaze-based attention detection + Human assistance
S5: Facial-based emotion detection + System hint	S6: Facial-based emotion detection + Human Assistance

^a For the first two scenarios (S1 or S2, presented in randomised order), participants were not informed that sensing was involved. For the remaining trials (S3-S6, represented in randomised order), participants were explicitly informed that AI-based sensing was operating. These scenarios varied by sensing approach and intervention form.

In stage 1, participants watched six scenarios (Table.1) and were asked to imagine themselves using the systems displayed. After each scenario, participants responded to a set of five-point Likert items to record their beliefs about the likely learning impact of the technologies and their affective responses. Upon completing all 6 scenarios, participants

were given an open-ended question that invited them to enter free text to describe in richer detail their responses in relation to all the systems viewed. This step was included to gain qualitative insights to interpret patterns in the Likert ratings, and for the researchers to learn about any aspects of experience that had been overlooked.

3.1.3 Measures and scale validation

We designed a set of Likert-scale items to assess participants' beliefs about learning and affective responses to sensing and intervention technologies. Existing technology acceptance scales (e.g., [75]) were not suitable to capture the specific affective and pedagogical concerns raised in prior observational and qualitative studies of AIED systems (e.g., [56, 59]). Therefore, we devised new items informed by the literature reviewed in Section 2.1.2 (first column of Table.2) and grounded in relevant theoretical frameworks on technology perception and user experience. Each item was mapped to an established psychological concept or cognitive theory (second column of Table.2) to support construct validity. The items were piloted with 10 students to ensure clarity and ease of interpretation, with minor wording revisions made based on their feedback.

Table 2: Likert items presented after each scenario presentation

Constructs and Likert Statements	Related Concepts
<i>Belief about learning [Cronbach's Alpha = 0.948, high reliability]</i>	
♦ personalisation [53, 118] - 'This system would be useful in meeting my individual learning needs.'	Perceived usefulness [37]
♦ efficiency [53, 118] - 'Using this system could make my learning process easier.'	Job-fit [127]
♦ control of pace [56, 59] - 'I would have complete control of my own learning at the pace I want when using this system.'	Perceived behavioural control [73]
♦ distraction [56] - 'It would distract me to think that my learning progress was observed.'	
♦ grade performance [82, 86] - 'Using this system could help me gain better grades on the learning tasks.'	Social-cognitive theory (individual esteem, sense of accomplishment) [32]
♦ over-helping [59] - 'Using this system would give me so much help that I don't really learn.'	
<i>Affective response [Cronbach's Alpha = 0.957, high reliability]</i>	
♦ enjoyable [56] - 'My learning experience with this new feature is more enjoyable.'	Attitude toward behaviour [38, 144]
♦ confident [83, 86, 99] - 'Using this system helps me feel more confident in my learning.'	
♦ anxious [56] - 'I feel anxious about using the system to learn.'	Emotional reaction to technology (ERT) model [17]
♦ uncomfortable [56] - 'It is uncomfortable to think that my learning progress is observed.'	Positive affect negative affect schedule (PANAS scale) [133]
♦ worried [56] - 'I am worried about the system flagging that I am struggling.'	Students' emotional responses to observational AI [56]
♦ fearing mistakes [56] - 'I fear making mistakes, as it is observing my learning progress.'	

Internal consistency of the scales was assessed using Cronbach's alpha [34]. Both constructs of *Beliefs about Learning* and *Affective Responses* showed high reliability. Each construct combined six items with both positive and negative items. We applied reverse-scoring to the negative items to align their direction with the positive items. Cronbach's Alpha was then computed, resulting in reliability scores of 0.948 for *Beliefs about Learning*, and 0.957 for *Affective Responses*, both well above the recommended threshold of 0.70 [103]. These reliability scores indicate that the items consistently measured students' perceptions of learning impacts and affective responses, which we used to test whether these two dimensions were viewed positively.

3.1.4 Analysis

For Likert-scale quantitative responses, we conducted non-parametric Mann–Whitney U tests with Bonferroni corrections to compare conditions across items. For open-ended qualitative responses, we employed thematic analysis following Braun and Clarke’s approach [22]. The first author inductively coded the data and generated initial themes. The research team then discussed and reflected on the themes through multiple rounds, iteratively refining the themes through a review and writing process.

3.2 Stage 2: Ethics Pairwise Ranking Tasks

After completing the scenario-based evaluations in Stage 1, participants proceeded to Stage 2, which involved two sequential tasks.

3.2.1 Design and procedure

First, participants rated their level of concern about 18 statements describing potential ethical issues, using a five-point Likert scale (1=not at all concerned, 5=extremely concerned). These statements were grouped under six broad ethical concepts: Privacy, Autonomy, Fairness, Accuracy, Transparency, and Learning Beneficence (Table.3). This task aimed to provide a descriptive overview of participants’ ethical concerns depicted across all the viewed scenarios in Stage 1.

Table 3: Statements of potential ethical concerns

Statement Tag	Statement
Privacy_1	Protecting my personal data from breaches.
Privacy_2	Sharing information about my activities with others, including teachers.
Privacy_3	Sharing information about my emotional states with others, including teachers
Autonomy_1	Having the fact that I am interacting with AI hidden from me.
Autonomy_2	Not being asked for my consent before using the system.
Autonomy_3	Not being given the option to opt out of using the system.
Fairness_1	Not getting the same amount of support as other students get.
Fairness_2	The system discriminating against students based on, e.g., their race and gender.
Fairness_3	The system only being available to privileged students.
Accuracy_1	The system inaccurately interpreting my behaviour and emotions.
Accuracy_2	The system failing to recognise when I need help.
Accuracy_3	The system offering too much help when I don’t need it.
Transparency_1	Not really understanding how the system’s AI algorithm works.
Transparency_2	Not being informed about how all the data collected about me is being used.
Transparency_3	Not being informed about the potential risks and flaws of the system.
Learning_1	Not having access to this system and therefore waiting for help for too long.
Learning_2	Not having access to this system, and therefore not getting enough attention from teaching staff.
Learning_3	The system not respecting my individual learning pace and preferences.

Next, participants completed 15 pairwise comparisons between representative statements from six core AI ethics concepts: Privacy, Autonomy, Fairness, Accuracy, Transparency, and Learning Beneficence (Table.4). In each comparison, participants were asked to select which of the two presented statements they were more concerned about. Each item in Table 4 was a single overarching statement designed to broadly represent the core ethical concern it belonged to. The order of both rating items and pairings was randomised for each participant.

Finally, participants completed a brief demographic questionnaire reporting their gender, age range, major and degree of study, and self-evaluation of general AI knowledge.

Table 4: Ethical constructs and their representative statements of concern, for paired comparison ranking

Ethical Concept	Presented Statement
Privacy:	Students not controlling their private information held by the technology.
Autonomy:	Students not having choices about whether and how this technology is used.
Fairness:	Students not being treated equally and fairly by the technology.
Accuracy:	The technology not accurately determining when students need help.
Transparency:	Students not knowing how the technology evaluates their attention and emotions.
Learning Beneficence:	The technology not helping students to learn effectively.

3.2.2 Item justification

The items and statements used in Stage 2 reflect our review of salient ethical principles from established frameworks (Section 2.2.1). In the rating task in Table 3, each principle was represented by three specific statements. Internal consistency of the scales in Table 3 was assessed using Cronbach’s alpha [34], demonstrating high and acceptable internal consistency for *Privacy* (0.90), *Autonomy* (0.82), *Transparency* (0.84), and *Fairness* (0.72), which exceeded the recommended threshold of 0.70 [103]. *Accuracy* (0.68) and *Learning* (0.67) showed reliability levels close to the 0.70 threshold. For *Learning*, it is likely because the three items pull in different directions: *Learning_1* and *Learning_2* focus on risks of not using AI, while *Learning_3* concerns risks when AI is used. We therefore present the Learning dimension result at the item level rather than relying on mean scores [103].

3.2.3 Analysis

Participants’ concern levels across 18 rated items (Table.3) were presented using descriptive statistics. Pairwise comparisons (Table.4) were analysed with the Bradley-Terry Model [21] to estimate the probability of each concern being prioritised over another.

3.3 Participants

We recruited 141 participants via the crowdsourcing platform Prolific, with 9 excluded based on an attention check to ensure data quality. A power analysis suggested a minimum of 106 participants were required, based on a medium effect size [31] $f=0.35$ with a significance level of 0.05 at 80%. Participants were all university students in Australia, with English as their first language and a minimum approval rate of 95% on the Prolific platform. We did not screen participants for AI literacy, as the aim was to reflect the real-world diversity of AI understanding within higher education settings (however, they were asked to self-evaluate their general AI knowledge in the demographic questionnaire at the end). In our sample, 50% identified as male, 47% identified as female, and 3% identified as non-binary or third/gender. All participants were compensated for their time and contribution with a payment of GBP £13.64 per hour, matching the local minimum wage in Australia.

4 FINDINGS

Our final dataset included 132 participants aged 18~24 (51%), 25~35 (44%), 45~54 (5%), with none below 18 or above 55. All participants were current university students, including undergraduates (67%), postgraduates (17%), PhD (9%), and others such as honours and diplomas (7%). Majors taken were arts and humanities (10%), business (8%), education (4%), engineering (11%), computer science and IT (16%), health sciences (20%), natural sciences (5%), social sciences (15%), and other (11%). In terms of AI literacy, most participants reported basic (43%) or intermediate knowledge (43%), with fewer indicating advanced (11%), no knowledge (2%), or expert knowledge (1%).

4.1 Stage 1: Believed learning impacts and affective response towards AI sensing-intervention

We first present students' quantitative ratings from Stage 1 to address RQ1: *How do students perceive AI sensing and intervention systems with various features?* Hypothesis testing was based on two constructs: Beliefs about Learning and Affective Responses, each consisting of multiple items. For construct scores, items with negative valence (e.g., "distraction") were reverse-coded so that higher values consistently reflect more positive perceptions. Item-level scores were reported in their original direction to provide nuanced insights into specific perceptions.

4.1.1 Sensing-in-use vs. not-in-use (Hypothesis 1)

Table 5 shows Likert ratings when sensing was not-in-use (scenarios 1-2) versus sensing-in-use (scenarios 3-6).

Table 5. Effect of Sensing Use on Beliefs about Learning and Affective Responses ^a

	Sensing Not-In-Use		Sensing In-Use		Significance		
	Mean	SD	Mean	SD	U-statistic	P-value	Adjusted P-value
<i>Belief about Learning</i>	3.28	0.89	2.51	0.89	105309.50	0.00	0.00
personalisation	3.59	1.12	2.63	1.30	99703.50	0.00	0.00
efficiency	3.60	1.15	2.56	1.30	101653.00	0.00	0.00
control of pace	3.04	1.36	2.22	1.23	94830.00	0.00	0.00
grade performance	3.50	1.03	2.66	1.18	98763.50	0.00	0.00
distraction	2.97	1.17	3.90	1.13	39976.50	0.00	0.00
over-helping	3.08	1.20	3.12	1.16	69395.00	0.65	1.00
<i>Affective Responses</i>	2.85	1.07	1.96	0.95	36202.50	0.00	0.00
enjoyable	3.07	1.21	2.00	1.13	104535.50	0.00	0.00
confident	2.77	1.09	2.09	1.08	95773.00	0.00	0.00
anxious	3.03	1.41	4.22	1.13	36894.50	0.00	0.00
uncomfortable	3.26	1.39	4.33	1.06	38535.00	0.00	0.00
worried	2.70	1.25	4.00	1.18	32420.50	0.00	0.00
fearing mistakes	3.01	1.34	3.82	1.17	46517.00	0.00	0.00

^a The response score is based on the Five-point Likert scale (1=Strongly disagree, 5=Strongly agree). Construct scores (shaded rows) were calculated by reverse-coding negatively worded items (e.g., 'distraction') so that higher values indicate more positive perceptions. Item-level scores are presented in their original direction.

In the sensing not-in-use condition, participants rated their beliefs on learning impacts relatively positively. They agreed that the intervention system would support *personalisation*, *efficiency* and *grade performance*, while holding neutral views on having *self-control of their learning pace*, being *distracted*, and receiving *too much help* from the system. Their affective responses were also generally neutral.

In the sensing-in-use condition, ratings shifted sharply negative on both believed learning impacts and affective responses. Participants disagreed that the system would support *personalisation*, *efficiency*, and *self-control of learning pace*. They deemed it *distracting*, finding the experience *unenjoyable* and not *confidence-boosting*, while anticipating feeling *anxious*, *uncomfortable*, *worried about being flagged as struggling*, and *afraid of making mistakes*.

Five of six learning beliefs items (*personalisation*, *efficiency*, *control of pace*, *grade performance*, *distraction*) and all six items from affective responses (*enjoyable*, *confident*, *anxious*, *uncomfortable*, *worried*, *fearing mistakes*) were significantly less positive when sensing was disclosed, with only *over-helping* unaffected. For example, perceptions on improved *efficiency* declined from Mean=3.60 (agree) in sensing not-in-use scenarios to Mean=2.56 (disagree) in sensing-in-use

scenarios. Disclosure of sensing led to negative responses on every item apart from *over-helping*. These results support *H1 - Students perceive AI systems that rely on sensing more negatively than systems that do not use sensing, both in terms of (a) potential learning impacts and (b) affective responses.*

4.1.2 Gaze-based attention detection vs. Facial-based emotion detection (Hypothesis 2)

Table 6 compares sensing approaches across scenarios 3–6 when students were informed that sensing was in operation.

Table 6: Effect of Sensing Approach on Beliefs about Learning and Affective Responses ^a

	Gaze Attention		Facial Emotion		Significance		
	Mean	SD	Mean	SD	U-statistic	P-value	Adjusted P-value
Belief about Learning	2.50	0.95	2.52	0.93	34814.40	0.75	1.00
personalisation	2.57	1.29	2.70	1.30	33503.50	0.27	1.00
efficiency	2.49	1.29	2.62	1.30	33385.00	0.25	1.00
control of pace	2.17	1.22	2.27	1.23	33643.00	0.31	1.00
grade performance	2.67	1.21	2.66	1.15	35443.50	0.97	1.00
distraction	3.85	1.13	3.94	1.12	33481.00	0.26	1.00
over-helping	3.05	1.17	3.20	1.15	32660.50	0.11	1.00
Affective Responses	1.94	0.93	1.97	0.96	34791.00	0.74	1.00
enjoyable	1.95	1.09	2.05	1.17	34034.00	0.42	1.00
confident	2.07	1.07	2.11	1.09	34726.50	0.70	1.00
anxious	4.24	1.11	4.19	1.16	36181.50	0.61	1.00
uncomfortable	4.34	1.06	4.31	1.07	35970.00	0.70	1.00
worried	4.00	1.18	4.00	1.17	35370.00	1.00	1.00
fearing mistakes	3.81	1.15	3.83	1.20	34583.00	0.64	1.00

^a The response score is based on the Five-point Likert scale (1=Strongly disagree, 5=Strongly agree). Construct scores (shaded rows) were calculated by reverse-coding negatively worded items (e.g., ‘distraction’) so that higher values indicate more positive perceptions. Item-level scores are presented in their original direction.

Across both conditions, participants reported relatively negative perceptions in both learning beliefs and affective dimensions. No significant differences were found between gaze-based attention detection and facial-based emotion detection on any of the items. Thus, we found no support for *H2 - Students perceive gaze-based attention detection more positively than facial-based emotion detection, both in terms of (a) potential learning impacts and (b) affective responses.*

4.1.3 System-generated vs. Human-mediated assistance (Hypothesis 3)

Table 7 compares intervention approaches when students were not informed that sensing was in operation (scenarios 1-2). Table 8 compares intervention approaches when sensing-in-use (scenarios 3-6).

In the sensing not-in-use condition (Table.7), system-generated hints were perceived significantly more positively than human teacher assistance for 3 items of believed learning impacts (*efficiency, control of pace, distraction*), with no difference for 3 items (*personalisation, grade performance, over-helping*). System-generated hints were also rated significantly more positively in terms of all the items of affective response. In the sensing-in-use condition (Table.8), system-generated hints again were rated significantly more positively than human teacher assistance, across nearly all items in believed learning impacts except *over-helping*, and all items in affective responses. These results support *H3 - Students perceive system-generated assistance more positively than human-delivered assistance, both in terms of (a) potential learning impacts and (b) affective responses.*

Table 7. Effect of Intervention Form (Sensing Not-In-Use Scenario) on Beliefs about Learning and Affective Responses ^a

Sensing Not-In-Use Scenario	Human Assist		System Hint		Significance		
	Mean	SD	Mean	SD	U-statistic	P-value	Adjusted P-value
Belief about Learning	3.09	0.92	3.47	0.82	6513.00	0.00	0.00
personalisation	3.41	1.24	3.77	0.97	7549.50	0.02	0.29
efficiency	3.29	1.25	3.91	0.95	6377.50	0.00	0.00
control of pace	2.48	1.29	3.59	1.19	4831.00	0.00	0.00
grade performance	3.44	1.19	3.56	0.96	8366.00	0.42	1.00
distraction	3.17	1.19	2.77	1.13	10604.00	0.00	0.05
over-helping	2.94	1.21	3.23	1.19	7734.50	0.07	0.83
Affective Responses	2.26	0.95	3.47	0.94	3104.00	0.00	0.00
enjoyable	2.62	2.00	3.52	1.06	5153.00	0.00	0.00
confident	2.45	1.00	3.10	1.09	5868.50	0.00	0.00
anxious	3.84	1.26	2.22	1.03	14538.50	0.00	0.00
uncomfortable	3.96	1.18	2.56	1.23	13897.00	0.00	0.00
worried	4.05	1.09	2.84	1.26	13457.50	0.00	0.00
fearing mistakes	3.68	1.22	2.33	1.08	13891.50	0.00	0.00

Table 8. Effect of Intervention Form (Sensing-In-Use Scenario) on Beliefs about Learning and Affective Responses ^a

Sensing-In-Use Scenario	Human Assist		System Hint		Significance		
	Mean	SD	Mean	SD	U-statistic	P-value	Adjusted P-value
Belief about Learning	2.29	0.86	2.73	0.97	25527.50	0.00	0.00
personalisation	2.36	1.24	2.90	1.30	27214.50	0.00	0.00
efficiency	2.23	1.20	2.88	1.31	25629.00	0.00	0.00
control of pace	1.86	1.04	2.58	1.29	24052.50	0.00	0.00
grade performance	2.49	1.18	2.84	1.15	29229.00	0.01	0.00
distraction	4.05	1.08	3.74	1.15	41088.00	0.01	0.00
over-helping	3.17	1.16	3.08	1.17	36966.50	0.36	1.00
Affective Responses	1.70	0.88	2.21	0.95	21834.50	0.00	0.00
enjoyable	1.73	1.02	2.27	1.17	25475.50	0.00	0.00
confident	1.89	1.00	2.30	1.12	27571.00	0.00	0.00
anxious	4.50	1.00	3.93	1.189	46439.50	0.00	0.00
uncomfortable	4.55	0.93	4.11	1.138	44194.50	0.00	0.00
worried	4.28	1.08	3.72	1.206	45604.00	0.0	0.00
fearing mistakes	4.10	1.11	3.54	1.169	45789.00	0.00	0.00

^a The response score is based on the Five-point Likert scale (1=Strongly disagree, 5=Strongly agree). Construct scores (shaded rows) were calculated by reverse-coding negatively worded items (e.g., 'distraction') so that higher values indicate more positive perceptions. Item-level scores are presented in their original direction.

A notable perception shift was observed when comparing intervention ratings across the sensing not-in-use condition (Table.7) and sensing-in-use condition (Table.8). As shown, perceptions of both intervention approaches worsened substantially once sensing was disclosed, across all items in learning impacts and affective responses (Appendix B presents a side-by-side comparison from synthesising Tables 7 and 8 to illustrate this shift).

When sensing was not-in-use, participants viewed system-generated hints relatively positively. They tended to agree that these hints supported *personalisation* (mean=3.77) and *efficiency* (mean=3.91). Regarding affective responses, they disagreed that such intervention would cause *anxiety* (mean=2.22), *discomfort* (mean=2.56) and *fear of making mistakes* (mean=2.33). However, once informed that sensing underpinned these interventions, evaluations worsened substantially on all items. System hints were now seen as more *distracting* (mean=3.74, changed from 2.77) and emotionally taxing, with sharp increases in *anxiety* (mean=3.93, changed from 2.22), *discomfort* (mean=4.11, changed from 2.56), and *fear of mistakes* (mean=3.54, changed from 2.33).

A similar pattern held for human teacher assistance. Initially rated neutral-to-negative when not using sensing, these ratings became even more negative when linked to sensing, with high levels of *anxiety* (mean=4.50) and *discomfort* (mean=4.55). Together, these results show that while students may accept intervention features on their own, once tied to sensing, they triggered strongly negative reactions, undermining both perceived learning benefits and emotional comfort.

4.2 Stage 1: Qualitative reports about learning impacts and affective responses to AI sensing-intervention

We next present qualitative findings from participants' open-ended responses, where the thematic analysis provides insights that help explain the quantitative preferences in Section 4.1.

4.2.1 Anxiety and discomfort about constant monitoring (N=78)

The most expressed concern was discomfort at constant observation by AI. Participants frequently reported feelings of "anxious" (P50), "uncomfortable" (P39), and "distracted" (P90). One called it "an invasion of personal space and autonomy" (P24). Participants voiced strong reactions to both sensing methods and their intrusiveness:

"That is horrible and dystopian, having a system constantly track and analyse my physical attributes would be extremely anxiety-inducing and stop my learning altogether. If this happened in my classes, I would start coming to class wearing masks and sunglasses to stop it, or I would not come to class at all" (P41)

4.2.2 Disrupted learning experience, learning autonomy, and control (N=66)

Participants felt that being monitored would distract them from learning. Constant sensing could cause pressure to constantly perform well or appear focused. With either attention or emotion detection, participants described feeling hyper-conscious of their appearance rather than concentrating on learning.

"I have to act and force myself to look more attentive, which is more distracting." (P90)

"Tracking eye gaze and emotions ... pressured me to be fully focused on my work for the entire class, which sometimes is just not possible." (P32)

Participants were dissatisfied with potential disruptions to the learning process from teacher assistance triggered automatically by the AI system. Many felt uncomfortable with additional social interactions with teachers when they were "not ready to be helped" (P128). They preferred self-control and autonomy over when and how assistance was provided, rather than passively receiving interventions initiated by a teacher.

"Having a computer assume I need help is a little patronising. It could be beneficial to suggest a hint, but the computer calling over a teacher for me is not okay... University students are adults and should be respected as such. I want to choose when I feel that I need help and ask for it myself." (P122)

"I like the system where it lets me (not the teacher) know that I am struggling and gives me hints. It would be even better if it asks me whether I need the hint or not, and then I can decide when to use the hint or whether to use the hint at all. Calling the teacher without asking me is distracting and feels forced." (P62)

Perceived disruptions to learning autonomy often added stress to learning:

"Alerting a teacher without my consent would make me feel very anxious as I usually try to get the task done alone" (P50).

"Having my teacher automatically notified every time I struggle or lose focus would be very distracting and anxiety-inducing." (P91)

4.2.3 Social dynamics and implications: peer perception, judgment, and embarrassment (N=25)

Another key concern was being identified as struggling or flagged for needing help in front of others, particularly teachers. Participants feared unwanted attention, embarrassment, and peer judgment. One said: "If everybody could see the teacher coming to help me whenever AI detected I was struggling, that has a certain social impact." (P68). Another worried it would "make it easier for peers to identify who's struggling and cause more bullying" (P93). Another said:

"I feel having the teacher notified whenever the AI thinks that I'm struggling is detrimental to the class and to my learning experience. Having the teacher attend to a particular student over and over again would draw the teacher's attention away from other students and could cause the student receiving ongoing assistance embarrassment or discomfort." (P36)

Students preferred keeping their emotional and learning status private. One suggested how the system could be improved:

"If it gathered classroom data on average completion time and offered a hint without anyone knowing, that would be the best method." (P31)

4.2.4 System inclusiveness and accommodating diverse learner needs (N=32)

Participants questioned whether the system could accommodate diverse learning styles, disabilities, cultural nuances, and neurodivergence. They worried it lacked flexibility and could misinterpret behaviours, disadvantaging those not conforming to expected patterns.

"I am on the autism spectrum—my learning style would trigger unnecessary attention. I often display unusual facial expressions and don't make normal eye contact." (P25)

"It's not inclusive. It will make some students, like those with ADHD, very embarrassed and more anxious because their behaviour might be misinterpreted by the AI." (P60)

"Facial recognition may be designed from a white Anglo-Saxon perspective." (P71)

Participants also worried about the system's inability to account for everyday emotional fluctuations, leading to misinterpretation of their emotional state and inappropriate intervention.

"We shouldn't be shamed for having a bad day. This system pressures us to be constantly alert and emotionally stable, which is unrealistic." (P55)

“My facial expressions don’t always show what I feel...if I am very focused, I could look upset/angry, especially to a computer.” (P122)

These qualitative findings help explain students’ negative reactions to sensing (H1) regardless of the two approaches (H2), which they found intrusive and disruptive to learning. Social dynamics surrounding teacher intervention explained students’ preference for system-generated hints (H3), which allowed them to retain control and avoid social consequences such as peer judgment and embarrassment at being publicly flagged for struggling. Overall, these themes highlight autonomy and privacy as central concerns. Many also raised broader ethical concerns around privacy and data misuse (44 participants) and accuracy in interpreting emotional or cognitive states (58 participants). We next explore ethics further.

4.3 Stage 3: Students’ ethical concerns

We now present findings relating to RQ2 (*How do students prioritise ethical concerns regarding AI sensing and intervention systems?*). Figure 2 plots the median Likert ratings and interquartile range for each statement of a possible ethical concern in Table 3. Overall, statements relating to Privacy (mean=4.14) and Autonomy (mean=4.10) garnered the highest level of concern from participants, while Fairness (mean=2.73) was rated the least concerning.

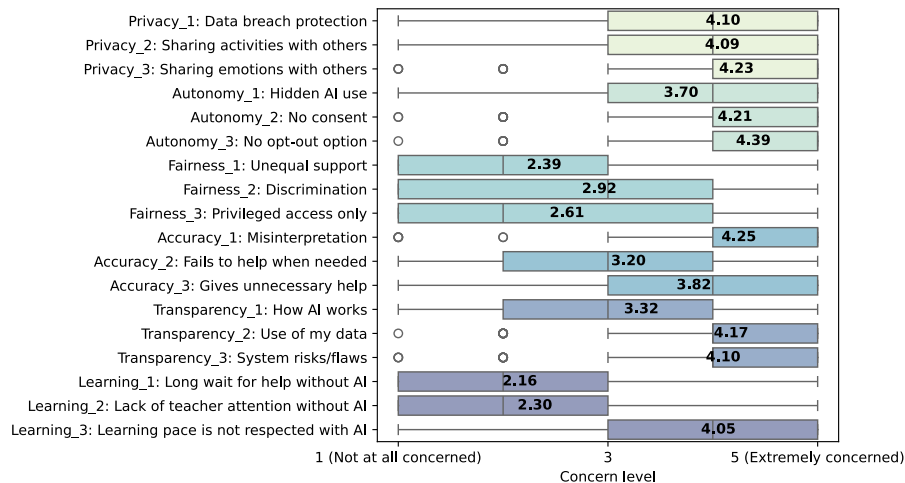


Figure 2. Mean ratings for statements of ethical concerns (see Table.3 for full item wording). The response score is based on the Five-point Likert scale rating from 1 (Not at all concerned) to 5 (Extremely Concerned).

There were greater variations in Learning, Transparency, and Accuracy. For Learning, concerns diverged depending on item framing: students were less worried about missing out on support without AI (e.g., Learning_1 and Learning_2, which are risks of not having AI), but expressed stronger concern that AI may not respect their individual learning pace (Learning_3, which is about a risk when AI is used). This suggests that students feared negative impacts when control over learning was ceded to the system more than they valued the potential benefits of AI. A similar pattern appeared in Transparency and Accuracy. Items reflecting system-centred performance with more distant concerns, such as understanding how the algorithm works (Transparency_1) or the system failing to recognise when help is needed (Accuracy_2), drew relatively lower concern and more variation. In contrast, items with immediate personal

consequences, such as not being informed about how data is used (Transparency_2), or misinterpreting behaviour and emotions (Accuracy_1), elicited consistently higher concern. These results together highlight that students' priorities focus less on the technical promises of AIED and more on the direct risks to their autonomy, privacy, and learning experience.

To better understand students' ethical priorities, we analysed the results from 1980 (15*132) paired comparisons participants made between two of the six key ethical principles of Autonomy, Transparency, Privacy, Fairness, Accuracy, and Learning Beneficence (Table.4). We first calculated the total 'wins' for each concept across each forced pairwise choice, which indicates how often this concept was prioritised as more important than another construct. The results reveal that Autonomy (N=515) and Privacy (N=462) received the highest total wins, indicating that participants placed the most importance on them. Transparency (N=269), Learning Beneficence (N=261), and Accuracy (N=248) scored lower, with Fairness (N=240) last. These results are consistent with the descriptive results from students' ratings.

To further quantify participants' preferences regarding specific ethical concerns, we applied the Bradley-Terry Model [21], which estimates the probability that one concept is more concerning than the alternative option and thus more likely to be selected. The probability of 'wins' confirmed that Autonomy (0.39) and Privacy (0.27) were by far the most important concerns, while Transparency (0.09), Learning Beneficence (0.09), Accuracy (0.08), and Fairness (0.08) were consistently less prioritised.

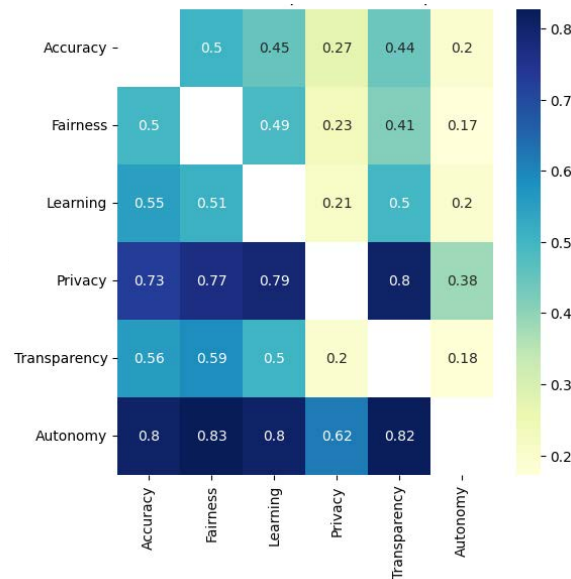


Figure 3. Ethical construct pairwise comparison heat map. Darker shades indicate a higher probability that the construct on the y-axis was selected as more concerning than the construct on the x-axis (See Table.4 for full ethical construct wording).

To visualise these preferences, Figure 3 presents a pairwise comparison heat map, showing the probability that one ethical construct was prioritised over another. The results confirm that Autonomy and Privacy were consistently judged as a higher ethical concern relative to Fairness, Transparency, Accuracy, and Learning Beneficence. It should be noted that even though Autonomy and Privacy are clearly more important to participants, the highest probability of wins in this context was 39%, indicating that no single concern completely overshadows the others within our participant sample.

4.4 Summary of Findings

Overall, the findings reveal that sensing causes strong negative affective responses and beliefs about learning experiences (H1 supported), with no significant differences between gaze-based attention detection and facial-based emotion detection (H2 rejected). Furthermore, the qualitative data consistently find discomfort, anxiety, and perceived disruption of learning caused by monitoring. Students expressed clear preferences for system-generated hints over human teacher assistance (H3 supported), largely due to a desire for agency and social concerns. Ethically speaking, students were most concerned about autonomy and privacy, suggesting that students highly value control over their data and learning process in AIEd interactions.

5 DISCUSSION

We now discuss the significance of our study findings for AI sensing-intervention in terms of student affective responses, beliefs about learning impacts (5.1), and ethical concerns (5.2). Implications for the design and deployment of these new technologies follow (5.3).

5.1 Possibilities and challenges regarding AI sensing-intervention systems

Students recognised potential benefits of AI systems, especially if private and non-intrusive. This was most evident in their preference for system-generated hints over teacher-initiated assistance, reflected consistently in both quantitative and qualitative findings. In scenarios without intrusive sensing, students leaned towards agreeing that the system could provide personalised support and render learning easier. These results highlight the educational value of such AI.

Yet these perceived benefits proved fragile once students realised that interventions derived from sensing. Positive perceptions shifted to neutral or negative, with both emotion and attention detection provoking discomfort, anxiety, and beliefs that the system would be distracting. Thus, the very mechanism of sensing that promises personalised, responsive teaching also risks undermining students' learning experiences [56, 59]. AIEd design must therefore focus on preserving the benefits of intervention while avoiding intrusive sensing by exploring alternative sources of insight to guide learning support (see also Section 5.3).

Students also resisted teachers being summoned automatically, preferring control over when and how help was triggered. This reflected a strong need for learning autonomy [59]. Students also feared unintended social consequences, such as embarrassment when teachers are alerted. While sensing technology might enhance teacher awareness of students' learning [65, 141] and emotional status [44, 124], our participants framed such visibility as undermining privacy and autonomy. This again underscores the importance of situating AIEd within the classroom's social dynamics, not just its technical potential [57, 123].

Our results suggest that affective responses and learning beliefs tended to align. Positive emotions corresponded with more positive learning perceptions, and vice versa. This resonates with theories that learning is partly an emotional experience [116] and that students' feelings in learning processes strongly shape engagement and outcomes [80, 116]. AIEd designs that foster positive emotional experiences may therefore enhance students' motivation and learning [80].

5.2 Shape of students' ethical concerns

Students expressed clear ethical concerns about AI sensing-intervention systems. Ranked highest were the lack of consent before deployment and opt-out options (autonomy), followed by not being informed of data use and associated risks (one kind of transparency), and having learning struggles or emotions revealed to others (privacy). Concerns about erroneous interpretation of behaviour and emotions (accuracy) were moderate, while understanding how algorithms

work (another kind of transparency) and fairness issues, such as algorithmic bias, evoked the least concern. Pairwise comparisons reinforced this hierarchy, with students prioritising autonomy and privacy over potential learning benefits.

The prominence of consent and opt-out options reflect not only students' strong desire for greater agency in how technology is integrated into their learning, but also their foundational role: without being informed and able to choose, students cannot meaningfully evaluate or respond to other risks. Meanwhile, our measures captured only one aspect of autonomy – the ability to accept or decline system use. As both prior research [59] and our qualitative findings emphasise, autonomy in education also encompasses broader control over what and how one studies, and the ability to make learning choices without undue direction from technologies [59]. This aligns with higher education students' broader expectations of independence [136], and their preference for agency in AI-assisted classrooms [141]. Our participants echoed these wishes by suggesting improvements like making hints optional and allowing control over teacher requests.

Privacy concerns extended beyond data-centric issues [62] to social privacy. Students were particularly uneasy about learning struggles or emotional states being revealed to peers and teachers (Privacy_3), a concern rated even higher than data breach protection (Privacy_1). This resonates with prior HCI findings that technology use can create unexpected breaches of personal boundaries in everyday contexts [110, 121, 123], again highlighting the need to account for the social contexts of technology use [123].

Fairness and equal treatment received lower priority, which is somewhat surprising given well-documented risks of algorithmic bias (e.g., [12, 16, 43]) causing racial and gender discrimination (e.g., [25]). One potential explanation is students' limited awareness of this problem [112]. However, it is also possible that students were more focused on avoiding sensing elements altogether, prioritising opting out over equal access and treatment. Nonetheless, participants did raise inclusiveness concerns, particularly regarding neurodiversity. The fact that fairness was rated lower does not imply it was unimportant, but that other issues felt more pressing.

Taken together, students' ethical priorities around autonomy and privacy directly shape their negative responses to sensing and teacher-initiated interventions, and explain their preference for private, system-generated hints. These links between ethical considerations and affective and learning needs suggest design implications.

5.3 Design Implications for AI sensing-intervention in education

We distil students' preferences and concerns into five design directions to guide the development of student-centred AIED systems. The aim is to support student autonomy, privacy, and well-being, while maintaining the pedagogical effectiveness of AIED.

1. *Minimise intrusiveness in sensing.* Given students' aversion to both gaze-based attention and facial-based emotion detection, designers should prioritise alternative, less invasive ways to gauge learning progress. For instance, tracking time taken to complete tasks or responses to in-platform questions may be preferable. Assuming sensing technologies are used in learning and teaching, one-on-one individualised sensing deployed on personal devices may be preferable to classroom-wide sensing. Sensing on individual devices like laptops could allow students to opt in (e.g., choose whether to install or use the application) and configure how and when sensing is enabled. Such individual sensing would offer greater autonomy than indiscriminate classroom-wide sensing, which makes opting-out difficult.

2. *Refining system-delivered interventions.* Students generally favoured system-generated hints, but only when these preserved autonomy and comfort. Hints should therefore be student-triggered, lightweight, and customisable. Students wanted the option to decide when to receive support rather than having it imposed automatically. This suggests value in offering 'opt-in' prompts, allowing students to accept or dismiss the hints. Hints should be delivered unobtrusively, but targeted to aid progress without being overwhelming or distracting. Students should be able to adjust intervention

frequency and style, thus retaining control. Finally, giving students transparency about why hints appear (e.g., “You’ve been on this step for a while, would you like a tip?”) may help build trust and reduce feelings of surveillance [56].

3. *Redesigning teacher-mediated interventions.* Although students expressed discomfort with automatic teacher involvement, this intervention approach could be redesigned rather than discarded. First, students wanted control over when teachers were notified, a preference that runs counter to a current trend in AIEd where systems typically notify teachers directly (e.g., [33, 64]). Instead, systems could preserve and enhance student agency by enabling students to request teacher help themselves or to explicitly consent beforehand. Second, interventions could be less visible to peers. For example, a private digital chat or a subtle in-platform message rather than a teacher walking over in class seems desirable. Third, teachers could be provided with aggregated or anonymised signals (e.g., class-level trends of confusion) rather than individual flags, so they can adjust teaching without singling out students. Finally, when one-on-one support is necessary, teachers should be trained to frame interventions as collaborative offers of help rather than corrections, to reduce peer embarrassment.

4. *Dynamic consent and transparency.* Students sought agency not just at the start (opt-in/opt-out), but dynamically throughout use. Instead of a one-off consent form, systems could implement layered consent prompts at key moments, such as starting a difficult task (e.g., “Do you want sensing enabled during this task?”, “Would you like your observed attentional status to be shared with your teacher?”). This dynamic consent approach respects students’ varying comfort levels in different learning situations and reinforces trust in the system.

5. *Provide private feedback channels before involving teachers.* The system could first signal to students privately (e.g., “You may be struggling, would you like a tip?”) before alerting teachers (e.g., “Would you like your teacher to help?”). This gives students a buffer stage to self-correct, accept a hint, or decide whether the teacher gets involved. This two-step approach could address both privacy and autonomy concerns while still enabling responsive interventions.

6 LIMITATIONS AND FUTURE WORK

This study used video scenarios to elicit perceptions of AI sensing-intervention during puzzle tasks. While effective for capturing reactions, we did not fully simulate the experience of wrestling with a learning activity. Future work could employ actual systems to gather richer responses.

Our sample focused on Australian higher education students. Expectations, cultural norms, and ethical concerns may vary across countries, educational systems, and age groups. Broader studies, including those of K-12 learners, could examine global and contextual differences.

Finally, our study reveals students’ immediate reactions to hypothetical scenarios. The long-term effects of using AI systems in education, such as changes in behaviour or attitudes over time, remain unexplored. Future research might consider longitudinal studies to assess how students’ experiences evolve with continued use. Nonetheless, the discomfort and ethical concerns we observed provide valuable guidance for developing technology that fosters students’ positive attitudes and better aligns with their needs, values, and overall well-being in the long term.

7 CONCLUSION

This study investigated student perceptions and reactions to AI sensing-intervention used in higher education, focusing on affective responses, beliefs about learning impacts, and ethical concerns. While students valued the potential for personalised support, they consistently reacted negatively to gaze-based attention and facial-based emotion detection, citing anxiety, discomfort, and threats to autonomy and privacy. Preferences leaned toward system-delivered hints, provided these were private, lightweight, and under student control. Ethical priorities, especially autonomy and privacy,

outweighed perceived learning benefits, with fairness, transparency, and accuracy considered secondary. These findings underscore that AIED design must move beyond technical optimisation to embed agency, privacy, and social sensitivity at its core. More broadly, this study contributes an empirical approach for examining student attitudes and value trade-offs in AIED, offering guidance for creating systems that are both pedagogically effective and ethically robust.

REFERENCE

- [1] Karan Ahuja, Dohyun Kim, Francesca Xhakaj, Virag Varga, Anne Xie, Stanley Zhang, Jay Eric Townsend, Chris Harrison, Amy Ogan and Yuvraj Agarwal. 2019. Edusense: Practical Classroom Sensing at Scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3, 1-26.
- [2] Karan Ahuja, Deval Shah, Sujeath Paredy, Francesca Xhakaj, Amy Ogan, Yuvraj Agarwal and Chris Harrison. 2021. Classroom Digital Twins with Instrumentation-Free Gaze Tracking. In *Proceedings of Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama, Japan, Article 484. <http://dx.doi.org/10.1145/3411764.3445711>
- [3] Ecenaz Alemdag and Kursat Cagiltay. 2018. A Systematic Review of Eye Tracking Research on Multimedia Learning. *Computers & Education* 125, 413-428.
- [4] Bryan Alexander, Kevin Ashford-Rowe, Noreen Barajas-Murph, Gregory Dobbin, Jessica Knott, Mark McCormack, Jeffery Pomerantz, Ryan Seilhamer and Nicole Weber. 2019. Horizon Report 2019 Higher Education Edition. 2019.
- [5] Cristina Almaraz-López, Fernando Almaraz-Menéndez and Carmen López-Esteban. 2023. Comparative Study of the Attitudes and Perceptions of University Students in Business Administration and Management and in Education toward Artificial Intelligence. *Education Sciences* 13, 6, 609.
- [6] Nese Alyuz, Eda Okur, Utiku Genc, Sinem Aslan, Cagri Tanriover and Asli Arslan Esme. 2017. An Unobtrusive and Multimodal Approach for Behavioral Engagement Detection of Students. In *Proceedings of Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education*. Association for Computing Machinery, Glasgow, UK, 26-32. <http://dx.doi.org/10.1145/3139513.3139521>
- [7] Dana Alzoubi, Jameel Kelley, Evrim Baran, Stephen B. Gilbert, Aliye Karabulut Ilgu and Shan Jiang. 2021. Teachactive Feedback Dashboard: Using Automated Classroom Analytics to Visualize Pedagogical Strategies at a Glance. In *Proceedings of Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama, Japan, Article 312. <http://dx.doi.org/10.1145/3411763.3451709>
- [8] Pengcheng An, Saskia Bakker, Sara Ordanovski, Ruurd Taconis, Chris LE Paffen and Berry Eggen. Year. Unobtrusively Enhancing Reflection-in-Action of Teachers through Spatially Distributed Ambient Information. In *Proceedings of Proceedings of the 2019 CHI conference on human factors in computing systems*. 1-14.
- [9] Cecilia Aragon, Clayton Hutto, Andy Echenique, Brittany Fiore-Gartland, Yun Huang, Jinyoung Kim, Gina Neff, Wanli Xing and Joseph Bayer. Year. Developing a Research Agenda for Human-Centered Data Science. In *Proceedings of Proceedings of the 19th ACM conference on computer supported cooperative work and social computing companion*. 529-535.
- [10] Alejandro Barredo Arrieta, Natalia Diaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina and Richard Benjamins. 2020. Explainable Artificial Intelligence (Xai): Concepts, Taxonomies, Opportunities and Challenges toward Responsible Ai. *Information fusion* 58, 82-115.
- [11] Ivon Arroyo, David G Cooper, Winslow Burleson, Beverly Park Woolf, Kasia Muldner and Robert Christopherson. Year. Emotion Sensors Go to School. In *Proceedings of Artificial intelligence in education*. Ios Press, 17-24.
- [12] T. S. Ashwin and Gautam Biswas. Year. Identifying And mitigating Algorithmic Bias In student Emotional Analysis. In *Proceedings of Springer Nature Switzerland*, 89-103.
- [13] Sinem Aslan, Nese Alyuz, Cagri Tanriover, Sinem E Mete, Eda Okur, Sidney K D'Mello and Asli Arslan Esme. Year. Investigating the Impact of a Real-Time, Multimodal Student Engagement Analytics Technology in Authentic Classrooms. In *Proceedings of Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1-12.
- [14] Ryan S Baker and Aaron Hawn. 2022. Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education* 32, 4, 1052-1092.
- [15] Toby Baker and Laurie Smith. 2019. Educ-Ai-Tion Rebooted? Exploring the Future of Artificial Intelligence in Schools and Colleges. Retrieved from https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf.
- [16] BBC. 2018. Amazon Scrapped 'Sexist Ai' Tool. Retrieved from <https://www.bbc.com/news/technology-45809919>.
- [17] Anne Beaudry and Alain Pinsonneault. 2010. The Other Side of Acceptance: Studying the Direct and Indirect Effects of Emotions on Information Technology Use. *MIS Quarterly* 34, 4, 689-710. <http://dx.doi.org/10.2307/25750701>
- [18] Melissa Bond, Hassan Khosravi, Maarten De Laat, Nina Bergdahl, Violeta Negrea, Emily Oxley, Phuong Pham, Sin Wang Chong and George Siemens. 2024. A Meta Systematic Review of Artificial Intelligence in Higher Education: A Call for Increased Ethics, Collaboration, and Rigour. *International Journal of Educational Technology in Higher Education* 21, 1, 4.
- [19] Alan Borning and Michael Muller. Year. Next Steps for Value Sensitive Design. In *Proceedings of Proceedings of the SIGCHI conference on human factors in computing systems*. 1125-1134.
- [20] Nigel Bosch, Yuxuan Chen and Sidney D'Mello. Year. It's Written on Your Face: Detecting Affective States from Facial Expressions While Learning Computer Programming. In *Proceedings of International Conference on Intelligent Tutoring Systems*. Springer, 39-44.
- [21] Ralph Allan Bradley and Milton E Terry. 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* 39, 3/4, 324-345.
- [22] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative research in psychology* 3, 2, 77-101.
- [23] Kathleen T Brinko. 1993. The Practice of Giving Feedback to Improve Teaching: What Is Effective? *The Journal of Higher Education* 64, 5, 574-593.
- [24] Christine Buchanan, M Lyndsay Howitt, Rita Wilson, Richard G Booth, Tracie Risling and Megan Bamford. 2021. Predicted Influences of Artificial Intelligence on Nursing Education: Scoping Review. *JMIR nursing* 4, 1, e23933.
- [25] Joy Buolamwini and Timnit Gebru. Year. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of Conference on fairness, accountability and transparency*. PMLR, 77-91.
- [26] Jake Chanenson, Brandon Sloane, Navaneeth Rajan, Amy Morril, Jason Chee, Danny Yuxing Huang and Marshini Chetty. Year. Uncovering Privacy and Security Challenges in K-12 Schools. In *Proceedings of Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-28.

- [27] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O'connell, Terrance Gray, F Maxwell Harper and Haiyi Zhu. Year. Explaining Decision-Making Algorithms through Ui: Strategies to Help Non-Expert Stakeholders. In *Proceedings of Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1-12.
- [28] Sribala Vidyadhari Chinta, Zichong Wang, Zhipeng Yin, Nhat Hoang, Matthew Gonzalez, Tai Le Quy and Wenbin Zhang. 2024. Fairaid: Navigating Fairness, Bias, and Ethics in Educational Ai Applications. *arXiv preprint arXiv:2407.18745*.
- [29] Hui-Chun Chu, Gwo-Haur Hwang, Yun-Fang Tu and Kai-Hsiang Yang. 2022. Roles and Research Trends of Artificial Intelligence in Higher Education: A Systematic Review of the Top 50 Most-Cited Articles. *Australasian Journal of Educational Technology* 38, 3, 22-42.
- [30] Simon Coghlan, Tim Miller and Jeannie Paterson. 2021. Good Proctor or "Big Brother"? Ethics of Online Exam Supervision Technologies. *Philosophy & Technology* 34, 4, 1581-1606.
- [31] Jacob Cohen. 2013. *Statistical Power Analysis for the Behavioral Sciences*. routledge.
- [32] Deborah R Compeau and Christopher A Higgins. 1995. Computer Self-Efficacy: Development of a Measure and Initial Test. *MIS quarterly*, 189-211.
- [33] Scotty Craig, Arthur Graesser, Jeremiah Sullins and Barry Gholson. 2004. Affect and Learning: An Exploratory Look into the Role of Affect in Learning with Autotutor. *Journal of educational media* 29, 3, 241-250.
- [34] Lee J Cronbach. 1951. Coefficient Alpha and the Internal Structure of Tests. *psychometrika* 16, 3, 297-334.
- [35] Sidney D'Mello and Art Graesser. 2012. Dynamics of Affective States During Complex Learning. *Learning and Instruction* 22, 2, 145-157.
- [36] Sidney D'Mello, Andrew Olney, Claire Williams and Patrick Hays. 2012. Gaze Tutor: A Gaze-Reactive Intelligent Tutoring System. *International Journal of human-computer studies* 70, 5, 377-398.
- [37] Fred D Davis. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS quarterly*, 319-340.
- [38] Fred D Davis, Richard P Bagozzi and Paul R Warshaw. 1992. Extrinsic and Intrinsic Motivation to Use Computers in the Workplace 1. *Journal of applied social psychology* 22, 14, 1111-1132.
- [39] Edward Dieterle, Chris Dede and Michael Walker. 2024. The Cyclical Ethical Effects of Using Artificial Intelligence in Education. *AI & society* 39, 2, 633-643.
- [40] Martin Ebers. 2023. The European Commission's Proposal for an Artificial Intelligence Act. In *Research Handbook on Eu Internet Law*, Edward Elgar Publishing, 271-292.
- [41] Autumn Edwards, Chad Edwards, Patric R Spence, Christina Harris and Andrew Gambino. 2016. Robots in the Classroom: Differences in Students' Perceptions of Credibility and Learning between "Teacher as Robot" and "Robot as Teacher". *Computers in Human Behavior* 65, 627-634.
- [42] Chad Edwards, Autumn Edwards, Fatima Albrehi and Patric Spence. 2021. Interpersonal Impressions of a Social Robot Versus Human in the Context of Performance Evaluations. *Communication Education* 70, 2, 165-182.
- [43] Virginia Eubanks. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press,
- [44] Mohamed Ez-Zaouia, Aurélien Tabard and Elise Lavoué. 2020. Emodash: A Dashboard Supporting Retrospective Awareness of Emotions in Online Learning. *International Journal of Human-Computer Studies* 139, 102411.
- [45] Taciana Pontual Falcão, Rafael Ferreira Mello, Rodrigo Lins Rodrigues, Juliana Regueira Basto Diniz, Yi-Shan Tsai and Dragan Gašević. Year. Perceptions and Expectations About Learning Analytics from a Brazilian Higher Education Institution. In *Proceedings of Proceedings of the tenth international conference on learning analytics & knowledge*. 240-249.
- [46] Kenneth R Fleischmann, Sherri R Greenberg, Danna Gurari, Abigale Stangl, Nitin Verma, Jaxsen R Day, Rachel N Simons and Tom Yeh. Year. Good Systems: Ethical Ai for Cscw. In *Proceedings of Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing*. 461-467.
- [47] Luciano Floridi and Josh Cows. 2022. A Unified Framework of Five Principles for Ai in Society. *Machine learning and the city: Applications in architecture and urban design*, 535-545.
- [48] Luciano Floridi, Josh Cows, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo and Francesca Rossi. 2018. Ai4people—an Ethical Framework for a Good Ai Society: Opportunities, Risks, Principles, and Recommendations. *Minds and machines* 28, 689-707.
- [49] Guillaume Fontaine, Sylvie Cossette, Marc-André Maheu-Cadotte, Tanya Mailhot, Marie-France Deschênes, Gabrielle Mathieu-Dupuis, José Côté, Marie-Pierre Gagnon and Veronique Dubé. 2019. Efficacy of Adaptive E-Learning for Health Professionals and Students: A Systematic Review and Meta-Analysis. *BMJ open* 9, 8, e025252.
- [50] Jodi Forlizzi and Katja Battarbee. Year. Understanding Experience in Interactive Systems. In *Proceedings of Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*. 261-268.
- [51] Dragan Gasevic, Jelena Jovanovic, Abelardo Pardo and Shane Dawson. 2017. Detecting Learning Strategies with Analytics: Links with Self-Reported Measures and Academic Performance. *Journal of Learning Analytics* 4, 2, 113-128-113-128.
- [52] Victor González-Calatayud, Paz Prendes-Espinosa and Rosabel Roig-Vila. 2021. Artificial Intelligence for Student Assessment: A Systematic Review. *Applied Sciences* 11, 12, 5467. <http://dx.doi.org/10.3390/app11125467>
- [53] Geraldine Gray, Ana Elena Schalk, Gordon Cooke, Phelim Murnion, Pauline Rooney and Kevin O'Rourke. 2022. Stakeholders' Insights on Learning Analytics: Perspectives of Students and Staff? *Computers & Education*, 104550.
- [54] Eetu Haataja, Jonna Malmberg and Sanna Järvelä. 2018. Monitoring in Collaborative Learning: Co-Occurrence of Observed Behavior and Physiological Synchrony Explored. *Computers in Human Behavior* 87, 337-347.
- [55] Laura Hakimi, Rebecca Eynon and Victoria A Murphy. 2021. The Ethics of Using Digital Trace Data in Education: A Thematic Review of the Research Landscape. *Review of Educational Research*, 00346543211020116.
- [56] Bingyi Han, George Buchanan and Dana McKay. 2023. Learning in the Panopticon: Examining the Potential Impacts of Ai Monitoring on Students. In *Proceedings of Proceedings of the 34th Australian Conference on Human-Computer Interaction*. Association for Computing Machinery, Canberra, ACT, Australia, 9-21. <http://dx.doi.org/10.1145/3572921.3572937>
- [57] Bingyi Han, Simon Coghlan, George Buchanan and Dana McKay. 2025. Who Is Helping Whom? Student Concerns About Ai-Teacher Collaboration in Higher Education Classrooms. *Proceedings of the ACM on Human-Computer Interaction* 9, 2, 1-32.
- [58] Bingyi Han, Sadia Nawaz, George Buchanan and Dana McKay. Year. Ethical and Pedagogical Impacts of Ai in Education. In *Proceedings of International Conference on Artificial Intelligence in Education*. Springer, 667-673.
- [59] Bingyi Han, Sadia Nawaz, George Buchanan and Dana McKay. 2025. Students' Perceptions: Exploring the Interplay of Ethical and Pedagogical Impacts for Adopting Ai in Higher Education. *International Journal of Artificial Intelligence in Education*. <http://dx.doi.org/10.1007/s40593-024-00456-4>
- [60] Wayne Holmes, Francisco Iniesto, Stamatina Anastopoulou and Jesus G Boticario. 2023. Stakeholder Perspectives on the Ethics of Ai in Distance-Based Higher Education. *International Review of Research in Open and Distributed Learning* 24, 2, 96-117.

- [61] Wayne Holmes and Kaška Porayska-Pomsta. 2022. *The Ethics of Artificial Intelligence in Education: Practices, Challenges, and Debates*. Taylor & Francis.
- [62] Wayne Holmes, Kaska Porayska-Pomsta, Ken Holstein, Emma Sutherland, Toby Baker, Simon Buckingham Shum, Olga C Santos, Mercedes T Rodrigo, Mutlu Cukurova and Ig Ibert Bittencourt. 2021. Ethics of Ai in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, 1-23.
- [63] Kenneth Holstein, Vincent Aleven and Nikol Rummel. Year. A Conceptual Framework for Human-Ai Hybrid Adaptivity in Education. In *Proceedings of Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part I* 21. Springer, 240-254.
- [64] Kenneth Holstein, Bruce M. McLaren and Vincent Aleven. Year. Student Learning Benefits of a Mixed-Reality Teacher Awareness Tool in Ai-Enhanced Classrooms. In *Proceedings of Springer International Publishing*, 154-168.
- [65] Kenneth Holstein, Bruce M. McLaren and Vincent Aleven. 2019. Designing for Complementarity: Teacher and Student Needs for Orchestration Support in Ai-Enhanced Classrooms Springer International Publishing, 157-171. http://dx.doi.org/10.1007/978-3-030-23204-7_14
- [66] Stephen Hutt, Kristina Krasich, James R. Brockmole and Sidney K. D'Mello. Year. Breaking out of the Lab: Mitigating Mind Wandering with Gaze-Based Attention-Aware Technology in Classrooms. In *Proceedings of Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [67] Yeonju Jang, Seongyune Choi and Hyeoncheol Kim. 2022. Development and Validation of an Instrument to Measure Undergraduate Students' Attitudes toward the Ethics of Artificial Intelligence (at-Eai) and Analysis of Its Difference by Gender and Experience of Ai Education. *Education and Information Technologies* 27, 8, 11635-11667.
- [68] Anna Jobin, Marcello Ienca and Effy Vayena. 2019. The Global Landscape of Ai Ethics Guidelines. *Nature machine intelligence* 1, 9, 389-399.
- [69] Kyle ML Jones, Andrew Asher, Abigail Gobin, Michael R Perry, Dorothea Salo, Kristin A Briney and M Brooke Robertshaw. 2020. "We're Being Tracked at All Times": Student Perspectives of Their Privacy in Relation to Learning Analytics in Higher Education. *Journal of the Association for Information Science and Technology* 71, 9, 1044-1059.
- [70] Ashish Kapoor, Winslow Burleson and Rosalind W Picard. 2007. Automatic Prediction of Frustration. *International journal of human-computer studies* 65, 8, 724-736.
- [71] Hilmi Karaca, Erhan Ertekin and Kursat Cagiltay. 2025. Investigating Middle School Students' Eye Movements on the Mathematical Representations: An Eye-Tracking Study. *Education and Information Technologies*, 1-22.
- [72] Argyrios Katsantonis and Ioannis Katsantonis. 2024. University Students' Attitudes toward Artificial Intelligence: An Exploratory Study of the Cognitive, Emotional, and Behavioural Dimensions of Ai Attitudes.
- [73] Sherrie-Anne Kaye, Ioni Lewis, Sonja Forward and Patricia Delhomme. 2020. A Priori Acceptance of Highly Automated Cars in Australia, France, and Sweden: A Theoretically-Informed Investigation Guided by the Tpb and Utaut. *Accident Analysis & Prevention* 137, 105441.
- [74] Kimon Kieslich, Birte Keller and Christopher Starke. 2021. Ai-Ethics by Design. Evaluating Public Perception on the Importance of Ethical Design Principles of Ai. *arXiv preprint arXiv:2106.00326*.
- [75] Jihyun Kim, Kelly Merrill, Kun Xu and Deanna D Sellnow. 2020. My Teacher Is a Machine: Understanding Students' Perceptions of Ai Teaching Assistants in Online Education. *International Journal of Human-Computer Interaction* 36, 20, 1902-1911.
- [76] Yelin Kim, Tolga Soyata and Reza Feyzi Behnagh. 2018. Towards Emotionally Aware Ai Smart Classroom: Current Issues and Directions for Engineering and Education. *IEEE Access* 6, 5308-5331.
- [77] Abirami Kirubakaran, Dylan Young, Shawn Khan, Noelle Crasto, Mara Sobel and Dafna Sussman. 2022. Artificial Intelligence and Surgical Education: A Systematic Scoping Review of Interventions. *Journal of Surgical Education* 79, 2, 500-515.
- [78] Ekaterina Kochmar, Dung Do Vu, Robert Belfer, Varun Gupta, Iulian Vlad Serban and Joelle Pineau. Year. Automated Personalized Feedback Improves Learning Gains in an Intelligent Tutoring System. In *Proceedings of international conference on artificial intelligence in education*. Springer, 140-146.
- [79] Kenneth R Koedinger, Emma Brunskill, Ryan Sjd Baker, Elizabeth A McLaughlin and John Stamper. 2013. New Potentials for Data-Driven Intelligent Tutoring System Development and Optimization. *AI Magazine* 34, 3, 27-41.
- [80] Jeya Amantha Kumar, Balakrishnan Muniandy and Wan Ahmad Jaafar Wan Yahaya. 2019. Exploring the Effects of Emotional Design and Emotional Intelligence in Multimedia-Based Learning: An Engineering Educational Perspective. *New Review of Hypermedia and Multimedia* 25, 1-2, 57-86.
- [81] Priya C Kumar, Marshini Chetty, Tamara L Clegg and Jessica Vitak. Year. Privacy and Security Considerations for Digital Technology Use in Elementary Schools. In *Proceedings of Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1-13.
- [82] Annabel Latham, Keeley Crockett, David McLean and Bruce Edmonds. 2012. A Conversational Intelligent Tutoring System to Automatically Predict Learning Styles. *Computers & Education* 59, 1, 95-109.
- [83] Annabel Latham and Sean Goltz. Year. A Survey of the General Public's Views on the Ethics of Using Ai in Education. In *Proceedings of Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, June 25-29, 2019, Proceedings, Part I* 20. Springer, 194-206.
- [84] LuEttamae Lawrence, Vanessa Echeverria, Kexin Yang, Vincent Aleven and Nikol Rummel. 2024. How Teachers Conceptualise Shared Control with an Ai Co-Orchestration Tool: A Multiyear Teacher-Centred Design Process. *British Journal of Educational Technology* 55, 3, 823-844.
- [85] Xiaotian Vivian Li, Mary Beth Rosson and Jenay Robert. Year. A Scenario-Based Exploration of Expected Usefulness, Privacy Concerns, and Adoption Likelihood of Learning Analytics. In *Proceedings of Proceedings of the Ninth ACM Conference on Learning@ Scale*. 48-59.
- [86] Hao-Chiang Koong Lin, Chih-Hung Wu and Ya-Ping Hsueh. 2014. The Influence of Using Affective Tutoring System in Accounting Remedial Instruction on Learning Performance and Usability. *Computers in Human Behavior* 41, 514-522.
- [87] Adam Linson, Yucheng Xu, Andrea R English and Robert B Fisher. Year. Identifying Student Struggle by Analyzing Facial Movement During Asynchronous Video Lecture Viewing: Towards an Automated Tool to Support Instructors. In *Proceedings of International Conference on Artificial Intelligence in Education*. Springer, 53-65.
- [88] Alex Jiahong Lu, Tawanna R Dillahun, Gabriela Marcu and Mark S Ackerman. 2021. Data Work in Education: Enacting and Negotiating Care and Control in Teachers' Use of Data-Driven Classroom Surveillance Technology. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2, 1-26.
- [89] Alex Jiahong Lu, Gabriela Marcu, Mark S. Ackerman and Tawanna R Dillahun. 2021. Coding Bias in the Use of Behavior Management Technologies: Uncovering Socio-Technical Consequences of Data-Driven Surveillance in Classrooms. In *Proceedings of Designing Interactive Systems Conference 2021*. Association for Computing Machinery, Virtual Event, USA, 508-522. <http://dx.doi.org/10.1145/3461778.3462084>
- [90] Rose Luckin, Wayne Holmes, Mark Griffiths and Laurie B Forcier. 2016. Intelligence Unleashed: An Argument for Ai in Education.
- [91] David Lundie. 2016. Authority, Autonomy and Automation: The Irreducibility of Pedagogy to Information Transactions. *Studies in philosophy and education* 35, 3, 279-291.

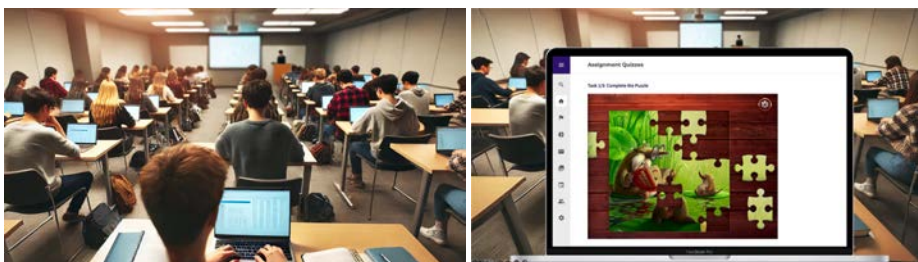
- [92] Qianru Lyu, Wenli Chen, Junzhu Su, Kok Hui John Gerard Heng and Shuai Liu. Year. How Peers Communicate without Words-an Exploratory Study of Hand Movements in Collaborative Learning Using Computer-Vision-Based Body Recognition Techniques. In *Proceedings of Artificial Intelligence in Education*. Springer Nature Switzerland, 316-326.
- [93] Roberto Martinez-Maldonado, Andrew Clayphan and Judy Kay. 2015. Deploying and Visualising Teacher's Scripts of Small Group Activities in a Multi-Surface Classroom Ecology: A Study in-the-Wild. *Computer Supported Cooperative Work (CSCW)* 24, 2, 177-221. <http://dx.doi.org/10.1007/s10606-015-9217-6>
- [94] Roberto Martinez-Maldonado, Vanessa Echeverria, Gloria Fernandez-Nieto, Lixiang Yan, Linxuan Zhao, Riordan Alfredo, Xinyu Li, Samantha Dix, Hollie Jaggard and Rosie Wotherspoon. 2023. Lessons Learnt from a Multimodal Learning Analytics Deployment in-the-Wild. *ACM Transactions on Computer-Human Interaction* 31, 1, 1-41.
- [95] Bahar Memarian and Tenzin Doleck. 2023. Fairness, Accountability, Transparency, and Ethics (Fate) in Artificial Intelligence (Ai), and Higher Education: A Systematic Review. *Computers and Education: Artificial Intelligence*, 100152.
- [96] Danaë Metaxa-Kakavouli, Kelly Wang, James A Landay and Jeff Hancock. Year. Gender-Inclusive Design: Sense of Belonging and Bias in Web Interfaces. In *Proceedings of Proceedings of the 2018 CHI Conference on human factors in computing systems*. 1-6.
- [97] Fengchun Miao, Wayne Holmes, Ronghuai Huang and Hui Zhang. 2021. *Ai and Education: A Guidance for Policymakers*. UNESCO Publishing,
- [98] Jessica Morley, Luciano Floridi, Libby Kinsey and Anat Elhalal. 2020. From What to How: An Initial Review of Publicly Available Ai Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and engineering ethics* 26, 4, 2141-2168.
- [99] Maria Moundridou and Maria Virvou. 2002. Evaluating the Persona Effect of an Interface Agent in a Tutoring System. *Journal of computer assisted learning* 18, 3, 253-261.
- [100] Chantal Mutimukwe, Olga Viberg, Lena-Maria Oberg and Teresa Cerratto-Pargman. 2022. Students' Privacy Concerns in Learning Analytics: Model Development. *British Journal of Educational Technology*.
- [101] Andy Nguyen, Ha Ngan Ngo, Yvonne Hong, Belle Dang and Bich-Phuong Thi Nguyen. 2023. Ethical Principles for Artificial Intelligence in Education. *Education and Information Technologies* 28, 4, 4221-4241.
- [102] Giovanna Nunes Vilaza, Kevin Doherty, Darragh McCashin, David Coyle, Jakob Bardram and Marguerite Barry. Year. A Scoping Review of Ethics across Sigchi. In *Proceedings of Proceedings of the 2022 ACM Designing Interactive Systems Conference*. 137-154.
- [103] Jum C Nunnally. 1978. *Psychometric Theory* (2nd Ed.).
- [104] Amy Ogan. 2019. Reframing Classroom Sensing: Promise and Peril. *Interactions* 26, 6, 26-32.
- [105] Prasoon Patidar, Tricia J Ngoon, John Zimmerman, Amy Ogan and Yuvraj Agarwal. 2024. Classid: Enabling Student Behavior Attribution from Ambient Classroom Sensing Systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2, 1-28.
- [106] Yanyi Peng, Masato Kikuchi and Tadachika Ozono. Year. Development and Experiment of Classroom Engagement Evaluation Mechanism During Real-Time Online Courses. In *Proceedings of International Conference on Artificial Intelligence in Education*. Springer, 590-601.
- [107] Filipe Dwan Pereira, Samuel C Fonseca, Elaine HT Oliveira, Alexandra I Cristea, Henrik Bellhäuser, Luiz Rodrigues, David BF Oliveira, Seiji Isotani and Leandro SG Carvalho. 2021. Explaining Individual and Collective Programming Students' Behavior by Interpreting a Black-Box Predictive Model. *IEEE Access* 9, 117097-117119.
- [108] Daniel Pinto dos Santos, Daniel Giese, Sebastian Brodehl, Seung-Hun Chon, W Staab, R Kleinert, D Maintz and B Baeßler. 2019. Medical Students' Attitude Towards Artificial Intelligence: A Multicentre Survey. *European radiology* 29, 1640-1646.
- [109] Michael I Posner and Mary K Rothbart. 2007. Research on Attention Networks as a Model for the Integration of Psychological Science. *Annu. Rev. Psychol.* 58, 1, 1-23.
- [110] Jenny Preece, Yvonne Rogers and Helen Sharp. 2002. *Interaction Design: Beyond Human-Computer Interaction*. John Wiley,
- [111] Paul Prinsloo and Sharon Slade. 2016. Student Vulnerability, Agency and Learning Analytics: An Exploration. *Journal of Learning Analytics* 3, 1, 159-182. <http://dx.doi.org/10.18608/jla.2016.3.1.10>
- [112] Shalini Ramachandran, Steven Matthew Cutchin and Sheree Fu. 2021. Raising Algorithm Bias Awareness among Computer Science Students through Library and Computer Science Instruction.
- [113] Lee Richstone, Michael J Schwartz, Casey Seideman, Jeffrey Cadeddu, Sandra Marshall and Louis R Kavoussi. 2010. Eye Metrics as an Objective Assessment of Surgical Skill. *Annals of surgery* 252, 1, 177-182.
- [114] Samantha Robertson, Tonya Nguyen and Niloufar Salehi. Year. Modeling Assumptions Clash with the Real World: Transparency, Equity, and Community Challenges for Student Assignment Algorithms. In *Proceedings of Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1-14.
- [115] Paul Rodway and Astrid Schepman. 2023. The Impact of Adopting Ai Educational Technologies on Projected Course Satisfaction in University Students. *Computers and Education: Artificial Intelligence* 5, 100150.
- [116] Jerry Rosiek. 2003. Emotional Scaffolding: An Exploration of the Teacher Knowledge at the Intersection of Student Emotion and the Subject Matter. *Journal of Teacher Education* 54, 5, 399-412.
- [117] Xingran Ruan, Charaka Palansuriya and Aurora Constantin. Year. Affective Dynamic Based Technique for Facial Emotion Recognition (Fer) to Support Intelligent Tutors in Education. In *Proceedings of International Conference on Artificial Intelligence in Education*. Springer, 774-779.
- [118] Clara Schumacher and Dirk Ifenthaler. 2018. Features Students Really Expect from Learning Analytics. *Computers in human behavior* 78, 397-407.
- [119] Donghee Shin. 2020. User Perceptions of Algorithmic Decisions in the Personalized Ai System: Perceptual Evaluation of Fairness, Accountability, Transparency, and Explainability. *Journal of Broadcasting & Electronic Media* 64, 4, 541-565.
- [120] Donghee Shin and Yong Jin Park. 2019. Role of Fairness, Accountability, and Transparency in Algorithmic Affordance. *Computers in Human Behavior* 98, 277-284.
- [121] Stephen Snow, Kenneth Radke, Dhaval Vyas and Margot Brereton. Year. Privacy in the New Era of Visible and Sharable Energy-Use Information. In *Proceedings of Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*. 342-351.
- [122] Christopher Starke, Janine Baleis, Birte Keller and Frank Marcinkowski. 2022. Fairness Perceptions of Algorithmic Decision-Making: A Systematic Review of the Empirical Literature. *Big Data & Society* 9, 2, 20539517221115189.
- [123] Lucy A Suchman. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge university press,
- [124] Gahyun Sung, Tianyi Feng and Bertrand Schneider. 2021. Learners Learn More and Instructors Track Better with Real-Time Gaze Sharing. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 134. <http://dx.doi.org/10.1145/3449208>
- [125] Lin Tang and Yu-Sheng Su. 2024. Ethical Implications and Principles of Using Artificial Intelligence Models in the Classroom: A Systematic Literature Review.
- [126] Arnout Terpstra, Alwin De Rooij and Alexander Schouten. Year. Online Proctoring: Privacy Invasion or Study Alleviation? Discovering Acceptability Using Contextual Integrity. In *Proceedings of Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-20.

- [127] Ronald L. Thompson, Christopher A Higgins and Jane M Howell. 1991. Personal Computing: Toward a Conceptual Model of Utilization. *MIS quarterly*, 125-143.
- [128] Yi-Shan Tsai, Carlo Perrotta and Dragan Gašević. 2020. Empowering Learners with Personalised Learning Approaches? Agency, Equity and Transparency in the Context of Learning Analytics. *Assessment & Evaluation in Higher Education* 45, 4, 554-567. <http://dx.doi.org/10.1080/02602938.2019.1676396>
- [129] Yi-Shan Tsai, Diego Rates, Pedro Manuel Moreno-Marcos, Pedro J Muñoz-Merino, Ioana Jivet, Maren Scheffel, Hendrik Drachsler, Carlos Delgado Kloos and Dragan Gašević. 2020. Learning Analytics in European Higher Education—Trends and Barriers. *Computers & Education* 155, 103933.
- [130] Yi-Shan Tsai, Alexander Whitelock-Wainwright and Dragan Gašević. 2020. The Privacy Paradox and Its Implications for Learning Analytics. In *Proceedings of Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. Association for Computing Machinery, Frankfurt, Germany, 230–239. <http://dx.doi.org/10.1145/3375462.3375536>
- [131] William Villegas-Ch, Diego Buenano-Fernandez, Alexandra Maldonado Navarro and Aracely Mera-Navarrete. 2025. Adaptive Intelligent Tutoring Systems for Stem Education: Analysis of the Learning Impact and Effectiveness of Personalized Feedback. *Smart Learning Environments* 12, 1, 1-31.
- [132] Zhifeng Wang, Longlong Li, Chunyan Zeng and Jialong Yao. 2023. Student Learning Behavior Recognition Incorporating Data Augmentation with Learning Feature Representation in Smart Classrooms. *Sensors* 23, 19, 8190.
- [133] David Watson, Lee Anna Clark and Auke Tellegen. 1988. Development and Validation of Brief Measures of Positive and Negative Affect: The Panas Scales. *Journal of personality and social psychology* 54, 6, 1063.
- [134] Jacob Whitehill, Zewelanjji Serpell, Yi-Ching Lin, Aysha Foster and Javier R Movellan. 2014. The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions. *IEEE Transactions on Affective Computing* 5, 1, 86-98.
- [135] Alexander Whitelock-Wainwright, Yi-Shan Tsai, Hendrik Drachsler, Maren Scheffel and Dragan Gašević. 2021. An Exploratory Latent Class Analysis of Student Expectations Towards Learning Analytics Services. *The Internet and Higher Education*, 100818.
- [136] Alexander Whitelock-Wainwright, Dragan Gašević, Ricardo Tejeiro, Yi-Shan Tsai and Kate Bennett. 2019. The Student Expectations of Learning Analytics Questionnaire. *Journal of Computer Assisted Learning* 35, 5, 633-666. <http://dx.doi.org/10.1111/jcal.12366>
- [137] Ben Williamson. 2019. Datafication of Education: A Critical Approach to Emerging Analytics Technologies and Practices. *Rethinking pedagogy for a digital age*, 212-226.
- [138] Billy Tak-Ming Wong, Kam Cheong Li and Simon K. S. Cheung. 2022. An Analysis of Learning Analytics in Personalised Learning. *Journal of Computing in Higher Education*. <http://dx.doi.org/10.1007/s12528-022-09324-3>
- [139] World-Economic-Forum. 2019. Generation Ai: Establishing Global Standards for Children and Ai. . Retrieved from <https://www.weforum.org/publications/generation-ai-establishing-global-standards-for-children-and-ai/>.
- [140] Lixiang Yan, Roberto Martinez-Maldonado, Linxuan Zhao, Xinyu Li and Dragan Gašević. Year. Physiological Synchrony And arousal As indicators Of stress And learning Performance In embodied Collaborative Learning. In *Proceedings of Artificial Intelligence in Education*. Springer Nature Switzerland, 602-614.
- [141] Kexin Bella Yang, Vanessa Echeverria, Zijiang Lu, Hongyu Mao, Kenneth Holstein, Nikol Rummel and Vincent Aleven. 2023. Pair-Up: Prototyping Human-Ai Co-Orchestration of Dynamic Transitions between Individual and Collaborative Learning in the Classroom. In *Proceedings of Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Hamburg, Germany, Article 453. <http://dx.doi.org/10.1145/3544548.3581398>
- [142] Mireia Yurrita, Tim Draws, Agathe Balayn, Dave Murray-Rust, Nava Tintarev and Alessandro Bozzon. Year. Disentangling Fairness Perceptions in Algorithmic Decision-Making: The Effects of Explanations, Human Oversight, and Contestability. In *Proceedings of Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1-21.
- [143] Olaf Zawacki-Richter, Victoria I. Marín, Melissa Bond and Franziska Gouverneur. 2019. Systematic Review of Research on Artificial Intelligence Applications in Higher Education – Where Are the Educators? *International Journal of Educational Technology in Higher Education* 16, 1. <http://dx.doi.org/10.1186/s41239-019-0171-0>
- [144] Tingru Zhang, Da Tao, Xingda Qu, Xiaoyan Zhang, Rui Lin and Wei Zhang. 2019. The Roles of Initial Trust and Perceived Risk in Public's Acceptance of Automated Vehicles. *Transportation research part C: emerging technologies* 98, 207-220.

APPENDICES

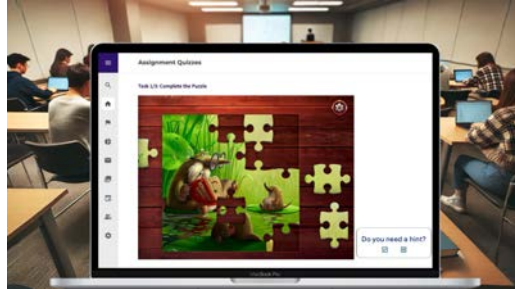
APPENDIX A. EXPERIMENTAL SCENARIOS, VISUAL MATERIALS AND TEXT INSTRUCTIONS

We set the scene for the context, telling participants the scenario is in the tutorial class and as students they are working on their own tasks. *Background picture and puzzle pictures were generated by DALL-E.



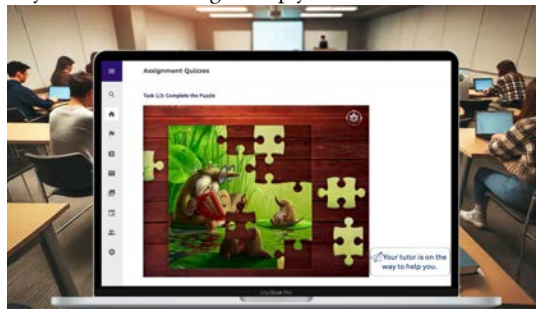
Scenario 1 *Sensing not in use + system hint:*

“Imagine that your university has introduced a new learning management system. You are in the tutorial class, practicing puzzle tasks. When the system thinks you are struggling, it sends you an automatic hint to help you.”



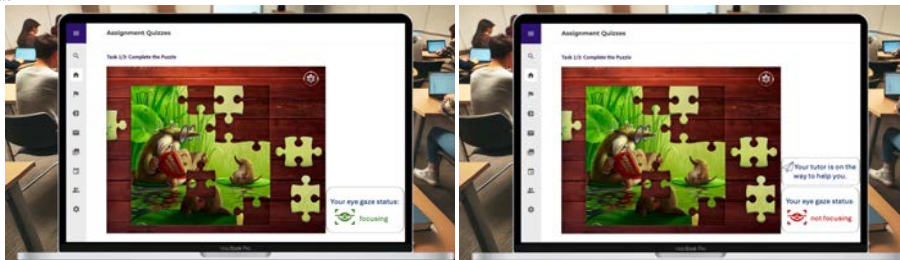
Scenario 2 *Sensing not in use + teacher assistance:*

“Imagine that your university has introduced a new learning management system. You are in the tutorial class, practicing the puzzle tasks. When the system thinks you are struggling, it notifies your tutor. A notification then pops up, letting you know that your tutor is coming to help you.”

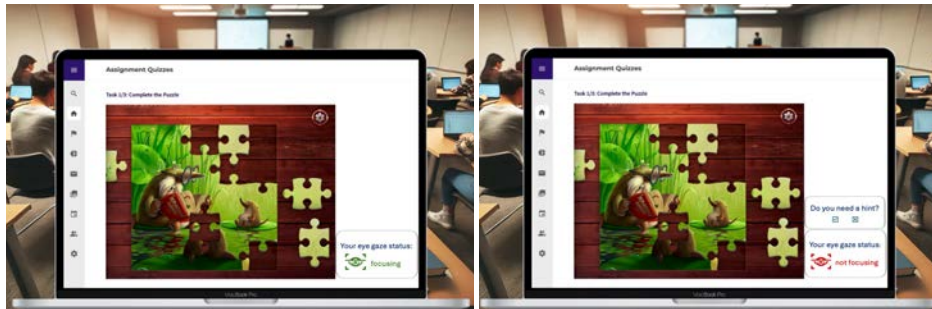


After completing the first two scenarios, we present the text: “In another system, the system observes your status to decide if you need help.”

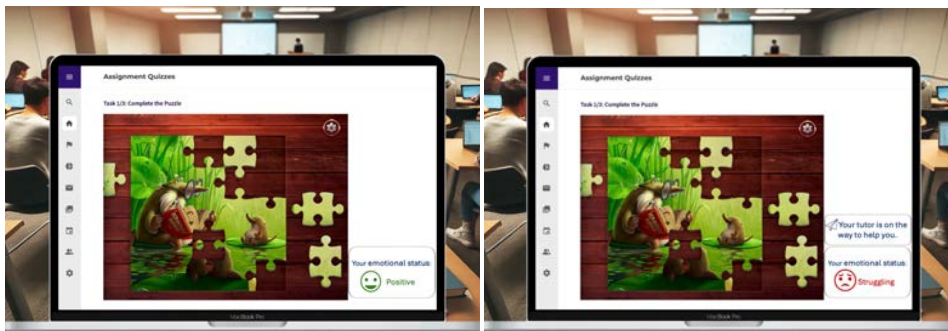
Scenario 3 *gaze-based attention + teacher assistance:* “You are in the tutorial class, practicing puzzle tasks. Your eye gaze is being tracked. When the system thinks you are struggling, it notifies your tutor. Your tutor will then come to help you.”



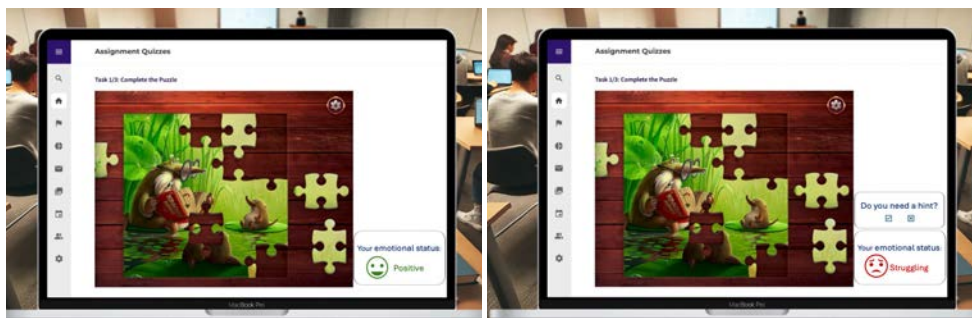
Scenario 4 *gaze-based attention detection + system hint*: “You are in the tutorial class, practicing puzzle tasks. Your eye gaze is being tracked. When the system thinks you are struggling, it sends you an automatic hint to help you.”



Scenario 5 *facial emotion-detection + teacher assistance*: “You are in the tutorial class, practicing puzzle tasks. The camera on your laptop has read the emotions on your face. When the system thinks you are struggling, it notifies your tutor. Your tutor will then come to help you.”



Scenario 6 *facial emotion-detection + system hint*: “You are in the tutorial class, practicing puzzle tasks. The camera on your laptop has read the emotions on your face. When the system thinks you are struggling, it sends you an automatic hint to help you.”



APPENDIX B. SYNTHESISED VIEW OF INTERVENTION RATINGS (TABLES 7 AND 8)

Table. Comparison of Ratings on the Intervention Form in Sensing Not-In-Use Scenario and Sensing In-Use Scenario

Comparing results in Table 7 & Table 8	Human Assist (No Sensing Scenario)		Human Assist (Sensing In-Use Scenario)		System Hint (No Sensing Scenario)		System Hint (Sensing In-Use Scenario)	
	Table 7		Table 8		Table 7		Table 8	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<i>Belief about Learning</i>	3.09	0.92	2.29	0.86	3.47	0.82	2.73	0.97
personalisation	3.41	1.24	2.36	1.24	3.77	0.97	2.90	1.30
efficiency	3.29	1.25	2.23	1.20	3.91	0.95	2.88	1.31
control of pace	2.48	1.29	1.86	1.04	3.59	1.19	2.58	1.29
grade performance	3.44	1.19	2.49	1.18	3.56	0.96	2.84	1.15
distraction	3.17	1.19	4.05	1.08	2.77	1.13	3.74	1.15
over-helping	2.94	1.21	3.17	1.16	3.23	1.19	3.08	1.17
<i>Affective Responses</i>	2.26	0.95	1.70	0.88	3.47	0.94	2.21	0.95
enjoyable	2.62	2.00	1.73	1.02	3.52	1.06	2.27	1.17
confident	2.45	1.00	1.89	1.00	3.10	1.09	2.30	1.12
anxious	3.84	1.26	4.50	1.00	2.22	1.03	3.93	1.189
uncomfortable	3.96	1.18	4.55	0.93	2.56	1.23	4.11	1.138
worried	4.05	1.09	4.28	1.08	2.84	1.26	3.72	1.206
fearing mistakes	3.68	1.22	4.10	1.11	2.33	1.08	3.54	1.169

The response score is based on the Five-point Likert scale (1=Strongly disagree, 5=Strongly agree). Construct scores (shaded rows) were calculated by reverse-coding negatively worded items (e.g., 'distraction') so that higher values indicate more positive perceptions. Item-level scores are presented in their original direction.

6.4 ADDITIONAL REFLECTIONS

This study offered several moments of insight, surprise, and learning during its development and analysis, which were not included in the published manuscript but are worth reflecting on here.

One surprising finding was the consistently negative responses toward both emotion detection and gaze tracking features. I had initially assumed that students might view gaze tracking more favourably, as it appears less intrusive than emotion recognition, which involves reading one's facial expressions. However, the data revealed no significant difference between the two – both forms of monitoring were perceived negatively. This suggests a more fundamental discomfort with being monitored while studying, regardless of the sensing method. This insight highlights the need for AIED designers to explore less intrusive alternatives for understanding students' learning progress - an idea I further discuss in the paper in Section 6.3.

Another unexpected result concerned students' preferences for system-generated hints over tutor assistance. Based on Study 1, I anticipated that students would prefer more teacher involvement, given some participants' concerns about reduced student-teacher interaction (seen in Chapter 4). However, the statistical results showed a stronger preference for system hints. I was pleased to see that the open-ended responses provided a compelling explanation: students worried that tutor intervention might inadvertently expose their learning struggles in front of peers, potentially causing embarrassment or even social stigma. This finding underscores the importance of considering the subtle social dynamics of classroom environments - something often overlooked in conventional AIED system design.

This explanation for student preference for hints over tutor assistance also prompted a methodological reflection, highlighting the value of integrating qualitative methods into this study. Originally, the experiment was conceived as a purely quantitative survey. The study was reshaped by my concern that statistical results alone would offer only a partial understanding - revealing what students think or feel, but not why they feel that way. This inclusion of open-ended questions enabled participants to articulate the reasons behind their responses, offering deeper insights into their concerns and ethical priorities. Study 3's findings ultimately validated this design choice.

Chapter 7: Discussion

This thesis aims to answer the overarching research question: “*What are the ethical challenges in adopting artificial intelligence in education (AIEd), from the perspectives of different stakeholder groups?*” The three studies of the thesis addressed this question by exploring the views of students and other key stakeholders, including school governors and teaching representatives, AIEd industrial practitioners, and parents. The five papers presented in this thesis offer critical insights into the ethical and sociotechnical dimensions of AIEd design and adoption.

While this chapter is primarily structured around the three overarching Aims of the thesis, these Aims are directly informed by the four core Problems identified in Chapter 1 and are operationalised through the three sub-Research Questions. The Aims reflect a broader contribution this thesis seeks to make, while the Research Questions serve as more specific inquiries, each addressed through an empirical study. In this way, the Discussion chapter uses the Aims as a higher-level structure to synthesise and integrate findings across the studies, while also responding to the Research Questions and addressing the original Problems. This structure enables a cohesive discussion that links the thesis’s empirical contributions to its broader theoretical and practical implications. (See Fig. 1 in Chapter 1 for a visual overview of the thesis structure and logic).

This chapter synthesises findings from three research studies of this thesis and offers a holistic discussion. Section 7.1 addresses Aim 1 of the thesis, discussing ethical challenges by synthesising insights across all stakeholder groups. Section 7.2 addresses Aim 2, presenting implications for understanding AIEd ethics through a human-centred ethical approach. Section 7.3 addresses Aim 3, offering recommendations for future ethical AIEd design and governance. Section 7.4 then reflects on methodological considerations.

7.1 SYNTHESISING PERSPECTIVES: IDENTIFYING UNIQUE ETHICAL CHALLENGES

This section addresses *Aim 1: Examine how different stakeholder groups perceive and respond to AIEd adoption and its impacts.*

Interconnected Ethical Impacts

The ethical impacts identified in this thesis are not isolated; they are deeply interconnected, as demonstrated in the three papers from Study 1 (Chapter 4). The consequences of technology adoption can never be fully anticipated, and Chapter 4 has provided vivid examples of how students may perceive and respond to AIED adoption across various dimensions. For instance, when evaluation criteria lack transparency, students may perceive monitoring-oriented AIED systems as creating a panoptic environment. This perception can lead to psychological impacts such as anxiety, and may prompt students to alter their behaviour, either to avoid mistakes or to conform to the system's expectations (Section 4.5). These behavioural changes, in turn, undermine their autonomy and motivation in the pedagogical dimension, since they no longer learn for personal growth but to satisfy the perceived system's demands (Section 4.6). Furthermore, the data from these altered behaviours feeds back into the AI models, potentially undermining their effectiveness and creating a negative feedback loop (Section 4.7). These examples show that ethical concerns are interwoven, where one concern often triggers or intensifies another. We must reconsider AIED as a holistic ecosystem rather than simply a tool with fixed intended outcomes.

The findings also revealed ethical issues rooted in social needs arising from the situated context of use, rather than from the merely technical aspects of AI. These social aspects have often been overlooked by existing ethical frameworks, as discussed in Chapter 2. For instance, Chapter 6 highlighted students' privacy concerns about human tutor assistance, particularly the risk of unintentionally revealing their learning struggles to peers, leading to social embarrassment. Similarly, in Chapter 4, autonomy was disrupted not by AI decision-making, but by the psychological pressure students felt to meet AIED expectations. Concerns around equity and fairness arose from introverted students regarding how AIED systems might favour extroverted students, who are more expressive. Introverted students may feel forced to adapt to evaluation methods that quantify performance and don't align with their natural behaviours. AI's ability to mitigate teacher bias in evaluations was also noted, but it might unintentionally introduce its own biases, influencing teachers' perceptions before they fully assess students. This highlights the need for AIED systems to better account for nuanced social dynamics within the educational system.

To reflect on these interconnected impacts, it is essential to recognise that the pedagogical environment (discussed in Section 4.6) is shaped by the individuals

(discussed in Section 4.5) within it and their interactions. The way individuals react to AIED directly impacts the dynamics of their environment (discussed in Section 4.7), as their behaviour, influenced by AI, could also affect how other individuals in the broader social context engage with the technology, such as their peers and educators. These interactions are mediated by the AIED technology, which in turn shapes the overall educational experience, encompassing both teaching and learning. Therefore, understanding AIED's impact requires looking beyond its technical features to consider the social and psychological dimensions that influence and are influenced by it. This thesis provides a bottom-up perspective, grounded in students' lived experiences, to underscore the need for a more holistic and deliberate approach to AIED design and adoption.

Tensions Among Stakeholder Groups

All stakeholder groups in this study acknowledged students' potential resistance to monitoring-oriented AIED adoption. Beyond students themselves, school decision-makers and teacher representatives, parents, and developers all recognised that this type of AIED could trigger negative emotional responses, such as discomfort, anxiety, frustration, and privacy concerns. This seems promising for the ethical implementation of AIED. However, while there was a broad recognition of these potential impacts, stakeholder groups diverged significantly in how they perceived and prioritised these concerns.

For parents and school decision-makers - who are the primary guardians of students' well-being - student resistance is often seen as a transitional phase rather than a valid critique of the system itself. Their focus remains on the potential educational benefits of AIED, often framing students' concerns as obstacles to be overcome or managed rather than ethical issues requiring deeper reflection. Chapter 5 has demonstrated that decisions about designing and deploying AIED are primarily driven by goals such as teaching efficiency, scalability, and legal compliance. Meanwhile, developers and decision-makers focus on data management and system performance, such as accuracy and system effectiveness, rather than students' lived experiences. Consequently, potential impacts - ranging from behavioural changes to disrupted learning autonomy and compromised learning experience - are largely overlooked. These stakeholders aim to persuade students to embrace AIED by emphasising its learning benefits and expecting students to eventually adapt.

However, as Chapter 6 highlights, this assumption does not align with students' perspectives. Students do not believe these systems will meaningfully enhance their learning and are unwilling to trade their privacy and autonomy for promised benefits. Unlike decision-makers and developers who prioritise system optimisation, students place greater value on autonomy and privacy than on technical performance metrics like accuracy and fairness. This misalignment in priorities underscores a fundamental gap in how different stakeholders interpret the ethical and practical implications of AIED adoption.

The divergence in perceptions and priorities is further complicated by the inherent power dynamics observed in AIED design and adoption. The findings suggest that the current AIED adoption structure may offer students limited influence over how these technologies are implemented, with institutional needs and decision-makers' priorities often overshadowing student values. While many students in this study expressed a strong desire for control over their engagement with AIED technologies - particularly the ability to opt in or out rather than being passive recipients - this level of autonomy appears to be rarely reflected in decision-making processes (as illustrated in Chapter 5). These patterns, while context-specific, raise important questions about the inclusiveness and responsiveness of current AIED governance practices.

This challenge is reinforced by the often overly optimistic or ethically superficial views held by more powerful stakeholders, including developers, school decision-makers, and parents, who may neglect broader socio-ethical concerns tied to students' learning experiences and well-being. As a result, students – who have the least influence and agency in these processes – are expected to bear the greatest impact of AIED adoption. This raises a critical ethical question: If student needs are consistently overshadowed by institutional priorities, can AIED truly serve students' best interests? If their needs are disregarded, what is the point of designing technology that claims to enhance their educational experiences?

This thesis exposes a fundamental issue in AI adoption within multi-stakeholder power structures: the authenticity of acceptance, especially for disempowered groups. In the AIED context, students' perceptions and affective attitudes - key attributes of user acceptance (Sekhon et al., 2017) - are often dismissed as a natural part of a technological transition. All stakeholder groups endorsed the established authority power of schools and parents over students, limiting students' agency in decision-making. This raises a

critical question: Can consent be considered genuine under such power imbalances? Similar concerns extend beyond education to other AI adoption contexts, such as healthcare (doctors vs. patients), finance (evaluators vs. loan applicants), and hiring (HR vs. job candidates), where vulnerable users are expected to comply with AI systems designed without their meaningful input. If these users' needs and concerns remain sidelined, the AI system risks not only failing to enhance societal wellbeing but also perpetuating structural inequalities and ethical harms, reinforcing a broader pattern of disempowerment in AI adoption.

7.2 IMPLICATIONS FOR AIED ETHICS: HUMAN-CENTRED APPROACH

This section addresses *Aim 2: Deepen the understanding of the ethical and sociotechnical consequences of AIED adoption and expand the ethical discourse of AIED through a human-centred approach.*

The ethical discourse surrounding AIED has traditionally been dominated by concerns about data-centric and computational issues (Holmes et al., 2021), as discussed in Chapter 2. While critical, these approaches largely overlook the direct, lived experiences of students – how AIED systems shape their learning, relationships, and well-being. Chapters 4 and 5 showed that students' perceptions of ethical impacts are under-represented, even though these perceptions strongly influence acceptance and learning outcomes. Focusing only on technical or institutional concerns leaves a critical blind spot in understanding and addressing the human dimensions of AIED.

Holmes et al. (2021) describe this blind spot as the “unknown unknowns”, represented by a question mark at the centre of their AIED ethics diagram (Fig.3 in Section 2.3.1; also shown on the left side of the adapted version in Fig.5 below), where data, computation, and education intersect. This thesis argues that this missing space should be understood as the **Ethics of Human Factors** – the ethical dimensions that emerge from how people actually encounter and live with AIED technologies.

To fill this gap, this thesis proposes the **Ethics of Human Factors Framework** (Fig.5), reframing the “unknown unknown” into three interrelated domains:

- **Student-AI interactions:** how students engage with, are affected by, and make sense of AIED systems, highlighting student-centred values (such as autonomy, privacy, and fairness), and emotional well-being, which this thesis identifies as a critical but underexplored ethical dimension. While often viewed separately

from values like privacy or fairness, well-being emerges in this framework as both a consequence of ethical (or unethical) design and a value in its own right, essential to ensuring that students can thrive in AI-mediated learning environments.

- **Learning and teaching context:** how AIED technologies align with pedagogy and reshape dynamics relationships among students, teachers, and educational systems.
- **Power and governance:** how power imbalances among key stakeholders can shape (un)ethical outcomes.

Together, these domains extend existing frameworiking by addressing what has been overlooked – the human and relational aspects of AIED – and by encouraging more inclusive, reflexive, and context-sensitive AIED development.

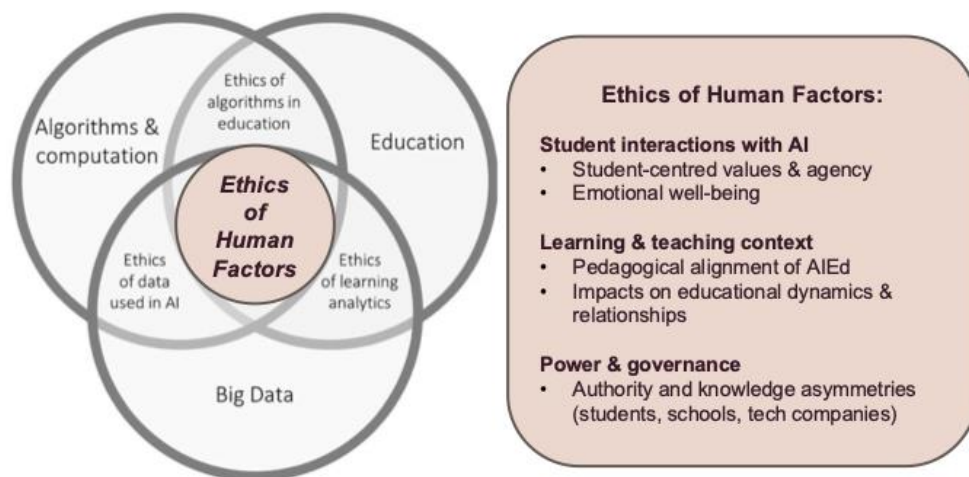


Fig. 5. Human-Centred AIED Ethics, adapted from Holmes et al. (2021)’s AIED Ethics Framework

User well-being must be central to ethics of human factors in AIED. The concept of promoting well-being has long been advocated in existing ethical AI principles like *beneficence*, supported with various demonstrations or proposed subtopics such as benefitting humanity and dignity (Floridi et al., 2018), protecting fundamental rights (Prem, 2023), and doing good for human society (Beijing AI Principles, 2019). These frameworks also emphasise respecting human autonomy (European Commission, 2019), and including affected stakeholders in decision-making (Morley et al., 2020). However, these principles under beneficence or the idea of doing good often lack specificity, making them difficult to implement in practice, as these general terms do not address the specific needs and concerns of individual users in specific contexts. For instance, what

do well-being, humanity, dignity, or justice mean in education? What fundamental rights should be protected? And what constitutes “doing good” in a classroom shaped by AI? This thesis offers an empirically grounded response to these questions, with its findings providing concrete insights into how students experience AIED systems and what they perceive as supportive or harmful to their well-being and sense of justice.

As introduced in Chapter 2, the earliest AIED research has been central to *beneficence* through a learner-centred approach (Kay, 2012), ensuring AI support individual students with a strong emphasis on “care” for the learner (e.g., (Self, 1998)). However, as the potential contribution of AIED has been recognised, and the scope of AIED has expanded to include broader educational ecosystems - including educators, families and external influencers such as institutional goals, the focus shifted from individual students’ needs to more systemic concerns. These systemic concerns are centred on algorithm models and computational approaches that aim to provide technological support to learning, which prioritise technical efficiency over human experience. This thesis calls for a return to a learner-centred approach, ensuring that beneficence in AIED truly prioritises students’ well-being and educational needs.

This thesis reinforces ongoing warnings from scholars within the AIED research community for AIED innovation and ethics being overly driven by technical goals (e.g., system efficiency) and addressing only technical concerns (Holmes et al., 2021). Beyond these critiques, this work argues that AIED ethics must account for the inherently social nature of learning. As scholars have emphasised, learning is a deeply social experience. Therefore, design considerations of AIED innovation cannot be entirely reduced to data-driven solutions (Fox, 2022; Fynn, 2016; Holmes et al., 2021; Siemens, 2013). Rather, it needs a deeper understanding of individuals’ learning experiences, practices and social interactions (Bond et al., 2019; Veletsianos et al., 2015). The findings from this thesis have provided empirical evidence that many ethical concerns in AIED arise not just from technical flaws but also from disruptions to the social fabric of education. Ethical considerations in AIED must therefore extend beyond individual learning outcomes to include broader social dimensions that define students’ well-being and educational experiences.

Incorporating emotional well-being into AIED ethics is vital. While students’ educational needs can often be addressed in terms of cognitive or performance outcomes, their emotional well-being (e.g., how they feel while engaging with the technology) is

equally critical. However, this emotional well-being has been frequently overlooked in existing AIED frameworks and technology evaluation metrics. As shown in Chapters 4 and 6, emotional responses, such as stress, anxiety, and feelings of surveillance, are not merely incidental effects of technology use but are fundamental to how students experience AIED systems. These affective attitudes, as primary attributes of acceptance towards the technology (Sekhon et al., 2017), also influence students' willingness to engage with the technology. As this thesis has argued, ethical AIED adoption requires genuine student acceptance, which cannot be achieved without addressing these affective and psychological dimensions. As such, AIED ethics must expand to more explicitly include what might be termed as emotional ethics, that is, ethical considerations around how technologies affect users' emotional and psychological states. This approach is to ensure that AIED supports not only learning outcomes but also students' broader well-being.

Finally, this thesis challenges the narrow conception of ethics often seen in AIED literature (as confirmed by (Hakimi et al., 2021; Holmes et al., 2021; Murchan & Siddiq, 2021; Tzimas & Demetriadis, 2021)). It raises a critical question: if ethics should go beyond data-centric or computation-centric technical concerns, what does it truly mean to approach the ethics of AI and AIED comprehensively? Ethics in this context is not just about the technology itself but also about its broader consequences of adoption and use. Then the question comes to that: Should emotional well-being and broader sociotechnical impacts be treated as ethical imperatives rather than peripheral concerns?

Ethical imperatives such as through legal frameworks often emerge from social consensus, gaining urgency only when harm becomes undeniable, and regulation becomes inevitable (e.g., (Buolamwini & Gebru, 2018; Noble, 2018)). Unlike data privacy, bias, and algorithmic fairness where explicit criteria, quantitative benchmarks, and technical solutions can be applied (e.g., (Ashwin & Biswas, 2024)), the human-centred ethical principles that this thesis advocates, such as beneficence and well-being, present distinct challenges. The harms they seek to prevent are difficult to identify, measure, and remediate once they occur. This makes them harder to regulate, as their urgency is often unrecognised until tangible harm is widespread. Moreover, well-being is socially and culturally constructed (Nussbaum, 2000; Wilkinson & Pickett, 2009), meaning perceptions of harm and ethical obligations vary across contexts, further complicating consensus and policy development. However, waiting for consensus or

regulatory frameworks to catch up is not an option. By the time ethical concerns surrounding AIEd become urgent enough to demand action, harm may already be embedded into educational systems. While there is growing advocacy for “designing technology for good”, this ambition does not always translate into concrete, sustained practice - particularly without careful consideration and critical examination. Thus, this thesis further argues that ethics in AIEd must move beyond reactive regulation and instead adopt a proactive, human-centred approach, where well-being and sociotechnical consequences are considered fundamental from the outset of designing AIEd, not as afterthoughts.

A context-specific approach to human-centred AIEd ethics is critical. While many existing ethical frameworks are grounded in general principles, this thesis argues that these must be supplemented with empirical insights to meaningfully account for the social and cultural variability of ethical concerns – particularly those related to well-being, autonomy, and privacy. As demonstrated in Chapter 5, stakeholders’ attitudes toward privacy are shaped by their cultural contexts; such perceptions and reactions to privacy concerns may differ significantly in other cultures. Fox (2022) has similarly highlighted the need for ethical flexibility, arguing that researchers must balance adherence to broader ethical values while respecting local norms and diverse stakeholder perspectives. In AIEd, this means that ethical considerations in AIEd design must be adaptive, not prescriptive, and responsive to the unique experiences of students and educators across different educational systems.

To further illustrate this, Fox (2022)’s reflections on the CERD framework for ethical thinking (consequential, ecological, relational, and deontological) provide a useful lens. Instead of treating AIEd ethics as a static set of principles, this framework calls for a dynamic, anticipatory, and context-sensitive approach to ethical design. It moves beyond merely addressing the technical challenges and invites attention to the broader relational and situated implications of AI systems. For example, future ethical AIEd design should focus on the need to balance maximising benefits with minimising harm (consequential thinking). This requires not just a harm-benefit evaluation but flexible design practices that consider alternative perspectives in the evolving user contexts. It should also identify those being affected (ecological thinking) and attend to stakeholder relationships and power dynamics (relational thinking). AIEd researchers and practitioners must embrace deontological thinking by identifying their obligations and

navigating ethical decisions holistically in AIEd design and development. This thesis's approach has been aligned with this framework, investigating the consequences of AIEd adoption from students' perspectives - who will be affected directly (Chapters 4 & 6), and examining how ethical decisions are shaped by stakeholder dynamics (Chapter 5). This approach also aligns with a key argument made throughout this thesis: the ethical considerations in AIEd must not be treated as an afterthought but require proactive engagement, where ethical risks are anticipated, acknowledged and mitigated before harm occurs.

7.3 IMPLICATIONS FOR THE FUTURE OF AIED

This section addresses *Aim 3: Offering actionable implications and recommendations for ethical AIEd design and governance that are grounded in stakeholder concerns and ethical considerations*

7.3.1 Aspirations for Future AIEd: Potential Benefits if Designed Well

This thesis has demonstrated that, while there are significant concerns about the adoption of AIEd, particularly from students, there are also considerable positive expectations about its potential. These aspirations are rooted in the belief that AIEd, if designed with ethical and human-centred principles, can enhance learning experiences and create new opportunities for engagement. The following aspirations highlight what AIEd could achieve if developed with a thoughtful, balanced approach.

AI-Enabled Personalised Learning

Findings from Chapters 4 and 6 indicate that students are optimistic about AIEd's capacity to deliver personalised learning experiences tailored to individual strengths, weaknesses, and needs. Unlike traditional classroom settings, where teachers have limited capacity to provide individualised attention to each student, AIEd could offer equal and undivided attention and timely and targeted support to all students without overwhelming teachers (Holmes et al., 2019). Additionally, students value AIEd-generated insights based on their learning progress and foresee that it could promote self-reflection for their self-directed learning, making their learning journey more accessible.

Balancing Personalised Learning and Student Autonomy

However, a key concern regarding personalised learning interventions is their impact on students' autonomy. While students value targeted support, they stress the need

for customizable settings within AIED systems to maintain control over their learning experience. Chapter 6 demonstrates that students prefer system-generated hints over direct tutor interventions, as the former allows them to decide when and how they engage with support. This suggests that AIED should carefully design intervention mechanisms - offering suggestions rather than imposing them. Students should have the agency to determine when, how, and whether they want assistance, including the option to decline unnecessary interventions.

Chapter 4 highlights students' diverse needs: some prefer more independent exploration, while others may benefit from guided interventions, depending on the learning contexts. While timely support is helpful for some, others expressed concerns that frequent automatic interventions could undermine their ability to engage in independent thinking and problem-solving. It might lead to over-reliance on assistance or perceiving interventions as intrusive or even manipulative. To address this, AIED systems must balance fostering independence with providing support. Again, granting students control over their engagement with the system ensures flexibility in meeting varied needs and preferences. This also empowers students to shape their own learning experiences and educational journeys, reinforcing their autonomy in education.

AIED as a Collaborative Learning Companion, not a Supervising Tool

To truly promote student learning autonomy and empower students, AIED must function as a collaborative learning companion rather than as a supervising tool. A central tension in current AIED design is whether students are passive recipients of AI-driven interventions or active agents in shaping their learning experiences. Many existing AIED tools, especially those focused on monitoring, are teacher-centred, aiming at enhancing educators' ability to understand students' progress and adapt instruction (as discussed in Chapter 2.1). While these systems may improve teaching efficiency, they risk undermining student agency, learning autonomy and well-being. Chapters 4 and 6 revealed that students view systems that track their emotional states and gaze patterns as controlling and judgmental rather than supportive. This creates an uncomfortable learning environment, leading to passive compliance where students feel pressured to conform rather than engage actively in learning. They feel these AIED innovations may not align with their needs and preferences.

As many scholars have argued (Brossi et al., 2022; Holmes et al., 2019), AIED should facilitate collaborative decision-making among students, teachers, and the system,

not just assist teachers in tracking student progress. While AIED's analytical ability to optimise education is valued by students and educational stakeholders like teachers and school governors, pursuing data-driven insights should not come at the expense of students' well-being and learning autonomy. This calls for a shift in how analytics are used to gauge student learning. We should find alternatives to gather insights that inform learning instead of relying on intrusive monitoring.

Enhancing Accessibility and Equity

AIED holds the potential to bridge educational disparities by providing adaptive learning tools that support students, particularly those with disabilities or from marginalised backgrounds. By tailoring educational experiences to their special needs, AIED could help ensure equitable access to personalised learning opportunities. However, achieving true equity requires more than just personalisation (Tsai, Perrotta, et al., 2020); it necessitates careful attention to potential biases in how the AIED mechanism interprets and evaluates student learning (I will detail in Section 7.3.2 - the risks of one-size-fits-all solutions that could unintentionally reinforce or even exacerbate existing inequalities). For instance, as highlighted in Chapter 4 and Chapter 6, neurodivergent students raised concerns that AIED's interpretation of their learning behaviours might misrepresent their abilities, leading to unfair evaluations. To prevent such biases, future AIED systems must be designed with inclusivity in mind, ensuring they actively support marginalised groups. This means taking into account cultural, social, and contextual diversity to create AI systems that are truly equitable and responsive to the needs of all student groups, regardless of their background or abilities.

Fostering Genuine Engagement and Meaningful Educational Experiences

AIED should not be solely driven by efficiency metrics such as task completion times and performance scores. While optimising learning processes is valuable, education extends beyond measurable outputs – it is about fostering deep engagement, critical thinking, and meaningful interactions between students, teachers, and educational systems. AIED must enrich the learning experience rather than reduce it to algorithmically optimised performance benchmarks. To achieve this, it should empower students and cultivate genuine engagement. AIED should facilitate collaborative learning environments that encourage active learning. By emphasising genuine engagement and the quality of interactions over rigid efficiency, AIED can contribute to a richer, more meaningful educational experience.

7.3.2 Challenges Future AIED Must Conquer

The implementation of AIED presents several challenges beyond the well-discussed data-centric and algorithmic concerns. This thesis highlights structural challenges spanning the design, adoption, and governance of AIED systems. This section explores these challenges from a broader perspective, including students' values, social dynamics, and the factors within the wider educational ecosystem. Alongside these discussions, I provide actionable recommendations to address these pressing issues.

Structural Limitations in AIED's One-Size-Fits-All Approach

A key challenge in achieving inclusivity in AIED lies in its inherent one-size-fits-all approach to tracking and inferring students' progress. While AIED promises to personalise learning, current models often rely on rigid, standardised patterns of interaction that may unintentionally reinforce inequalities rather than reduce them. As a result, students with diverse learning styles and needs (Baker & Hawn, 2022) - such as introverted or neurodivergent learners - may find themselves disadvantaged by systems that misinterpret their behaviours or favour more traditionally expressive forms of participation. For instance, some introverted students view AIED as offering a safer, more comfortable learning environment where they can receive help from non-human agents, without the stress of direct interactions with teachers or peers. However, they also fear that systems relying on behaviour tracking may inadvertently favour more outgoing students, whose engagement is more visible and thus more easily recognised by AI models. Similarly, neurodivergent students worry that their unique learning patterns may be flagged as anomalies, leading to incorrect assessments or biased interventions.

These structural limitations call for a more deliberate approach to AIED design – one that acknowledges and accommodates the diversity and complexity of real-world educational contexts, instead of enforcing predefined behavioural norms. Education involves students from various backgrounds, including differences in culture, race, gender, socioeconomic status, and cognitive differences. AIED should develop models that are capable of recognising and responding to these differences without reducing students to rigid behavioural metrics. To achieve this, future AIED systems must integrate inclusive design principles from the outset, ensuring that they do not unintentionally reinforce exclusionary biases under the guise of personalisation.

Lack of Rigorous Evaluation in AIED Innovation

Many AIED solutions are deployed without sufficient empirical validation, particularly regarding their sociotechnical implications. Bartoletti (2022) warns against scaling up pedagogical ideas without rigorous evidence, as this can result in unintended consequences for not only students but also the broader educational ecosystem. Chapter 4 has highlighted how intrusive AIED adoption can disrupt individual learning processes and pedagogical interactions. Despite the promise of AI-driven learning, there remains a lack of comprehensive research evaluating how these systems impact students' well-being, autonomy, and educational experiences - beyond basic metrics focusing on AIED performance.

As I argued in Section 7.3.1, AIED design should go beyond efficiency to foster deep engagement, critical thinking and meaningful interactions. The same principle must apply when validating new AIED innovation ideas. Evaluation must move beyond simplistic measures, such as students' task completion time and their scores. Rather, a holistic approach is required, assessing AIED's influence on students' learning autonomy, well-being, pedagogical interactions, and the broader educational environment, including shifts in student-teacher dynamics and changes in educational roles (these themes can refer to students' concerns in Chapter 4). This thorough evaluation ensures that unintended consequences are identified early, especially those harmful ones.

Rather than accelerating deployment with limited scrutiny, AIED requires a deliberate, ethically informed, and evidence-based approach where all potential impacts are thoroughly assessed before large-scale adoption. A proactive, anticipatory evaluation framework must be embedded into AIED design from the outset, ensuring that human-centred ethical considerations are prioritised. This means taking the necessary time to understand the full spectrum of AIED's effects. This empirical knowledge not only informs ethical AIED design but also prepares for potential adoption risks.

Lack of Ethical Reflexivity in AIED Development and Research

I have discussed the insufficient empirical evaluation of AIED innovation, but there is also a deeper concern about the absence of ethical reflexivity in AIED research and development. The prevailing mindset among some AIED researchers – driven by enthusiasm for technological advancements – prioritises innovation over critical reflection on its broader consequences. This optimism, while well-intentioned, risks accelerating the deployment of AIED without enough deliberation and adequate safeguards, overlooking potential ethical and societal impacts.

To prevent ethical missteps, AIED development must integrate interdisciplinary ethical oversight at every stage of the research and deployment process. AI developers, educational technologists, and learning designers should work collaboratively with ethical review committees to identify and mitigate ethical risks before AIED systems reach widespread adoption (Fox, 2022). However, ethical considerations must go beyond procedural checklists - they should be embedded in the research culture itself, fostering an ongoing commitment to anticipating, questioning, and addressing ethical risks in practice.

AIED researchers must also take responsibility for ensuring their innovations serve the broader well-being of students and society, rather than focusing solely on technical advancements. This requires shifting from a narrow, technology-driven perspective to a relationship-centred ethical approach (Fox, 2022) (as I introduced in Section 7.2), paying attention to the consequences and impacts, and recognising that AIED technologies do not exist in isolation but are deeply embedded in social, educational, and institutional contexts. Ethical reflexivity should be treated as an essential research tool - one that ensures AIED serves not just as an efficient educational technology, but as a responsible and equitable innovation that genuinely benefits learners.

Lack of Student-Centred Design Thinking

Current AIED design has largely been shaped by a teacher-centric mindset that prioritises monitoring student progress to support educators rather than fostering meaningful student engagement. This focus often leads to the pursuit of accuracy and efficiency in computational models while overlooking students' lived learning experiences and values. As Brossi et al. (2022) caution, such "adult-centric" (teacher-centric) approaches of AIED development might exclude student perspectives and reinforce passive learning dynamics. This is especially concerning given that modern education has increasingly embraced student-centred pedagogies, yet AIED design in this study appears to reverse this progress, re-centring control around educators and systems rather than learners. This unbalanced focus might risk creating AIED that treats students as mere subjects of technological advancements and a passive recipient of AIED's promised benefits. This approach can significantly impact students' motivation and engagement by reducing their role in the learning process, leading to a diminished sense of ownership over their education. It also skews the power structure, favouring educators and technology developers, and leaving students more vulnerable to ethical risks.

To address this challenge, AIED design must shift towards a student-centred approach that incorporates participatory methods. This means moving beyond positioning students as mere data subjects and instead empowering them as active co-creators of their learning journeys. Involving students in the design process can help ensure the AIED responds to their genuine needs and better anticipates potential ethical risks early. A participatory approach can also mitigate power imbalances by allowing students to influence decisions about AIED that directly affect their learning. This ensures that ethical considerations are not just imposed from above but are informed by the diverse experiences and perspectives of students themselves.

Lack of Genuine Consent and Acceptance in AIED Adoption

A significant challenge of AIED adoption is ensuring genuine informed consent and acceptance, particularly when AIED relies on sensitive student data (Li et al., 2021). Two major concerns arise. First, power imbalances in the educational system raise the question of whether student consent can ever be considered genuine. In many cases, AIED adoption is mandatory (e.g., (Jatileni et al., 2024)) – making authentic consent impossible (Berendt et al., 2020; Tschider, 2018). Even when not mandatory, institutional pressures and unequal power dynamics may still compromise students’ ability to make free choices (e.g., data collection in Virtual Learning Environments employed by some institutions is often set by default (Sclater, 2016)). If students do not accept these systems, do they have meaningful alternatives? This issue becomes even more complicated in K-12 education, where consent is typically granted by children’s guardians, such as parents. What, then, constitutes meaningful acceptance, and how can it be ensured that consent is truly voluntary rather than coerced or procedural?

Second, the overwhelming complexity of the terms and conditions often prevents students and their families from fully understanding the implications of providing consent, e.g., for data usage (Hartung, 2020). Many unknowingly consent to AI implementation in general due to information overload, making the current system of informed consent questionable, if not unethical. Given these complexities, it raises the question: who is ultimately responsible for safeguarding students’ rights and data?

To address these concerns, clearer guidelines must be established to ensure that students and their guardians truly understand AIED and its implications, including its outcomes and risks. New mechanisms are needed to guarantee valid consent, giving students and families meaningful opportunities to express concerns and participate in

decisions about whether AI systems should be implemented in their learning environments.

Lack of Valid Ethical Accountability in AIED Governance

AIED governance faces a critical gap in ethical accountability, as oversight mechanisms remain largely reactive rather than proactive. AI technology evolves at a pace that outstrips the development of ethical frameworks and regulatory structures, where governance mechanisms often emerge only after harm has been widely recognised. This lag in addressing potential ethical issues can lead to unintended harm by the time solutions are considered.

Without proactive governance, key decision-makers - including AI developers, school governors, and parents - often fail to ensure ethical safeguards before deployment. As demonstrated in Chapter 5, developers frequently prioritise project delivery efficiency, accelerating adoption without fully considering risks. Meanwhile, school decision-makers and parents, despite their role in overseeing AIED adoption, often lack sufficient AI literacy to critically evaluate its pedagogical and ethical implications. As a result, flawed decisions made without valid accountability leave significant ethical risks unaddressed, reinforcing a governance gap that favours rapid adoption over careful deliberation.

Addressing this challenge requires a proactive and informed approach, embedding ethical accountability at all stages of AIED development and implementation. Stronger oversight mechanisms must hold AI developers, schools, and parents accountable for ethical risks. However, accountability must also be valid - meaning those responsible for decision-making must be equipped with the necessary knowledge and tools to engage in ethical deliberation. This calls for targeted efforts to improve AI literacy among key stakeholders, particularly educators and school administrators, to enable them to critically assess AIED's risks and implications. This approach can also shift the power imbalance stemming from the knowledge gap and avoid the ethical debt from short-sighted decision-making, as demonstrated in Chapter 5. Additionally, clear and enforceable guidelines must be developed to ensure that ethical risks are identified and mitigated proactively rather than retrospectively. By shifting from a reactive to a proactive governance model, AIED can be implemented with greater ethical responsibility, reducing the likelihood of harmful consequences.

Bridging the Gap Between Ethical Principles and Practical Governance

A persistent challenge in AI governance is the disconnect between high-level ethical principles and their practical implementation in specific contexts, like education. Existing governance frameworks emphasise broad principles such as fairness, transparency, and beneficence, yet these abstract guidelines often fail to provide concrete solutions to address complex ethical dilemmas in real-world AIED adoption. As a result, regulation risks being too detached from the lived experiences of students, who, in practice, prioritise how these principles are enacted in relation to their autonomy, informed consent, and emotional well-being. What's needed is not just ethical principles, but mechanisms to translate them into meaningful protections in students' everyday educational experiences.

This thesis argues that governance should not merely dictate how AI operates but should ensure that students' genuine learning needs and rights, such as agency, are respected and protected. Instead of imposing ethical standards from the top down, policy mechanisms should create pathways for students to actively participate in governance discussions and decision-making about AIED implementation. This shift necessitates context-specific governance approaches, as advocated by Fox (2022), ensuring that policies accommodate the diverse educational environments, cultural differences, and localised ethical concerns that shape AIED's impact. Furthermore, ethical governance must prioritise a practical understanding of risks to break down and apply abstract ethical principles. Moving beyond general AI ethics frameworks, AIED governance should embed ethical considerations into every stage of design, development, and deployment, ensuring that ethical concerns are not an afterthought but a guiding force in shaping AIED's future.

7.3.3 Summary and Actionable Recommendations

Balancing the promises and challenges of AIED requires a fundamental shift in its design, development, and governance. While AIED holds the potential to foster personalised learning, empower learners, enrich learning experiences, enhance accessibility, and create more equitable educational opportunities, these benefits can only be realised if ethical risks and unintended consequences are proactively addressed. As this section has demonstrated, the risks of a one-size-fits-all approach, lack of rigorous evaluation, insufficient ethical reflexivity, and prioritisation of efficiency over the quality

of interactions present significant barriers to ethical AIED adoption. Table 6 summarises the core potential, challenges and recommendations for the ethical future of AIED.

Table 6. Potentials, challenges and recommendations for the future ethical AIED.

#	Potential	Challenges	Recommendation
1	AI-enabled personalised learning tailored to individual strengths and needs.	Risk of undermining autonomy; over-reliance on interventions; one-size-fits-all models.	Offer customisable settings; design interventions as suggestions rather than mandates; balance independence with guidance.
2	AIED as a supportive companion that empowers students.	Current systems often act as supervising tools for teachers, undermining student agency.	Reframe AIED as a learning partner that fosters collaborative decision-making among students, teachers, and AI; avoid intrusive monitoring.
3	Enhancing accessibility and equity through adaptive tools.	Standardised models and bias in interpreting behaviours risk reinforcing inequalities and disadvantaging less visible students.	Apply inclusive, context-sensitive, culturally responsive, and bias-aware design that supports diverse and marginalised groups.
4	Fostering genuine engagement and richer educational experiences beyond efficiency.	Overemphasis on task completion and performance scores reduces learning to optimisation.	Prioritise deep learning, critical thinking, and meaningful interactions over efficiency benchmarks in AIED design.
5	Responsible and ethical innovation.	Lack of rigorous evaluation and reflexivity in current deployments.	Establish holistic evaluation frameworks; embed ongoing ethical oversight.
6	Trustworthy adoption and governance.	Coerced consent, weak accountability, and abstract principles fail to protect students.	Ensure genuine informed consent; strengthen AI literacy among decision-makers; develop context-specific governance mechanisms.

Rather than focusing solely on technological optimisation, the future of AIED must adopt student-centred, participatory approaches that embed ethical considerations from the outset, rather than addressing them retrospectively. Governance frameworks must move beyond abstract ethical principles to develop meaningful, context-specific policies. This thesis advocates for a slower, more deliberate approach to AIED development - one that prioritises student well-being, ethical reflexivity, and human-centred values in both design and evaluation. By striking this balance, AIED can fulfil its potential to enrich education while safeguarding fundamental rights and ethical integrity.

7.4 METHODOLOGICAL REFLECTIONS ON STORY COMPLETION

This thesis employed story completion as a novel method to explore the ethical and sociotechnical implications of AIED adoption, marking, to the best of my knowledge, its first application within the AIED research community. Given its unique contribution, this subsection offers a focused reflection on the use of story completion in this context.

Story Stem Design: Minimal or Context-Rich?

In designing the story stems, inspiration was drawn from earlier HCI studies using story completion methods, many of which employed minimalistic, single-sentence prompts to provide a blank canvas for participant creativity. For example, in Cambre et al. (2020), the stem reads, “It’s the year 2050... Sam gets home from work and enters the living room, where Jamie is sitting. The voice assistant...” This approach left participants ample room to envision what the technology could be like freely. However, given my focus on capturing participants’ emotional responses, concerns, and attitudes toward a specific AIED system proposed and currently under advancement by research groups, more detailed scenarios were chosen. This approach guided participants into an immersive AI-driven context - a dream that I created for them - to better imagine living within it and respond more meaningfully.

In retrospect, this design choice underscores the importance of tailoring the story stem to the research purpose. Minimalist, open-ended prompts are effective for future-oriented design ideation (e.g., (Troiano et al., 2020; Troiano et al., 2021; Wood et al., 2017)). Context-rich scenarios (e.g., designed in this thesis), may be more appropriate when exploring lived or anticipated experiences with an established technology. The former approach is forward-looking and focused on the envisioned future; the latter, like mine, is retrospective, eliciting situated, emotionally grounded responses to realistic technology adoption.

Potential for Story Bias?

My initial review of participants’ responses revealed a notable shift from initial resistance in Scenario 1 (focused on individual monitoring) to a more accepting and even positive outlook by Scenario 3 (which featured AI-driven customisation). This pattern may reflect genuine differences in how students respond to various AIED features. However, upon reflection, it also raises the possibility of story bias: Scenario 3 may have

foregrounded more benefits than the earlier scenarios, inadvertently shaping participants' attitudes.

While the structured progression was intended to explore responses to distinct AIED functions (monitoring vs. customisation), the way different scenarios were sequenced and framed may have influenced participants' interpretations. This highlights an inherent challenge in speculative research methods: how narrative framing shapes participant reactions. To examine this further, Study 3 was designed to validate these patterns under more controlled conditions.

In response to the potential bias concern in the story completion methods, I extensively reviewed previous story completion studies and found that none have explicitly discussed story bias – perhaps due to their use of minimal prompts, which limits the language influence. Existing studies generally advocate for careful story design to ensure clarity and avoid narrowing participant responses (Clarke et al., 2019). This thesis thus contributes to an under-discussed area by raising methodological questions about how story prompt framing influences participants' speculative responses.

While acknowledging the potential for story bias, it does not undermine the findings. Rather, participants' negative reactions offer a lens into worst-case scenarios, pointing to ethical risks in AIED adoption, informing more responsible design. From a constructionist perspective, the research aim is not to uncover the objective truth but to surface diverse viewpoints and anticipate potential concerns. Acknowledging story bias serves not as a flaw but as a reminder of the interpretive nature of speculative methods, and a valuable opportunity to extend methodological discussion within the HCI and AIED communities.

Chapter 8: Conclusion

This chapter begins with a synthesis of the thesis contributions (Section 8.1), followed by a discussion of limitations and future directions (Section 8.2), then concludes with closing reflections (Section 8.3), bringing the chapter and the thesis to a close.

8.1 CONTRIBUTIONS

To consolidate the contributions, Table 7 summarises the key outcomes derived from Studies 1–3 and the integrative discussion. While each study made distinct advances, with evidence detailed in the results chapters and associated publications, the synthesis highlights how these findings collectively advance understanding of the ethical challenges of AIEd adoption. The table cross-references where each contribution is elaborated in the thesis (tables and figures within the respective chapters). Together, these contributions span conceptual (frameworks and models), methodological (validated scale), and practical (design implications and recommendations) advances, culminating in the proposed student-centred ethics in AIEd framework.

Table 7. Summary of Thesis Contributions

Key Contribution	Source
Contribution 1: Identified three novel ethical dimensions of potential consequences of AIEd adoption, including individual, pedagogical, and classroom dynamics, which form a comprehensive thematic framework and basis for a conceptual model to anticipate and evaluate risks.	Chapter 4 (Study 1 findings) Table 2. (Page 57)
Contribution 2: Developed a power dynamics model mapping authority and knowledge imbalances among stakeholders in AIEd design and adoption, explaining how ethical gaps arise and how they may be mitigated.	Chapter 5 (Study 2 findings) Fig. 4 (Page 137)
Contribution 3: a) Developed and validated a scale for assessing AIEd across learning beliefs and affective responses. b) Identified students' ranked ethical concerns (Autonomy and Privacy most salient). c) Proposed five student-centred design implications for AIEd.	Chapter 6 (Study 2 findings) Tables 3,4,5 (Pages 175-176)
Contribution 4: Proposed the Ethics of Human Factors framework for AIEd, reframing the “unknown unknowns” in existing ethics discourse, and extending existing frameworks by integrating human, relational, and contextual dimensions essential for ethical AIEd adoption.	Chapter 7 (Discussion) Fig.5 (Page 214)

Contribution 5: Offered a set of recommendations for balancing potentials and challenges in the future design and adoption of ethical AIEd.	Chapter 7 (Discussion) Table 6 (Page 227)
---	--

8.2 LIMITATIONS AND FUTURE DIRECTIONS

As with any research, there are limitations that suggest directions for further investigation. The following discussion outlines potential limitations related to the research approach, research context and scope, and philosophical and ethical positioning, along with future research avenues.

Research Approach

This thesis examines multiple stakeholder perspectives (students, parents, school representatives, and developers), with each group surveyed or interviewed separately. While this approach provides deep insights into their concerns, while their tensions are analysed and discussed, it may not fully capture the potential negotiations and dynamics that might emerge in direct interaction. Future research could extend this work by using participatory methodologies such as focus groups or stakeholder workshops, which bring students, teachers, developers, and school governors into dialogue. This would help identify whether conflicting priorities can be solved through discussion and whether engaging in a shared conversation influences each group's ethical stance on AIEd adoption.

This thesis primarily employs a speculative research approach using self-reported data, drawing on hypothetical AIEd scenarios (Chapter 4, Chapter 5) and video prototypes (Chapter 6) to probe participants' concerns and responses. These responses are based on their perceptions rather than direct experience. However, this should not be seen as a limitation. Rather, it is a deliberate and appropriate choice for the exploratory stages of studying controversial technologies in sensitive settings such as education. This approach reduces the risk of harm to participants, allowing for ethical, pre-emptive inquiry without requiring direct exposure of participants to these technologies. More realistic, experiential assessment of how students and other stakeholders actually interact with AIEd should follow speculative methods – once it is ethically assured that such interactions will not cause harm. Combining speculative and experiential approaches would enable future studies to examine the real-world implications of AIEd adoption while upholding ethical standards.

Similarly, this thesis does not focus on longitudinal data to track how student perceptions and reactions evolve as AIEd is implemented in real-world settings. The studies presented in this thesis primarily capture students' anticipatory perceptions and reactions (Chapter 4) and immediate attitudes (Chapter 6). While this approach allows for early intervention in ethical design, it is important to acknowledge that students' perceptions may shift over time as they could potentially adapt to these technologies, or as institutions modify their use based on feedback. Longitudinal studies could provide a dynamic understanding of AIEd's long-term ethical and social impacts. However, a critical ethical concern must be noted: longitudinal research should not involve technologies that have already been identified as harmful to students' well-being, even in the short term. For instance, intrusive monitoring-driven AIEd tools, which elicited strong negative emotional responses in this thesis, should not be deployed simply for the sake of long-term observation. Doing so would unethically compromise student well-being. Instead, longitudinal studies should focus on alternative, non-invasive AI tools to assess how students interact with ethical AIEd systems.

Research Context and Scope

This thesis collected data from multiple educational settings (higher education and K-12) and cultural contexts (Chinese and Australian), but did not conduct a systematic comparison. While Study 1 and Study 3 focused on university students, Study 2 shifted attention to K-12 education by including parents, school decision-makers, and developers. This shift aimed to capture the complex power dynamics in educational systems where students are typically younger, more vulnerable, and possess less autonomy. Although ethical concerns identified in higher education where student involvement is often optional, remain highly relevant, they become even more pressing in K-12 settings, where participation is compulsory. However, a limitation of this thesis lies in the asymmetry of stakeholder representation: K-12 students' voices were not directly included. Given their distinct levels of agency, cognitive development, and ethical sensitivities, future research should examine how younger students perceive AIEd adoption, particularly in relation to privacy, surveillance, and learning autonomy, and how these perspectives might differ from those of university students.

Study 1 and Study 2 were conducted in China (due to COVID-19 travel restrictions), while Study 3 was conducted in Australia. Though Study 1 involved participants from both countries, the initial data screening indicated substantial overlap in ethical concerns

across both cultural contexts. This led the analysis to focus on shared ethical challenges. As a result, this thesis does not conduct a systematic cultural comparison of AIED perceptions between China and Australia. Future studies could take a comparative approach, systematically investigating whether cultural, socio-political, or institutional differences shape AIED perceptions in meaningful ways. This would help determine whether ethical concerns around AIED are universal or highly context-dependent.

A key finding of Study 1 was the diversity in students' responses to AIED adoption. While many students expressed privacy concerns about AIED collecting their behavioural and emotional data, I observed that some - especially those lacking confidence in their learning abilities - were more willing to compromise privacy in exchange for personalised academic support. This variation highlights the complexity of student attitudes and suggests that AIED adoption is not experienced uniformly. However, this study did not systematically investigate the underlying factors that shape these trade-offs. It remains unclear which specific elements - such as students' learning confidence, past experiences with AI, trust in institutional oversight, or the level of control they feel over their data - drive their willingness to accept or resist AIED collecting their data and intervening in their learning. Future research should explore these factors in depth, identifying the conditions under which students feel comfortable engaging with AIED and the ethical boundaries they perceive as necessary. Such insights could inform the development of more personalised, adaptive AIED systems that accommodate diverse student needs while upholding ethical standards of autonomy and privacy.

A key highlight of Study 2 was that conventional participatory design approaches may fall short in real-world contexts shaped by pronounced power asymmetries and varying value priorities. While this thesis identifies these structural issues, it did not directly pursue design interventions to address them. Future research could build on this by exploring adapted forms of participatory design that explicitly account for these power differentials and ensure the meaningful inclusion of disempowered users, particularly students. Similarly, future research could investigate interventions or governance models that aim to realign stakeholder value priorities. Approaches such as participatory deliberation workshops, ethical impact assessments, or governance audits could serve as testbeds for designing more inclusive and accountable AIED decision-making processes.

Philosophical and Ethical Limitations

This thesis discussed findings with key Western ethical principles (e.g., autonomy, privacy, beneficence) and AI ethics frameworks. However, ethical considerations surrounding AI are not universal; different cultures prioritise different moral and epistemological concerns. For example, collectivist societies like in China may weigh community well-being more heavily than individual autonomy (Maznevski et al., 2002), leading to different interpretations of ethical AIED adoption. Study 2 was conducted in the Chinese context. While this cultural difference towards privacy was acknowledged, this thesis did not dive deep to analyse these considerations according to non-Western ethical philosophies – such as Confucian ethics. Future research could incorporate non-Western ethical philosophies to broaden AI ethics beyond Western paradigms. AIED governance should account for pluralistic ethical perspectives to ensure global inclusivity.

A notable contribution of this thesis is the emphasis on students' emotional well-being - an ethical dimension often overlooked in existing AIED ethics frameworks. While this thesis highlights its importance, it has not yet fully conceptualised the ethical dimension of emotional well-being or operationalised it within design and governance practices. Future research could build on this foundation by developing the concept of emotional ethics more systematically. Advancing this work would help move human-centred ethical frameworks beyond abstract principles toward more situated, affective dimensions of students' experiences.

8.3 CONCLUDING REMARKS

This thesis began with identifying four critical problems (in Chapter 1): a lack of understanding of AIED's unintended consequences; the lack of understanding of students' perspectives, despite being the primary recipient of AIED's impact; limited understanding of how stakeholder power dynamics shape ethical outcomes; and a narrow technology-centric framing of AIED ethics. In further exploring these problems, Chapter 2 reviewed literature from disciplines across AIED, HCI, and AI Ethics. It revealed a lack of empirical research on stakeholders' lived experiences towards AIED and highlighted that the existing ethical frameworks fail to address human-centred concerns. These gaps underscore the need for a human-centred approach to AIED design, adoption and governance, grounded in empirical evidence. In response, Chapter 3 outlined the thesis's overarching methodological approach, guided by a constructionist epistemology and

symbolic interactionism. This approach frames ethical challenges as emerging from stakeholders' interactions with AIED technologies. Grounded in this approach, this thesis aims to examine how different stakeholder groups perceive and respond to AIED adoption and its impacts, expand the ethical discourse through a human-centred lens, and offer actionable recommendations for ethical AIED. To address these aims, various methods are adopted across three empirical studies, drawing on story completion with 71 students, semi-structured interviews with 30 powerful stakeholders, and online experiment surveys with 131 students.

The three empirical studies, presented in Chapters 4, 5 and 6 through their respective publications and manuscripts, collectively reveal a cohesive and layered understanding of the ethical challenges surrounding AIED adoption. Chapter 4 demonstrated how students' concerns span psychological, behavioural, pedagogical and social dimensions, underscoring the interdependence of the situated impacts, and the importance of addressing the nuanced social elements in education when adopting AIED. Chapter 5 highlighted how students' concerns are often overlooked by more powerful stakeholders – developers, school decision-makers, and parents. It revealed how their decisions are shaped by divergent priorities within complex power dynamics, leading to ethical blind spots and accountability gaps that risk exposing students to unintended harm. While key decision-makers focus on learning gains and expect students to overcome their privacy concerns and emotional discomfort for potential benefits, Chapter 6 demonstrated that students prioritise autonomy and privacy over learning benefits, and desire control over how they engage with AIED and its adoption. Together, these findings revealed a troubling misalignment between value trade-offs presumed by powerful stakeholders and students' lived realities - one that must be addressed if AIED is to support, rather than undermine, students' educational well-being.

The Discussion Chapter consolidates these findings to illustrate how ethical challenges in AIED are not merely technical, but deeply sociotechnical and relational. It argues that ethical risks emerge not only from the design of technologies themselves, but also from the situated educational context where AIED is implemented, as well as the governance structures and power dynamics that shape its adoption. A fundamental and systematic issue is whether students' needs and concerns are meaningfully considered in AIED design, whether governance mechanisms mitigate the unintended harms of AIED adoption to effectively protect students, and whether student acceptance is genuine – not

simply assumed. In this thesis study, AIEd systems are designed under the premise of serving students' best interests while their perspectives are marginalised, despite students being the most affected by these technologies. Addressing these challenges requires embedding human-centred values, particularly emotional well-being and autonomy, throughout the entire lifecycle of AIEd design, adoption and governance. This thesis contributes to this fundamental shift by advocating for more proactive and inclusive frameworks that foreground students' lived experiences and support ethical, context-sensitive technology adoption.

Looking ahead, this thesis marks only the beginning of a broader effort to inform more human-centred, ethically-grounded AIEd development and adoption. Critical future work is needed to explore how lived experiences, user values and student agency can be meaningfully embedded into real-world decision-making processes. A key area of inquiry is how stakeholder power imbalances can be mitigated and how marginalised stakeholder groups like students can be empowered throughout AIEd design and governance. This could include developing practical interventions to operationalise inclusive governance frameworks, ensuring that student perspectives are not only heard but meaningfully integrated. Future research should also focus on building accountability mechanisms that enable more transparent and ethically robust oversight across AIEd development, implementation and evaluation stages. It could develop tools to equip key responsible decision-makers with AI literacy to help them in making informed decisions. Moreover, a deeper investigation is needed into the diverse factors that shape students' willingness to engage with AIEd, such as their learning confidence, trust in institutional processes, and perceived sense of control. Finally, while this thesis expands the ethical discourse around AIEd by proposing emotional well-being as a central element of human-centred AIEd ethics, this represents only a first step. Further work is needed to develop and operationalise emotional well-being as an ethical concern, both theoretically and practically, to ensure that AIEd systems support not only learning outcomes but also the holistic well-being of students in technology-mediated environments.

Reflecting on my positionality as both a student and a researcher contributing to the development of AIEd, this thesis carries academic as well as personal significance. As introduced in Chapter 1, my motivation has been shaped by a desire to ensure that future AI technologies in education will not only enhance learning outcomes but also respect and support students' needs and are genuinely welcomed by students. I believe students

should not feel that technology disrupts their well-being or marginalises their voices in shaping their future of education. They should not be treated as passive recipients of technology adoption but as active participants in decisions that affect their learning lives. This research journey has deepened my understanding of the tensions between technological innovation and ethical responsibility. This thesis is ultimately a product of that tension - a reflection of my conviction that ethical design is not only about building better systems, but about creating technologies that honour human dignity, agency, and context and that ultimately enrich everyday life.

A final caution this thesis offers, not only for AIEd but also for other emerging technologies introduced in high-stakes sensitive settings such as AI in healthcare, finance, and the workplace (as identified by AI Ethicists (Campolo et al., 2017)):

1. *Ethical practice must start at the outset.* AI adoption in sensitive settings demands a proactive, not reactive, approach. Ethical considerations must be embedded early in the process of innovation, design, and implementation, rather than retrofitted after harm has occurred. This requires slowing down the pace of technological rollout, grounding decisions in empirical evidence, and anticipating unintended consequences before deployment. Speculative methods can help surface hidden issues and communicate alternative futures (Vilaza, 2022). Designers must account for the situated realities within the context where technology is implemented - its social dynamics, contextual diversity, and the emotional stakes it carries.
2. *Reframe user-centred design: Beyond functionality, design for experience.* While early-stage ethics is essential, it must also shape what we design for. Technology innovation should not be driven solely by efficiency or task performance, it must enrich human experience and foster meaningful interactions. User-centred design and evaluation must move beyond usability to consider users' lived experience, the value trade-offs involved, and whether such compromises are acceptable, the deeper and nuanced impacts on users' well-being and whether the intended benefits truly outweigh potential harms.
3. *Centre well-being in ethical frameworks.* While ethical design must begin early and centre user experience, this thesis highlights a critical yet often overlooked dimension: the emotional impacts of AI systems. Students' feelings of anxiety and discomfort are not incidental, they shape how students experience and

respond to AIEd. We need to recognise these affective dimensions as integral to students' well-being. Ethical frameworks must move beyond technical safeguards to prioritise human-centred values and needs, especially students' emotional well-being and holistic experience. This ensures that AIEd systems genuinely promote students' engagement and flourishing.

Finally, as we stand at the threshold of an increasingly AI-mediated future, it is our shared responsibility to ensure that these technologies empower rather than silence, support rather than undermine, and ultimately serve the people they are meant to benefit, starting with those whose voices have too often been left unheard.

References

- Ahuja, K., Kim, D., Xhakaj, F., Varga, V., Xie, A., Zhang, S., Townsend, J. E., Harrison, C., Ogan, A., & Agarwal, Y. (2019). EduSense: Practical classroom sensing at Scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3), 1-26. <https://doi.org/10.1145/3351229>
- Ahuja, K., Shah, D., Paredy, S., Xhakaj, F., Ogan, A., Agarwal, Y., & Harrison, C. (2021). *Classroom Digital Twins with Instrumentation-Free Gaze Tracking* Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan. <https://doi.org/10.1145/3411764.3445711>
- Alexander, B., Ashford-Rowe, K., Barajas-Murph, N., Dobbin, G., Knott, J., McCormack, M., Pomerantz, J., Seilhamer, R., & Weber, N. (2019). *Horizon report 2019 higher education edition* (ISBN 978-1-933046-02-0). <https://library.educause.edu/resources/2019/4/2019-horizon-report>
- Almaraz-López, C., Almaraz-Menéndez, F., & López-Esteban, C. (2023). Comparative study of the attitudes and perceptions of university students in business administration and management and in education toward artificial intelligence. *Education Sciences*, 13(6), 609. <https://doi.org/10.3390/educsci13060609>
- Alyuz, N., Okur, E., Genc, U., Aslan, S., Tanriover, C., & Esme, A. A. (2017). *An unobtrusive and multimodal approach for behavioral engagement detection of students* Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education, Glasgow, UK. <https://doi.org/10.1145/3139513.3139521>
- Alzoubi, D., Kelley, J., Baran, E., Gilbert, S. B., Ilgu, A. K., & Jiang, S. (2021). *TeachActive Feedback Dashboard: Using Automated Classroom Analytics to Visualize Pedagogical Strategies at a Glance* Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan. <https://doi.org/10.1145/3411763.3451709>
- Aragon, C., Hutto, C., Echenique, A., Fiore-Gartland, B., Huang, Y., Kim, J., Neff, G., Xing, W., & Bayer, J. (2016). Developing a research agenda for human-centered data science. Proceedings of the 19th ACM conference on computer supported cooperative work and social computing companion. <https://doi.org/10.1145/2818052.2855518>
- Arroway, P., Morgan, G., O'Keefe, M., & Yanosky, R. (2015). *Learning analytics in higher education*. <https://doi.org/10.1002/AEHE.20121>
- Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. Artificial intelligence in education. <https://doi.org/10.3233/978-1-60750-028-5-17>
- Ashwin, T. S., & Biswas, G. (2024). Identifying and Mitigating Algorithmic Bias in Student Emotional Analysis. In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt, *Artificial Intelligence in Education* Cham. https://doi.org/10.1007/978-3-031-64299-9_7
- Aslan, S., Alyuz, N., Tanriover, C., Mete, S. E., Okur, E., D'Mello, S. K., & Arslan Esme, A. (2019). Investigating the impact of a real-time, multimodal student engagement analytics technology in authentic classrooms. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. <https://doi.org/10.1145/3290605.3300534>
- Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning*, 20(2), 13-29. <https://files.eric.ed.gov/fulltext/EJ1105911.pdf>
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, 563(7729), 59-64. <https://doi.org/10.1038/s41586-018-0637-6>

- Baker, R. S. (2016). Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education*, 26, 600-614. <https://doi.org/10.1007/s40593-016-0105-0>
- Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052-1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of educational data mining*, 1(1), 3-17. <https://doi.org/10.5281/zenodo.3554657>
- Baker, T., & Smith, L. (2019). *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges*. https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf
- Bartoletti, I. (2022). AI in education: An opportunity riddled with challenges. In *The ethics of artificial intelligence in education* (pp. 74-90). Routledge. <https://doi.org/10.4324/9780429329067>
- Bearman, M. (2019). Eliciting rich data: A practical approach to writing semi-structured interview schedules. *Focus on Health Professional Education: A Multi-Professional Journal*, 20(3), 1-11. <https://search.informit.org/doi/10.3316/informit.002757698372666>
- Beijing AI Principles. (2019). <https://ai-ethics-and-governance.institute/beijing-artificial-intelligence-principles/>
- Berendt, B., Littlejohn, A., & Blakemore, M. (2020). AI in education: Learner choice and fundamental rights. *Learning, Media and Technology*, 45(3), 312-324. <https://doi.org/10.1080/17439884.2020.1786399>
- Bond, M., Zawacki - Richter, O., & Nichols, M. (2019). Revisiting five decades of educational technology research: A content and authorship analysis of the British Journal of Educational Technology. *British Journal of Educational Technology*, 50(1), 12-63. <https://doi.org/10.1111/bjet.12730>
- Borning, A., & Muller, M. (2012). Next steps for value sensitive design. Proceedings of the SIGCHI conference on human factors in computing systems. <https://doi.org/10.1145/2207676.2208560>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589-597. <https://doi.org/10.1080/2159676x.2019.1628806>
- Braun, V., Clarke, V., Hayfield, N., Frith, H., Malson, H., Moller, N., & Shah-Beckley, I. (2019). Qualitative story completion: Possibilities and potential pitfalls. *Qualitative Research in Psychology*, 16(1), 136-155. <https://doi.org/10.1080/14780887.2018.1536395>
- Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). *Answers to frequently asked questions about thematic analysis*. <https://cdn.auckland.ac.nz/assets/psych/about/our-research/documents/Answers%20to%20frequently%20asked%20questions%20about%20thematic%20analysis%20April%202019.pdf>
- Brossi, L., Castillo, A. M., & Cortesi, S. (2022). Student-centred requirements for the ethics of ai in education. In *The Ethics of Artificial Intelligence in Education* (pp. 91-112). Routledge. <https://doi.org/10.4324/9780429329067>
- Brown, B., Taylor, A. S., Izadi, S., Sellen, A., Kaye, J. J., & Eardley, R. (2007). Locating family values: A field trial of the Whereabouts Clock. International Conference on Ubiquitous Computing. https://doi.org/10.1007/978-3-540-74853-3_21
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Conference on fairness, accountability and transparency. http://proceedings.mlr.press/v81/buolamwini18a.html?mod=article_inline&ref=akusion-ci-shi-dai-bizinesumedeia

- Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & quantity*, 56(3), 1391-1412. <https://doi.org/10.1007/s11135-021-01182-y>
- Cambre, J., Reig, S., Kravitz, Q., & Kulkarni, C. (2020). "All Rise for the AI Director": Eliciting Possible Futures of Voice Technology through Story Completion Proceedings of the 2020 ACM Designing Interactive Systems Conference, Eindhoven, Netherlands. <https://doi.org/10.1145/3357236.3395479>
- Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). *AI now 2017 report*. https://ainowinstitute.org/AI_Now_2017_Report.pdf
- Carbonell, J. R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE transactions on man-machine systems*, 11(4), 190-202. doi: 10.1109/TMMS.1970.299942.
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *Ieee Access*, 8, 75264-75278. doi:10.1109/ACCESS.2020.2988510
- Chu, H.-C., Hwang, G.-H., Tu, Y.-F., & Yang, K.-H. (2022). Roles and research trends of artificial intelligence in higher education: A systematic review of the top 50 most-cited articles. *Australasian Journal of Educational Technology*, 38(3), 22-42. <https://doi.org/10.14742/ajet.7526>
- Clarke, V., Braun, V., Frith, H., & Moller, N. (2019). Editorial introduction to the special issue: using story completion methods in qualitative research. In (Vol. 16, pp. 1-20): Taylor & Francis. <https://doi.org/10.1080/14780887.2018.1536378>
- Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., & Terem, E. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns*, 4(10). DOI: 10.1016/j.patter.2023.100857
- Cox, A., Cameron, D., Checco, A., Herrick, T., Mawson, M., & Steadman-Jones, R. (2022). Criteria of quality in fiction-based research to promote debate about the use of AI and robots in Higher Education. *Higher Education Research & Development*, 1-15. <https://doi.org/10.1080/07294360.2022.2087603>
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.
- Crotty, M. (1998). The foundations of social research: Meaning and perspective in the research process. In: Sage.
- D'mello, S., & Graesser, A. (2013). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(4), 1-39. <https://doi.org/10.1145/2395123.2395128>
- Denzin, N. K. (1974). The methodological implications of symbolic interactionism for the study of deviance. *The British journal of sociology*, 25(3), 269-282. <https://doi.org/10.2307/589394>
- Denzin, N. K. (2012). Triangulation 2.0. *Journal of mixed methods research*, 6(2), 80-88. <https://doi.org/10.1177/1558689812437186>
- Drachsler, H., & Greller, W. (2016). *Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics* Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, Edinburgh, United Kingdom. <https://doi.org/10.1145/2883851.2883893>
- Du Boulay, B. (2016). Artificial intelligence as an effective classroom assistant. *IEEE Intelligent Systems*, 31(6), 76-81. DOI: 10.1109/MIS.2016.93
- Engeström, Y. (2000). Activity theory and the social construction of knowledge: A story of four umpires. *Organization*, 7(2), 301-310. <https://doi.org/10.1177/135050840072006>
- European Commission. (2019). *Ethics guidelines for trustworthy AI*. <https://ec.europa.eu/futurium/en/ai-alliance-consultation/>
- Ez-Zaouia, M., Tabard, A., & Lavoué, E. (2020). EMODASH: A dashboard supporting retrospective awareness of emotions in online learning. *International Journal of Human-Computer Studies*, 139, 102411. <https://doi.org/10.1016/j.ijhcs.2020.102411>

- Falcão, T. P., Mello, R. F., Rodrigues, R. L., Diniz, J. R. B., Tsai, Y.-S., & Gašević, D. (2020). Perceptions and expectations about learning analytics from a Brazilian higher education institution. Proceedings of the tenth international conference on learning analytics & knowledge, <https://doi.org/10.1145/3375462.3375478>
- Fitzpatrick, G., & Ellingsen, G. (2013). A review of 25 years of CSCW research in healthcare: contributions, challenges and future agendas. *Computer Supported Cooperative Work (CSCW)*, 22, 609-665. <https://doi.org/10.1007/s10606-012-9168-0>
- Flanagan, M., Howe, D. C., & Nissenbaum, H. (2005). Values at play: Design tradeoffs in socially-oriented game design. Proceedings of the SIGCHI conference on human factors in computing systems. <https://doi.org/10.1145/1054972.1055076>
- Floridi, L., & Cows, J. (2022). A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design*, 535-545. <https://doi.org/10.1002/9781119815075.ch45>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., & Rossi, F. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Fontaine, G., Cossette, S., Maheu-Cadotte, M.-A., Mailhot, T., Deschênes, M.-F., Mathieu-Dupuis, G., Côté, J., Gagnon, M.-P., & Dubé, V. (2019). Efficacy of adaptive e-learning for health professionals and students: a systematic review and meta-analysis. *BMJ open*, 9(8), e025252. <https://doi.org/10.1136/bmjopen-2018-025252>
- Fox, A. (2022). Educational research and AIED: Identifying ethical challenges. In *The Ethics of Artificial Intelligence in Education* (pp. 47-73). Routledge.
- Fynn, A. (2016). Ethical considerations in the practical application of the Unisa socio-critical model of student success. *The International Review of Research in Open and Distributed Learning*, 17(6). <https://doi.org/10.19173/irrodl.v17i6.2812>
- Glaser, B., & Strauss, A. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge. <https://doi.org/10.4324/9780203793206>
- Gray, G., Schalk, A. E., Cooke, G., Murnion, P., Rooney, P., & O'Rourke, K. (2022). Stakeholders' insights on learning analytics: Perspectives of students and staff? *Computers & Education*, 104550. <https://doi.org/10.1016/j.compedu.2022.104550>
- Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 15(3), 42-57. <http://www.jstor.org/stable/jeductechsoci.15.3.42>
- Haataja, E., Malmberg, J., & Järvelä, S. (2018). Monitoring in collaborative learning: Co-occurrence of observed behavior and physiological synchrony explored. *Computers in Human Behavior*, 87, 337-347. <https://doi.org/10.1016/j.chb.2018.06.007>
- Hakimi, L., Eynon, R., & Murphy, V. A. (2021). The Ethics of Using Digital Trace Data in Education: A Thematic Review of the Research Landscape. *Review of Educational Research*. <https://doi.org/00346543211020116>
- Hartung, P. (2020). The children's rights-by-design standard for data use by tech companies. *Unicef Good Governance of Children's Data Project*. Available online: <https://tinyurl.com/2s42h5k4> (accessed on 12 September 2022).
- Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review*, 6(1), 27-54. <https://doi.org/10.1016/j.edurev.2010.07.004>
- Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. In. Center for Curriculum Redesign. <https://discovery.ucl.ac.uk/id/eprint/10139722>
- Holmes, W., & Porayska-Pomsta, K. (2023). *The ethics of artificial intelligence in education*. Routledge Taylor.
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., & Bittencourt, I. I. (2021). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, 1-23. <https://doi.org/10.1007/s40593-021-00239-1>

- Holstein, K., McLaren, B. M., & Aleven, V. (2018). Student Learning Benefits of a Mixed-Reality Teacher Awareness Tool in AI-Enhanced Classrooms. In C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. du Boulay, *Artificial Intelligence in Education* Cham. https://doi.org/10.1007/978-3-319-93843-1_12
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms. In (pp. 157-171). Springer International Publishing. https://doi.org/10.1007/978-3-030-23204-7_14
- Hutt, S., Krasich, K., R. Brockmole, J., & K. D'Mello, S. (2021). Breaking out of the lab: Mitigating mind wandering with gaze-based attention-aware technology in classrooms. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3411764.3445269>
- Ismail, A., Thakkar, D., Madhiwalla, N., & Kumar, N. (2023). Public Health Calls for/with AI: An Ethnographic Perspective. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW2), Article 354. <https://doi.org/10.1145/3610203>
- Jatileni, C. N., Sanusi, I. T., Olaleye, S. A., Ayanwale, M. A., Agbo, F. J., & Oyelere, P. B. (2024). Artificial intelligence in compulsory level of education: Perspectives from Namibian in-service teachers. *Education and Information Technologies*, 29(10), 12569-12596. <https://doi.org/10.1007/s10639-023-12341-z>
- Jiang, W., & Pardos, Z. A. (2021). Towards equity and algorithmic fairness in student grade prediction. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. <https://doi.org/10.1145/3461702.3462623>
- Jiménez-Mijangos, L. P., Rodríguez-Arce, J., Martínez-Méndez, R., & Reyes-Lagos, J. J. (2023). Advances and challenges in the detection of academic stress and anxiety in the classroom: A literature review and recommendations. *Education and Information Technologies*, 28(4), 3637-3666. <https://doi.org/10.1007/s10639-022-11324-w>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, 1(9), 389-399. <https://doi.org/10.1038/s42256-019-0088-2>
- Johnson, R. B., & Christensen, L. B. (2024). *Educational research: Quantitative, qualitative, and mixed approaches*. Sage publications.
- Jones, K. M., Asher, A., Goben, A., Perry, M. R., Salo, D., Briney, K. A., & Robertshaw, M. B. (2020). "We're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology*, 71(9), 1044-1059. <https://doi.org/10.1002/asi.24358>
- Katsantonis, A., & Katsantonis, I. (2024). University Students' Attitudes Toward Artificial Intelligence: An Exploratory Study of the Cognitive, Emotional, And Behavioural Dimensions of AI Attitudes. <https://doi.org/10.3390/educsci14090988>
- Kay, J. (2012). AI and education: Grand challenges. *IEEE Intelligent Systems*, 27(5), 66-69. doi: 10.1109/MIS.2012.92
- Kaye, J. J. (2006). *I just clicked to say I love you: rich evaluations of minimal communication* CHI '06 Extended Abstracts on Human Factors in Computing Systems, Montréal, Québec, Canada. <https://doi.org/10.1145/1125451.1125530>
- Kim, Y., Soyata, T., & Behnagh, R. F. (2018). Towards emotionally aware AI smart classroom: Current issues and directions for engineering and education. *Ieee Access*, 6, 5308-5331. <https://doi.org/10.1109/ACCESS.2018.2791861>
- Koedinger, K. R., Brunskill, E., Baker, R. S., McLaughlin, E. A., & Stamper, J. (2013). New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine*, 34(3), 27-41. <https://doi.org/10.1609/aimag.v34i3.2484>
- Kučak, D., Juričić, V., & Đambić, G. (2018). MACHINE LEARNING IN EDUCATION-A SURVEY OF CURRENT RESEARCH TRENDS. *Annals of DAAAM & Proceedings*, 29. <https://doi.org/10.2507/29th.daaam.proceedings.059>
- Kvale, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. Sage.

- Latham, A., Crockett, K., McLean, D., & Edmonds, B. (2012). A conversational intelligent tutoring system to automatically predict learning styles. *Computers & Education*, 59(1), 95-109. <https://doi.org/10.1016/j.compedu.2011.11.001>
- Lawson, C., Beer, C., Rossi, D., Moore, T., & Fleming, J. (2016). Identification of ‘at risk’ students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology Research and Development*, 64(5), 957-968. <https://doi.org/10.1007/s11423-016-9459-0>
- Lawson, S., Kirman, B., Linehan, C., Feltwell, T., & Hopkins, L. (2015). Problematising upstream technology through speculative design: the case of quantified cats and dogs. Proceedings of the 33rd annual ACM conference on human factors in computing systems, <https://doi.org/10.1145/2702123.2702260>
- Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. *Computers and Education: Artificial Intelligence*, 2, 100016. <https://doi.org/10.1016/j.caeai.2021.100016>
- Li, W., Sun, K., Schaub, F., & Brooks, C. (2021). Disparities in students’ propensity to consent to learning analytics. *International Journal of Artificial Intelligence in Education*, 1-45. <https://doi.org/10.1007/s40593-021-00254-2>
- Lin, H.-C. K., Wu, C.-H., & Hsueh, Y.-P. (2014). The influence of using affective tutoring system in accounting remedial instruction on learning performance and usability. *Computers in Human Behavior*, 41, 514-522. <https://doi.org/10.1016/j.chb.2014.09.052>
- Linson, A., Xu, Y., English, A. R., & Fisher, R. B. (2022). Identifying Student Struggle by Analyzing Facial Movement During Asynchronous Video Lecture Viewing: Towards an Automated Tool to Support Instructors. International Conference on Artificial Intelligence in Education, https://doi.org/10.1007/978-3-031-11644-5_5
- Lu, A. J., Dillahunt, T. R., Marcu, G., & Ackerman, M. S. (2021). Data Work in Education: Enacting and Negotiating Care and Control in Teachers' Use of Data-Driven Classroom Surveillance Technology. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1-26. <https://doi.org/10.1145/3479596>
- Lu, A. J., Marcu, G., Ackerman, M. S., & Dillahunt, T. R. (2021). *Coding Bias in the Use of Behavior Management Technologies: Uncovering Socio-technical Consequences of Data-driven Surveillance in Classrooms* Designing Interactive Systems Conference 2021, Virtual Event, USA. <https://doi.org/10.1145/3461778.3462084>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. <https://discovery.ucl.ac.uk/id/eprint/1475756>
- Lyu, Q., Chen, W., Su, J., Heng, K. H. J. G., & Liu, S. (2023, 2023//). How Peers Communicate Without Words-An Exploratory Study of Hand Movements in Collaborative Learning Using Computer-Vision-Based Body Recognition Techniques. *Artificial Intelligence in Education*, Cham. https://doi.org/10.1007/978-3-031-36272-9_26
- Madaio, M., Blodgett, S. L., Mayfield, E., & Dixon-Román, E. (2022). Beyond “fairness”: Structural (in) justice lenses on ai for education. In *The ethics of artificial intelligence in education* (pp. 203-239). Routledge.
- Manolev, J., Sullivan, A., & Slee, R. (2019). The datafication of discipline: ClassDojo, surveillance and a performative classroom culture. *Learning, Media and Technology*, 44(1), 36-51.
- Martinez-Maldonado, R., Clayphan, A., & Kay, J. (2015). Deploying and Visualising Teacher’s Scripts of Small Group Activities in a Multi-surface Classroom Ecology: a Study in-the-wild. *Computer Supported Cooperative Work (CSCW)*, 24(2), 177-221. <https://doi.org/10.1007/s10606-015-9217-6>
- Martinez-Maldonado, R., Echeverria, V., Fernandez-Nieto, G., Yan, L., Zhao, L., Alfredo, R., Li, X., Dix, S., Jaggard, H., & Wotherspoon, R. (2023). Lessons learnt from a multimodal learning analytics deployment in-the-wild. *ACM Transactions on Computer-Human Interaction*, 31(1), 1-41. <https://doi.org/10.1145/3622784>
- Maznevski, M. L., Gomez, C. B., DiStefano, J. J., Noorderhaven, N. G., & Wu, P.-C. (2002). Cultural dimensions at the individual level of analysis: The cultural orientations

- framework. *International journal of cross cultural management*, 2(3), 275-295. <https://doi.org/10.1177/147059580223001>
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research.
- Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: A guidance for policymakers*. UNESCO Publishing.
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and engineering ethics*, 26(4), 2141-2168. <https://doi.org/10.1007/s11948-019-00165-5>
- Moundridou, M., & Virvou, M. (2002). Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning*, 18(3), 253-261. <https://doi.org/10.1046/j.0266-4909.2001.00237.x>
- Murchan, D., & Siddiq, F. (2021). A call to action: a systematic review of ethical and regulatory issues in using process data in educational assessment. *Large-scale Assessments in Education*, 9(1). <https://doi.org/10.1186/s40536-021-00115-3>
- Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, 11(1). <https://doi.org/10.3390/app11010237>
- Nawaz, S., Kennedy, G., Bailey, J., Mead, C., & Horodyskyj, L. (2018). Struggle town? Developing profiles of student confusion in simulation-based learning environments. 35th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education, ASCILITE.
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28(4), 4221-4241. <https://doi.org/10.1007/s10639-022-11316-w>
- Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. In *Algorithms of oppression*. New York university press. <https://doi.org/10.18574/nyu/9781479833641.001.0001>
- Nunes Vilaza, G., Doherty, K., McCashin, D., Coyle, D., Bardram, J., & Barry, M. (2022). A scoping review of ethics across SIGCHI. Proceedings of the 2022 ACM Designing Interactive Systems Conference, <https://doi.org/10.1145/3532106.3533511>
- Nussbaum, M. C. (2000). *Women and human development: The capabilities approach* (Vol. 3). Cambridge university press.
- Patidar, P., Ngoon, T. J., Zimmerman, J., Ogan, A., & Agarwal, Y. (2024). ClassID: Enabling Student Behavior Attribution from Ambient Classroom Sensing Systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(2), 1-28. <https://doi.org/10.1145/3659586>
- Peng, Y., Kikuchi, M., & Ozono, T. (2023). Development and Experiment of Classroom Engagement Evaluation Mechanism During Real-Time Online Courses. International Conference on Artificial Intelligence in Education, https://doi.org/10.1007/978-3-031-36272-9_48
- Pinto dos Santos, D., Giese, D., Brodehl, S., Chon, S.-H., Staab, W., Kleinert, R., Maintz, D., & Baeßler, B. (2019). Medical students' attitude towards artificial intelligence: a multicentre survey. *European radiology*, 29, 1640-1646. <https://doi.org/10.1007/s00330-018-5601-1>
- Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction Design: Beyond Human-Computer Interaction*. John Wiley.
- Prem, E. (2023). From ethical AI frameworks to tools: a review of approaches. *AI and Ethics*, 1-18. <https://doi.org/10.1007/s43681-023-00258-9>
- Pyett, P. M. (2003). Validation of qualitative research in the “real world”. *Qualitative health research*, 13(8), 1170-1179. <https://doi.org/10.1177/1049732303255686>
- Robertson, S., Nguyen, T., & Salehi, N. (2021). Modeling assumptions clash with the real world: Transparency, equity, and community challenges for student assignment algorithms. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, <https://doi.org/10.1145/3411764.3445748>
- Robson, C., & McCartan, K. (2002). *Real world research* (Vol. 2). Blackwell Oxford.

- Rodway, P., & Schepman, A. (2023). The impact of adopting AI educational technologies on projected course satisfaction in university students. *Computers and Education: Artificial Intelligence*, 5, 100150. <https://doi.org/10.1016/j.caeai.2023.100150>
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27. <https://doi.org/10.1002/widm.1075>
- Ruan, X., Palansuriya, C., & Constantin, A. (2023). Affective Dynamic Based Technique for Facial Emotion Recognition (FER) to Support Intelligent Tutors in Education. *International Conference on Artificial Intelligence in Education*, https://doi.org/10.1007/978-3-031-36272-9_70
- Schepman, A., & Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and general trust. *International Journal of Human-Computer Interaction*, 39(13), 2724-2741. <https://doi.org/10.1080/10447318.2022.2085400>
- Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, 78, 397-407. <https://doi.org/10.1016/j.chb.2017.06.030>
- Sclater, N. (2016). Developing a Code of Practice for Learning Analytics. *Journal of Learning Analytics*, 3(1), 16-42. <https://doi.org/10.18608/jla.2016.31.3>
- Seale, C., & Silverman, D. (1997). Ensuring rigour in qualitative research. *The European journal of public health*, 7(4), 379-384. <https://doi.org/10.1093/eurpub/7.4.379>
- Seidman, I. (2006). Interviewing as qualitative research: A guide for researchers in education and the social sciences. *Teachers College*.
- Sekhon, M., Cartwright, M., & Francis, J. J. (2017). Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC health services research*, 17(1), 1-13. <https://doi.org/10.1186/s12913-017-2031-8>
- Self, J. (1998). The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. *International Journal of Artificial Intelligence in Education*, 10, 350-364.
- Sha, L., Rakovic, M., Das, A., Gasevic, D., & Chen, G. (2022). Leveraging Class Balancing Techniques to Alleviate Algorithmic Bias for Predictive Tasks in Education. *IEEE Transactions on Learning Technologies*, 1-12. <https://doi.org/10.1109/tlt.2022.3196278>
- Shneiderman, B. (2010). *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380-1400. <https://doi.org/10.1177/0002764213498851>
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5), 30.
- Smolaks, M. (2019). *Chinese university deploys facial recognition tech to monitor student behavior*. Retrieved 13th April from <https://aibusiness.com/verticals/chinese-university-deploys-facial-recognition-tech-to-monitor-student-behavior>
- Southgate, E., Blackmore, K., Pieschl, S., Grimes, S., McGuire, J., & Smithers, K. (2019). Artificial intelligence and emerging technologies in schools.
- Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.
- Suh, W., & Ahn, S. (2022). Development and validation of a scale measuring student attitudes toward artificial intelligence. *Sage Open*, 12(2). <https://doi.org/10.1177/21582440221100463>
- Terry, G., & Hayfield, N. (2020). Reflexive thematic analysis. *Handbook of qualitative research in Education*. Edward Elgar Publishing Limited, 430-441.
- Terry, G., & Hayfield, N. (2021). *Essentials of thematic analysis*. American Psychological Association.
- The-Institute-for-Ethical-AI-in-Education. (2021). *The Ethical Framework for AI in Education*. The Institute for Ethical AI in Education. <https://www.buckingham.ac.uk/research/research-in-applied-computing/the-institute-for-ethical-ai-in-education/>
- Thompson, T. L., & Prinsloo, P. (2022). Returning the data gaze in higher education. *Learning, Media and Technology*, 1-13. <https://doi.org/10.1080/17439884.2022.2092130>

- Trainor, L. R., & Bundon, A. (2021). Developing the craft: Reflexive accounts of doing reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 13(5), 705-726. <https://doi.org/10.1080/2159676X.2020.1840423>
- Troiano, G. M., Wood, M., & Hartevelde, C. (2020). " And This, Kids, Is How I Met Your Mother": Consumerist, Mundane, and Uncanny Futures with Sex Robots. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, <https://doi.org/10.1145/3313831.3376598>
- Troiano, G. M., Wood, M., Sonbudak, M. F., Padte, R. C., & Hartevelde, C. (2021). "Are We Now Post-COVID?": Exploring Post-COVID Futures Through a Gamified Story Completion Method. Designing Interactive Systems Conference 2021, <https://doi.org/10.1145/3461778.3462069>
- Tsai, Y.-S., & Gasevic, D. (2017). Learning analytics in higher education---challenges and policies: a review of eight learning analytics policies. Proceedings of the seventh international learning analytics & knowledge conference, <https://doi.org/10.1145/3027385.3027400>
- Tsai, Y.-S., Moreno-Marcos, P. M., Tammets, K., Kollom, K., & Gašević, D. (2018). SHEILA policy framework: informing institutional strategies and policy processes of learning analytics. Proceedings of the 8th international conference on learning analytics and knowledge, <https://doi.org/10.1145/3170358.3170367>
- Tsai, Y.-S., Perrotta, C., & Gašević, D. (2020). Empowering learners with personalised learning approaches? Agency, equity and transparency in the context of learning analytics. *Assessment & Evaluation in Higher Education*, 45(4), 554-567. <https://doi.org/10.1080/02602938.2019.1676396>
- Tsai, Y.-S., Whitelock-Wainwright, A., & Gašević, D. (2020). *The privacy paradox and its implications for learning analytics* Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, Frankfurt, Germany. <https://doi.org/10.1145/3375462.3375536>
- Tschider, C. A. (2018). The consent myth: improving choice for patients of the future. *Wash. UL Rev.*, 96, 1505.
- Tzimas, D., & Demetriadis, S. (2021). Ethical issues in learning analytics: A review of the field. *Educational Technology Research and Development*, 69(2), 1101-1133. <https://doi.org/10.1007/s11423-021-09977-4>
- UNESCO. (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development*.
- UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14>
- Van Manen, M. (2016). *Researching lived experience: Human science for an action sensitive pedagogy*. Routledge.
- Veletsianos, G., Collier, A., & Schneider, E. (2015). Digging deeper into learners' experiences in MOOC s: Participation in social networks outside of MOOC s, notetaking and contexts surrounding content consumption. *British Journal of Educational Technology*, 46(3), 570-587. <https://doi.org/10.1111/bjet.12297>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425-478. <https://doi.org/10.2307/30036540>
- Vierhauser, M., Sauerwein, C., Antensteiner, T., & Hatmanstorfer, S. (2024). Learning Analytics Support in Higher-Education: Towards a Multi-Level Shared Learning Analytics Framework. Proceedings of the 16th International Conference on Computer Supported Education, <https://doi.org/10.5220/0012744400003693>
- Vivar, J. M. F., & Peñalvo, F. J. G. (2023). Reflexiones sobre la ética, potencialidades y retos de la Inteligencia Artificial en el marco de la Educación de Calidad (ODS4). *Comunicar: Revista científica de comunicacion y educacion*(74), 37-47.
- Walker, D., & Myrick, F. (2006). Grounded theory: An exploration of process and procedure. *Qualitative health research*, 16(4), 547-559. <https://doi.org/10.1177/1049732305285972>

- Wang, Z., Li, L., Zeng, C., & Yao, J. (2023). Student learning behavior recognition incorporating data augmentation with learning feature representation in smart classrooms. *Sensors*, 23(19), 8190. <https://doi.org/10.3390/s23198190>
- Whitelock-Wainwright, A., Tsai, Y.-S., Drachsler, H., Scheffel, M., & Gašević, D. (2021). An exploratory latent class analysis of student expectations towards learning analytics services. *The Internet and Higher Education*, 100818. <https://doi.org/10.1016/j.iheduc.2021.100818>
- Whitelock - Wainwright, A., Gašević, D., Tejeiro, R., Tsai, Y. S., & Bennett, K. (2019). The Student Expectations of Learning Analytics Questionnaire. *Journal of Computer Assisted Learning*, 35(5), 633-666. <https://doi.org/10.1111/jcal.12366>
- Wilkinson, R. G., & Pickett, K. (2009). *The spirit level: Why more equal societies almost always do better* (Vol. 6). Allen Lane London.
- Williamson, B. (2019). Datafication of education: a critical approach to emerging analytics technologies and practices. *Rethinking pedagogy for a digital age*, 212-226.
- Wong, P.-H. (2020). Cultural differences as excuses? Human rights and cultural values in global ethics and governance of AI. *Philosophy & Technology*, 33(4), 705-715. <https://doi.org/10.1007/s13347-020-00413-8>
- Wood, M., Wood, G., & Balaam, M. (2017). " They're Just Tixel Pits, Man" Disputing the'Reality'of Virtual Reality Pornography through the Story Completion Method. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, <https://doi.org/10.1145/3025453.3025762>
- World-Economic-Forum. (2019). *Generation AI: Establishing global Standards for children and AI*. <https://www.weforum.org/publications/generation-ai-establishing-global-standards-for-children-and-ai/>
- Yan, L., Martinez-Maldonado, R., Zhao, L., Li, X., & Gašević, D. (2023, 2023//). Physiological Synchrony and Arousal as Indicators of Stress and Learning Performance in Embodied Collaborative Learning. *Artificial Intelligence in Education*, Cham. https://doi.org/10.1007/978-3-031-36272-9_49
- Yang, K. B., Echeverria, V., Lu, Z., Mao, H., Holstein, K., Rummel, N., & Aleven, V. (2023). *Pair-Up: Prototyping Human-AI Co-orchestration of Dynamic Transitions between Individual and Collaborative Learning in the Classroom* Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany. <https://doi.org/10.1145/3544548.3581398>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1). <https://doi.org/10.1186/s41239-019-0171-0>

Appendices

Appendix A

Book Chapter – AI and Academic Integrity

Appendix A contains a reflection paper that expands on the findings of Study 1, which was accepted as a book chapter in *Artificial Intelligence and Academic Integrity: Navigating Ethical Challenges of AI in Education*. This paper explores the value of academic integrity in the era of AI monitoring in classrooms.

The publication draft begins on the following page.

The Values of Academic Integrity in the Era of Monitoring AI in Classrooms

Bingyi Han, University of Melbourne, Australia

Simon Coghlan, University of Melbourne, Australia

George Buchanan, RMIT University, Australia

Dana McKay, RMIT University, Australia

The integration of Artificial Intelligence in Education (AIEd) into classrooms warrants critical evaluation. AI-enabled tools aim to enhance learning experiences through continuously monitoring, collecting, and analysing student in-class behaviour and performance data, providing real-time evaluation and intervention. However, even when well-intentioned, this monitoring-oriented AIEd introduces complex ethical and pedagogical challenges, particularly concerning academic integrity. This chapter discusses findings from our study involving 71 university students using the story completion method, which illustrates how AI potentially influences students to modify their behaviours in ways that compromise fundamental values of academic integrity, including honesty, fairness, trust, respect, and responsibility. We expand the traditional focus on academic integrity – typically centered on assessable work and issues like cheating and plagiarism – to include student learning behaviours and processes increasingly subject to AI monitoring and evaluation. We explore the subtle yet significant behavioural consequences of adopting monitoring-oriented AIEd in the classroom, drawing on students' perspectives, and highlighting how AIEd might challenge the pedagogical and ethical values they aim to support. We call for academics, developers and institutions to address these implications proactively, advocate for a more comprehensive understanding of academic integrity in the AI era, ensuring that AI deployment enriches the educational experience without compromising ethical standards.

Keyword: Academic Integrity, AI Monitoring, Surveillance AI, AIEd, Student Perceptions, Behaviour Change, Ethics, Values

1. Introduction

AI has been implemented in classrooms to enhance learning and teaching processes (e.g., (Smolaks, 2019)). Like generative AI, AI in classrooms will likely be deeply involved in students' learning journeys. The applications range from detecting students' learning progress to facilitating real-time pedagogical interventions. Both academia and industry are investing in technical advancement of AI. This includes developing techniques for monitoring students' in-class facial and emotional expressions (Ruan et al., 2023), tracking eye direction or gaze (Linson et al., 2022, Ahuja et al., 2021, Hutt et al., 2021), analysing students' awareness and engagement during lectures (Holstein et al., 2017, Holstein et al., 2018, Ez-Zaouia et al., 2020), providing automatic evaluation and timely intervention for students' learning progress (Yang and Ogata, 2022), prediction and early warning for students' learning outcome (Queiroga et al., 2022, Nam and Samson, 2019), and adaptive personalised recommendations (Altun et al., 2022).

Some or many of these techniques may improve student learning. However, AIED also introduces significant challenges. These extend beyond longstanding ethical concerns over data misuse (Hakimi et al., 2021), privacy (Korir, 2022), bias (Baker and Hawn, 2022), and algorithm fairness (Jiang and Pardos, 2021a), to less obvious challenges that result from reshaping the pedagogical landscape (Han et al., 2023). For instance, AIED could change approaches to the evaluation of students' work and performances and shift both educators' and students' expectations surrounding learning. Therefore, our understanding of academic integrity must evolve (Curtis, 2021) to encompass relevant changing interactions between students and learning environments (e.g., (Han et al., 2022)).

Although there is no agreed single definition for academic integrity (Eaton, 2024), the concept typically invokes several important values. The Australian Government's Tertiary Education Quality and Standards Agency (TEQSA) (2022), for example, states that academic integrity is "the expectation that teachers, students, researchers and all members of the academic community act with: honesty, trust, fairness, respect and responsibility". These five core values, which were initially identified by the Center for Academic Integrity (1999), serve as the foundation for academic conduct.

We shall also treat these values as foundational to academic integrity. However, TEQSA's guidelines on behaviours supporting academic integrity primarily focus on student activities related to scholarly outputs and assessments. This approach to defining academic integrity – starting from the established values but zooming in on student misconduct in scholarly work (such as using AI for completing assignments and exams) - is prevalent in current discussions (e.g., (Cotton et al., 2023)). This conceptualisation of academic integrity is arguably too narrow.

In this chapter, we shall define academic integrity somewhat more broadly, by moving beyond the prevailing emphasis solely on *Student Work – one way of evaluating the learning results*, to also encompass *Learning Behaviour – the learning*

process, especially within the classroom setting, which may well become increasingly subject to AI monitoring and evaluation.

This chapter draws upon our empirical research findings, which demonstrate how AI in classroom settings can undermine the fundamental values of academic integrity (Center for Academic Integrity, 1999) during students' learning. By engaging 71 university students and recent graduates through the story completion method, this study explores students' responses to AI adoption in classrooms, revealing a major concern that AI may lead students to deviate from the moral principles that are integral to academic integrity.

The rest of the chapter has the following sections. In section 2, we examine the current discussion on academic integrity. In section 3, we detail our story completion method, data collection, and analysis process. Then, in section 4, we present the findings regarding potential behaviour changes amongst students, and discuss these unintended consequences with reference to the fundamental values of academic integrity. Our examination highlights the risks that AIED can pose to academic integrity under our more expansive definition.

2. Academic Integrity

A very recent paper argues for the absence of a “singular universal definition of academic integrity” since definitions can vary according to values, principles, virtues and culture (Eaton, 2024). Historically, the concept has been associated with academic (Macfarlane et al., 2014)) and student conduct (e.g., (Nuss, 1984)), focusing on deterring acts of plagiarism or cheating. Various institutions (e.g., (Australian Government TEQSA, 2022)) have proposed diverse definitions, with many built upon the Fundamental Values of Academic Integrity developed by the International Center for Academic Integrity (ICAI) (Eaton, 2024). Five values - honesty, trust, fairness, respect, and responsibility – were initially introduced in 1999 (Center for Academic Integrity, 1999). The framework was revisited and expanded in 2014 to include a sixth value, courage (International Center for Academic Integrity (ICAI), 2014). Meanwhile, the European Network for Academic Integrity (ENAI) presented academic integrity as adherence to “ethical and professional principles, standards, practices and consistent system of values” to inform decision-making and actions in education, research and scholarship” (Tauginienė et al., 2018). Approaching academic integrity from its emphasised values and ethical standards, its practical implications should be expansive to various aspects of educational practices.

Macfarlane et al. (2014) conducted a comprehensive literature review on academic integrity among academics, encapsulating professional conduct across teaching, research, and service domains. They understood integrity as a holistic attribute, underlining the ethical standards and professional practices of academics, that integrate various aspects of one's conduct to uphold honesty and ethical principles. The review reveals a predominant focus on misconduct and academic corruption with

ethics of doing research, highlighting issues like fabrication and falsification of results, plagiarism, conflicts of interest in research sponsorship, and other ethical breaches including peer review manipulation, unethical treatment of subjects, and nepotism or cronyism in academic promotions. Beyond research ethics, the review highlights unethical behaviours for teaching, including power misuse in student-teacher relationships, problematic dual relationships between faculty and students that blur professional and personal boundaries, and the divergent perceptions of ethical teaching conduct between faculty and students. In the service domain, concerns include resource mismanagement, unfair faculty evaluation processes (e.g., untransparent criteria affect promotion clarity), and commercialisation of academic positions such as selling complimentary textbooks copies.

It is evident that the discussion of academic integrity for academic professionals encompasses a broad spectrum of conduct according to certain norms and values. Some apply elements of ethical theory such as virtue ethics to identify moral principles for academic integrity (e.g., (Corlett, 2005, Nixon, 2004)). However, when it comes to students, the academic integrity discussion often narrows to expectations of behaviour directly related to their scholarly outputs.

Many academic discussions on student academic integrity often proceed without a clear delineation of the concept. This ambiguity becomes particularly evident when addressing students' academic misconduct, which is predominantly framed around cheating and plagiarism. For example, this very early paper (Nuss, 1984) alternatively used the terms "academic integrity" and "academic dishonesty", without defining those concepts, nor establishing a definitive list of behaviours constituting academic dishonesty. Nevertheless, Nuss (1984) indicated that many studies relating to academic dishonesty refer simply to cheating while pointing out that interpretations of cheating vary widely among individuals. The paper then identified fourteen distinct forms of academic dishonesty, broadening the topic to relevant behaviours that disrupt the integrity of scholarly outputs and assessments.

In a more contemporary analysis, Gallant (2017) framed academic integrity as a matter of teaching and learning, moving beyond mere prevention and punitive measures against student cheating. Though the article did not precisely define either academic integrity, misconduct, or cheating behaviours, it outlines strategies to reduce instances of student academic dishonesty by fostering learning-oriented environments and improving instruction. The article emphasises transitioning from a performance-oriented to a mastery-focused educational environment, criticising superficial assessments that promote surface learning for external goals like scores (Gallant, 2017). These considerations are grounded in the fundamental values of education.

Today, in the era of AI, discussions predominantly revolve around the use of generative AI, narrowing the scope of academic integrity to scholarly pursuits and assessments. Yet, scholars like Israel (2015) and Eaton (2024) argue that academic integrity encompasses far more than just plagiarism or exam cheating and extends beyond mere regulatory compliance.

Eaton (2024) has developed a Comprehensive Academic Integrity (CAI) framework, which expands upon the traditional understanding of academic integrity to a

broader interpretation that encompasses all aspects of education and involves all stakeholders in education, emphasising ethical decision-making as crucial across the academy. Eaton suggests that integrity should be interwoven into the fabric of educational cultures, rather than viewed as a set of rules for students to follow. We agree that academic integrity for students should transcend scholarly pursuits and assessment to include greater attention to learning behaviours fundamentally rooted in ethical and pedagogical values.

Our empirical research supports this broader view by revealing that: 1) the pressure on students to perform in environments using AI for evaluation can inadvertently promote actions somewhat akin to "cheating", as student performance becomes a metric for assessment; 2) the use of AI monitoring in academic settings may routinely foster students' dishonest practices, eroding the core values of academic integrity and causing students to deviate from their moral principles or to not acquire those values and principles in the first place. It is possible that such dishonesty might also slip from the learning process to producing scholarly work. Again, these findings highlight concerns beyond mere scholarly outputs, touching on practices deeply intertwined with ethical standards. In the next section, we explain how we conduct this research and analysis.

3. Research Method

The objective of this research is to explore students' responses to adopting monitoring AI in higher education classrooms. This chapter presents and discusses our findings regarding how AI adoption can undermine students' academic integrity. In this section, we will introduce our data collection via the story completion method and data analysis process. This research was approved by the Human Research Ethics Committee from the University of Melbourne (Ethics ID Number: 22256).

3.1 Story Completion Method

Current students have yet to fully experience AI in classroom settings, although widespread adoption is a goal within the AIED community. Despite the rapid advancement of these technologies, there is a glaring paucity of research exploring students' perspectives and AI's implications for students, particularly given their vulnerable status as direct recipients of such interventions. The ethical considerations surrounding the deployment of these technologies are paramount; exposing students to potential negative consequences without prior understanding and mitigation planning is irresponsible, especially in sensitive high-stakes environments like education. To navigate these ethical waters and predict potential outcomes, we adopted a user-centered speculative approach known as the Story Completion Method.

The Story Completion Method draws from the domains of speculative design (Dunne and Raby, 2013, Wong and Khovanskaya, 2018) and research fiction (Blythe et al., 2018, Troiano et al., 2016). It has been used to critically evaluate the development (Lawson et al., 2015) and ethical and moral dilemmas (Dunne and

Raby, 2013) of current or future technologies. Participants are provided with Story Completion Tasks and are asked to complete narratives. Their stories can reflect their socially and culturally constructed values (Kitzinger and Powell, 1995, Wood et al., 2017). Rather than queries about “What’s your perspective?”, the indirect approach of third-person storytelling (Clarke et al., 2019) encourages participants to express their concerns more comfortably and openly.

We designed story completion tasks¹ based on both existing classroom monitoring and observation AIED artefacts and methodological insights from AIED research, as previously introduced in Section 1. It is further informed by the recent literature-based design fiction studies (e.g., [27]), which examine the implications of AI adoption in higher education. Our story prompts crafted scenarios that immerse participants in AI-enabled monitoring and intervention within higher education classrooms, helping them to understand our research context. The scenarios encapsulate the educational data collection on students’ in-class performance and engagement, through using AI monitoring tools and forms of observation such as facial recognition and emotion recognition. The scenario then introduces the customised AIED functionalities which are enabled by the collected data, including real-time analysis, personalised recommendations, predictions and interventions.

To foster a wide range of imaginative responses, we deliberately infused a degree of ambiguity when describing the AIED artefacts and functions (Braun et al., 2019). The story prompts concluded with open-ended questions to stimulate participants’ reflection, for example, “What happens next? How would the character think and behave?”

3.1 Data Collection and Analysis

71 stories were collected through anonymous surveys hosted on the protected Qualtrics online platform. Participants were first introduced to the study through a plain language statement, followed by a consent form. Upon consenting, participants were presented with the story stems and invited to complete the stories. The survey also collected participant’s demographic information including gender, age, and teaching experience, if any. Our participants were either currently enrolled in university studies or had recently graduated. Out of 71 collected stories, six were excluded due to incompleteness. Of the remaining 65 participants, 54% were male and 46% were female, none identified as non-binary.

Given the diverse AIED experiences among university students in real-world contexts, we chose not to collect data on participants’ majors, specific AIED experience or AI literacy levels. Our focus was on broader concerns relevant to all types of students. The speculative approach facilitated the exploration of students’ socially and culturally constructed views on the potential impacts of AI, regardless of their direct experience with AIED technologies.

The first author led the data analysis, with the contested themes or interpretations discussed collaboratively among the research team. We employed the Thematic

¹ The full scenarios are presented at: <https://sites.google.com/view/storystems>

Analysis (Braun and Clarke, 2006) to sift through the narrative data. We identified both explicit and latent content to generate lower-level codes. These codes were then generated into higher-level themes, providing structured insights into the participants' perceptions and the nuanced implications of monitoring-oriented AIED.

4. Findings & Discussion

Our findings uncover a complex interplay between AI adoption and student behaviour and adjustment. This included: decreased interaction among students with teachers; cheating the surveillance; conformity and a kind of pander performativity (forced performance) to escape monitoring; and rebellion against the system (manipulating monitoring mechanisms). We will demonstrate students' potential adaptive strategies in section 4.1. Worryingly, each of these strategies may undermine the fundamental values of academic integrity, which we discuss in section 4.2.

4.1 Students' Interactive Strategies in Responding to AI Monitoring

The predominant narratives reveal that students are troubled by the lack of transparency in the AI monitoring and evaluation mechanisms. This opacity creates a sense of uncertainty about the extent and impact of AI-driven monitoring and assessment, fostering a perception of the classroom environment as akin to a surveillance panopticon. Within this perceived panopticon, students harbour fears of their inadequacies being detected and documented by AI technologies. Such apprehensions lead to students assuming stringent requirements alongside a high standard for behaviour, which we will detail shortly.

Students articulated feelings of omnipresent monitoring, describing it as “ubiquitous” (P6) and feeling as if there are “eyes staring at me all the time” (P37). They expressed concern that any moment of inattention or deviation from the learning task – described as being caught in a “loaf moment” – would be unforgivably recorded, with one noting that “distraction or idle time is not permitted” (P26). Consequently, students perceive the learning environment as intense and strict, necessitating increased efforts for learning and a perpetual state of alertness in the class. As participants often said, a student “must be constantly switched on” (P56) and “act as a perfect self” (P27), while feeling compelled to meet a perceived, albeit unarticulated, standard of evaluation. This situation is described with negative emotions, characterising the learning experience as stressful, exhausting, and overly demanding.

Amidst these stringent perceived rules and high expectations, a notable decline in student confidence was observed. Students are concerned about the potential for an AI system to label a student as inadequate in a subject area and to deem them “a poor performer...or a dumb student” (P67). This apprehension directly leads to students' behaviour changes in classroom participation. In the narratives, students hesitate to freely speak and ask questions in class since they are constantly observed by AI, lest their contributions be judged as subpar. Such an environment may lead to a

cautious approach towards in-class expressions and behaviours, with students constantly aiming to eschew errors. Furthermore, stories also commonly show decreased interaction and communication with peers, driven by a fear that social interactions might be misinterpreted by the AI as irrelevant or disruptive behaviour.

A significant portion of participants disclosed adopting what they termed as “smart strategies” (P27) to respond to this perceived learning panopticon. A common theme among these strategies was the aim of securing favourable performance evaluation scores, leading to a spectrum of behaviours ranging from strategic compliance with performativity to covert resistance. For instance, the necessity to adapt to this monitoring regime can compel some students to act and perform, exhibiting ostensibly perfect engagement behaviours. As a student described: “a smarter student is likely to become an actor and pretend with perfect performances” (P27). Other stories described specific actions such as feigning active participation, querying, meticulous note-taking, and maintaining an outwardly positive demeanour.

There are also more inventive strategies that attempt to circumvent the AI evaluative gaze through the use of what might be termed as ‘cheating tools’. These ranged from the conceptualisation of an “Emotional Self-Check Software” designed to mask any detectable signs of negative emotions (P18), to the compilation of a “High-Score Dictionary”, aggregating terms believed to trigger high scores for real-time learning performance evaluation (with keywords spoken in the group discussion)(P19). As P19 mentioned, after graduation the student can sell this high-score dictionary, narratives from P52 described that the student would buy high-score notes from upper-year students “as a cheat sheet” to navigate stringent evaluations.

There are also privacy-preserving attempts to obscure students’ true performance or emotional states, as they fear poor performance being recorded and making mistakes in front of AI evaluation. Techniques included the use of non-standard languages (P28) or “undetectable accents and dialects” (P31) to elude voice recognition, and the concealment of their genuine emotional or facial expressions “... by putting a poker face” (P42) to evade emotion recognition.

Notably, despite the ingenious nature of these adaptations, they were frequently accompanied by expressions of discomfort and dissonance. Students articulated these strategies as deviations from their preferred modes of learning and interaction, often describing their actions as “adopting behaviours that felt unnatural or misaligned with their authentic selves” (P66).

In contrast to adaptive strategies aimed at navigating or complying with AI monitoring, some stories described a rebellion or an overt resistance against the system. These reactions ranged from subtle sabotage to explicit acts of defiance, each undermining the efficacy of the AI in educational settings (AIEd systems) in distinct ways.

A notable method of resistance involved the intentional corruption of data, with students deliberately presenting flawed data, introducing inaccuracies to “contaminate the data model” (P30). This approach represents an indirect yet impactful attack on the core functionality of AIEd systems, questioning their reliability and validity. There are also direct aggressions such as “...he walked to the front of the classroom

and unplugged the power source of the monitoring system with anger” (P2), or even more extreme measures such as attacking the database.

Beyond disrupting the system outcome, some students sought to leverage the system to their advantage, embracing a mindset of “manipulating the system instead of being manipulated” (P38). This manipulation was manifested in various forms, from using the system's feedback to prompt instructional improvements—illustrated by a student encouraging exaggerated expressions of boredom to signal disengagement to the instructor (P37)—to exploiting predictive failure analyses for personal gain, as seen in a student's strategic underperformance in algebra to facilitate dropping the subject (P48). Additionally, a creative use of the system's tracking capabilities was described, with a student intentionally focusing on illustrations to influence the inclusion of more visual content in textbooks (P29).

These interactive strategies - ranging from compliance through performative conformity (or cheating the surveillance) to evasion and outright rebellion – not only affect the students on an individual level such as changing their behaviours and influencing how they perceive the learning environment, but also carry significant implications for pedagogical practice and academic integrity.

4.2 Disrupted Pedagogical Values and Undermined Academic Integrity

The narratives from our participants show the potential shifts in students' behaviour as they adopt to AI monitoring and evaluation in the classroom. It also raises concerns regarding key pedagogical values such as learner autonomy and genuine engagement: The constant presence of AI monitoring and evaluation “encourages” students to conform to expected behaviours, not out of a genuine desire to learn or engage, but out of fear or being assessed and even graded for poor performance. This coercion into performative acts of engagement due to being constantly assessed by AI – ranging from overt acting to subtle pretense – erodes learner autonomy, restricting students' freedom to freely explore, inquire, and learn in a manner that resonates with their individual curiosities and learning styles.

The pressure to adhere to stringent rules and expectations of compliance transforms the learning environment into one that is perceived as intimidating rather than inviting. The joy and pleasure traditionally associated with discovery and learning are supplanted by a climate of apprehension. As students navigate this surveilled space, their actions are dictated by a strategic calculation while they become cautious about their behaviours and expressions to avoid making mistakes and negative evaluations. This undermines the authenticity of the learning process.

These unintended consequences of implementing this AI monitoring and evaluation go beyond learning autonomy, learning process and environment, to the very foundation of academic integrity. As noted earlier, the fundamental values of academic integrity (International Center for Academic Integrity (ICAI), 2021) encompass honesty, trust, fairness, respect, and responsibility. We will now discuss how each of these core principles may be undermined by AI monitoring and evaluation.

Honesty

To start with, classrooms with AI-driven monitoring might discourage students' honesty. It could promote a performative learning environment where students may feel compelled to act or pretend to engage, rather than participating authentically. Moreover, instead of fostering an environment where students can openly admit to gaps in their understanding and seek genuine improvement, the presence of AI monitoring may encourage a facade of comprehension. These behaviours, driven by fear of negative evaluation, challenge the value of intellectual and personal honesty in learning. To be fair, students may sometimes feel justified in trying to fool or deflect the AI systems used on them without their full consent. Nonetheless, such action may harm their learning.

It would be ironic if AI monitoring sought to promote honesty but instead facilitated student dishonesty and cheating behaviour. Moreover, while AI monitoring in the classroom can encourage students' dishonesty behaviour which causes students to deviate from their moral principles in their learning processes, this dishonesty might slip into their conduct in producing scholarly work. By making students more used to being dishonest in classroom learning activities, AI monitoring may end up damaging academic integrity more generally.

Trust

Implementing monitoring AI into the classroom may influence students' trust. The cornerstone of trust in educational contexts is the encouragement and support of the free exchange of ideas within clear, understood boundaries. Trust enables collaboration and information sharing not only among students but also between students and teaching agents. However, AI monitoring, if perceived as invasive, can disrupt this trust, making students cautious about freely expressing their thoughts or asking questions for fear of being judged by the system. Students' trust in teachers and the learning environment may be another of AIED's casualties. Our findings indicate a decreased willingness to interact between students for educational or social purposes. Some students even pointed out that they used to learn a lot from peers but that this pedagogical relationship involving an open exchange of ideas could be interrupted by AI monitoring.

Fairness

Monitoring AI could also disrupt practices of fairness. Students who act, pretend, cheat or fool the AI monitoring and evaluation systems may achieve higher marks, while those who engage authentically may face unfairly negative outcomes. This might encourage students to manipulate AIED systems, which violate fairness. Furthermore, participants have noted that the AI evaluation tends to favour extroverted students, whose performances are more visibly detectable, whereas introverted stu-

dents may be unjustly perceived as disengaged in learning. In this scenario, introverts may feel compelled to alter their natural learning behaviours unfairly with more effort, just to receive equitable treatment.

Additionally, fairness in education requires transparent expectations and standards. However, when in the learning environment the rules of engagement and assessment criteria are not clearly articulated or understood in what they perceive as a learning panopticon, students are not fairly treated. In addition, an AI mechanism that monitors and evaluates students by quantifying students' behaviours and facial expressions to score their engagement (Ahmad et al., 2022) has been recognised as posing transparency (Robertson et al., 2021) and bias problems (Sha et al., 2022, Jiang and Pardos, 2021b). For instance, an algorithm can work in ways that it was not originally designed for, e.g., absorbing learning biases from data and then exaggerating it further during the modelling process, which can result in certain ethnicities and demographics being disadvantaged (Baker and Hawn, 2022).

Specifically, AI's interpretation and grading on students' in-class engagement are based on calculations with quantitative indicators, where students can be categorised as 'disengaged' or 'struggling' (Nawaz et al., 2018, Aslan et al., 2019) and 'low-performing' or 'at-risk' (Namoun and Alshantiri, 2020). This evaluation result can then influence educational strategies (Gasevic et al., 2017, Alzoubi et al., 2021) and performance or drop-out prediction (Gašević et al., 2016, Xing et al., 2016) and can be supplied to human teachers (Aslan et al., 2019). When these data-driven interpretations of students' performance are ineffective, the evaluation becomes unfair, as do the subsequent consequences or actions derived from these assessments.

Respect

AIED might affect students' respect for their learning. Students can show respect by being interactive, cooperative, and participatory in learning with others, as well as taking an active role in their own education journey (International Center for Academic Integrity (ICAI), 2021). Meanwhile, both students and the learning environment consider, honour, and value diverse opinions and ideas. The surveillance culture fostered by AI monitoring and evaluation, however, might prioritise conformity and pandering performativity over creativity and experimentation, discouraging students' free expression of viewpoints. In an environment where students strive to avoid mistakes, they are no longer encouraged to challenge, test, and refine their ideas. Thus, students might be discouraged from being respectful of their own learning, and accepting diverse opinions that divergent from the standard "model answers".

Responsibility

Students need to be responsible for their own learning. However, monitoring and evaluation AIED in the classroom might undermine the development of such responsibility. There also should be a shared responsibility between the broader education system and students on learning outcomes. However, the introduction of AI

monitoring might shift this balance, placing undue emphasis on such monitoring as a means of tracking students' learning progress where students perceive that they are forced to learn. This disrupted learning autonomy leads to disengagement from the personal responsibility that students take on in their education—a responsibility that ought to be nourished rather than impeded. As one student expressed it: “I feel that I am not learning for myself but to completing tasks to satisfy the evaluation system” (P8).

Gallant (2017) pointed out that the whole educational system should be involved in encouraging academic integrity, not just the students. Especially, the organisational culture in the classroom can influence students learning perceptions and behaviours. Eaton (2024) emphasised the importance of ethical leadership, institutional support, and the active involvement of all stakeholders of the educational community in fostering an environment of integrity. Such an environment should encourage a more balanced approach to responsibility, ensuring that AI tools enhance rather than detract from the educational experience, and supporting a culture with emerging technology that has educational values at its core. Specifically, students should be encouraged to control their own learning, including their learning progress and approaches (Han et al., 2023), thereby taking responsibility for their education in ways that align with the fundamental values of academic integrity.

Summary

In this section, we discussed how adopting AI observation, monitoring and evaluation in the classroom can undermine the fundamental values of academic integrity: *Honesty* – Monitoring-oriented AIED could compromise student honesty as students are encouraged to perform and cheat the evaluation mechanism. This dishonest behaviour in daily learning could eventually permeate their approach to producing scholarly work, as students might become accustomed to dishonesty as a learning strategy; *Trust* – The use of monitoring AI might erode students' trust in their learning environment, where they may feel inhibited from freely expressing their thoughts for fear of getting punished for making mistakes. Eroding student trust could stifle open academic discourse; *Fairness* – AIED in classroom settings potentially undermines fairness for students. The quantitative measures and algorithms for evaluating students can raise concerns about transparency, bias, and effectiveness of these systems. This can lead to unfairness in how students are assessed and the interventions they receive, and again create a culture in which students are less inclined to act fairly in their studies and assessments; *Respect* – Under the gaze of machine monitoring, students may be discouraged from being interactive, cooperative, and participatory with their peers. Thus, such AIED may diminish students' respect for learning and producing academic work; *Responsibility* – AIED might undermine students' sense of responsibility for their own learning and work, insofar as they feel forced into learning in a way that disrupts their learning autonomy. In sum, while some monitoring AI might bring benefits for student learning, it also has the potential to damage not only their learning but their development of the values of academic integrity.

5. Conclusion

In this chapter, we showed how AIEd which observes, monitors, and assesses students in the classroom can potentially generate student behaviour change, which disrupts learning autonomy and undermines academic integrity. AIEd might cause students to abandon or fail to develop values inherent in that concept. These unintended consequences might not be anticipated by the designer and may even go against their intentions. We should think about this before any negative consequences really happen in education. Such consequences may be hard to remediate, and they will result in potentially avoidable harm to students, their learning, and their values. Our findings underscore the need for careful consideration of how AI technologies are integrated into educational settings to ensure they align with and support the fundamental values of academic integrity.

We raise the following recommendations based on our findings: It is crucial to consider ethical implications and both intended and unintended consequences before deploying AI technologies to prevent ethical harms that may be challenging to rectify later. Additionally, the design and development of technology in educational settings should be guided by the fundamental pedagogical values that the technology aims to support.

In this chapter, we extended the prevailing, broad definition of academic integrity, which traditionally focuses narrowly on student work. We widened the scope of academic integrity to include student learning behaviour that is subject to AI monitoring and evaluation, while focusing on the fundamental values inherent in the concept of academic integrity. This expanded definition underscores the importance of integrating values of honesty, trust, fairness, responsibility, and respect into every aspect of the educational journey, particularly in the era of AI.

References

- AHMAD, A., SCHNEIDER, J., GRIFFITHS, D., BIEDERMANN, D., SCHIFFNER, D., GRELLER, W. & DRACHSLER, H. 2022. Connecting the dots—A literature review on learning analytics indicators from a learning design perspective. *Journal of Computer Assisted Learning*.
- AHUJA, K., SHAH, D., PAREDDY, S., XHAKAJ, F., OGAN, A., AGARWAL, Y. & HARRISON, C. 2021. Classroom Digital Twins with Instrumentation-Free Gaze Tracking. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama, Japan: Association for Computing Machinery.
- ALTUN, D., KRAUSS, C., STREICHER, A., MUELLER, C., ATORF, D., RERHAYE, L. & KUNDE, D. Lessons Learned from Creating, Implementing and Evaluating Assisted E-Learning Incorporating Adaptivity, Recommendations and Learning Analytics. *International Conference on Human-Computer Interaction*, 2022. Springer, 257-270.

- ALZOUBI, D., KELLEY, J., BARAN, E., GILBERT, S. B., ILGU, A. K. & JIANG, S. 2021. TeachActive Feedback Dashboard: Using Automated Classroom Analytics to Visualize Pedagogical Strategies at a Glance. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama, Japan: Association for Computing Machinery.
- ASLAN, S., ALYUZ, N., TANRIOVER, C., METE, S. E., OKUR, E., D'MELLO, S. K. & ARSLAN ESME, A. Investigating the impact of a real-time, multimodal student engagement analytics technology in authentic classrooms. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019. 1-12.
- AUSTRALIAN GOVERNMENT TEQSA. 2022. Understanding Academic Integrity. Available: <https://www.teqsa.gov.au/students/understanding-academic-integrity/what-academic-integrity#:~:text=Academic%20integrity%20is%3A,%20or%20'academic%20dishonesty>.
- BAKER, R. S. & HAWN, A. 2022. Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32, 1052-1092.
- BLYTHE, M., ENCINAS, E., KAYE, J., AVERY, M. L., MCCABE, R. & ANDERSEN, K. Imaginary design workbooks: Constructive criticism and practical provocation. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018. 1-12.
- BRAUN, V. & CLARKE, V. 2006. Using thematic analysis in psychology. *Qualitative research in psychology*, 3, 77-101.
- BRAUN, V., CLARKE, V., HAYFIELD, N., FRITH, H., MALSON, H., MOLLER, N. & SHAH-BECKLEY, I. 2019. Qualitative story completion: Possibilities and potential pitfalls. *Qualitative Research in Psychology*, 16, 136-155.
- CENTER FOR ACADEMIC INTEGRITY. 1999. *The fundamental values of academic integrity: Honesty, trust, respect, fairness, responsibility*. [Online]. Available: <https://files.eric.ed.gov/fulltext/ED452778.pdf> [Accessed].
- CLARKE, V., BRAUN, V., FRITH, H. & MOLLER, N. 2019. Editorial introduction to the special issue: using story completion methods in qualitative research. Taylor & Francis.
- CORLETT, J. A. 2005. The good professor. *Journal of Academic Ethics*, 3, 27-54.
- COTTON, D. R., COTTON, P. A. & SHIPWAY, J. R. 2023. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1-12.
- CURTIS, G. 2021. Evolving an understanding of academic integrity: Keynote address. . *European Conference on Academic Integrity and Plagiarism (ECAIP)*.
- DUNNE, A. & RABY, F. 2013. *Speculative everything: design, fiction, and social dreaming*, MIT press.

- EATON, S. E. 2024. Comprehensive Academic Integrity (CAI): An Ethical Framework for Educational Contexts. *Second Handbook of Academic Integrity*. Springer.
- EZ-ZAOUIA, M., TABARD, A. & LAVOUÉ, E. 2020. EMODASH: A dashboard supporting retrospective awareness of emotions in online learning. *International Journal of Human-Computer Studies*, 139, 102411.
- GALLANT, T. B. 2017. Academic Integrity as a Teaching & Learning Issue: From Theory to Practice. *Theory Into Practice*, 56, 88-94.
- GAŠEVIĆ, D., DAWSON, S., ROGERS, T. & GASEVIC, D. 2016. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28, 68-84.
- GASEVIC, D., JOVANOVIĆ, J., PARDO, A. & DAWSON, S. 2017. Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics*, 4, 113–128-113–128.
- HAKIMI, L., EYNON, R. & MURPHY, V. A. 2021. The Ethics of Using Digital Trace Data in Education: A Thematic Review of the Research Landscape. *Review of Educational Research*, 00346543211020116.
- HAN, B., BUCHANAN, G. & MCKAY, D. Learning in the Panopticon: Examining the Potential Impacts of AI Monitoring on Students. Proceedings of the 34th Australian Conference on Human-Computer Interaction, 2022. 9-21.
- HAN, B., NAWAZ, S., BUCHANAN, G. & MCKAY, D. Ethical and Pedagogical Impacts of AI in Education. 2023 Cham. Springer Nature Switzerland, 667-673.
- HOLSTEIN, K., MCLAREN, B. M. & ALEVEN, V. SPACLE: Investigating learning across virtual and physical spaces using spatial replays. Proceedings of the Seventh International Learning Analytics & Knowledge Conference, 2017. 358-367.
- HOLSTEIN, K., MCLAREN, B. M. & ALEVEN, V. Student Learning Benefits of a Mixed-Reality Teacher Awareness Tool in AI-Enhanced Classrooms. 2018 Cham. Springer International Publishing, 154-168.
- HUTT, S., KRASICH, K., R. BROCKMOLE, J. & K. D'MELLO, S. Breaking out of the lab: Mitigating mind wandering with gaze-based attention-aware technology in classrooms. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021. 1-14.
- INTERNATIONAL CENTER FOR ACADEMIC INTEGRITY (ICAI) 2014. The fundamental values of academic integrity (2nd ed.). In: FISHMAN, T. (ed.).
- INTERNATIONAL CENTER FOR ACADEMIC INTEGRITY (ICAI) 2021. The Fundamental Values of Academic Integrity (3rd ed.).
- ISRAEL, M. 2015. Research ethics and integrity for social scientists: Beyond regulatory compliance (2nd ed.). SAGE Publications Ltd.
- JIANG, W. & PARDOS, Z. A. 2021a. Towards Equity and Algorithmic Fairness in Student Grade Prediction. *arXiv preprint arXiv:2105.06604*.

- JIANG, W. & PARDOS, Z. A. Towards equity and algorithmic fairness in student grade prediction. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 2021b. 608-617.
- KITZINGER, C. & POWELL, D. 1995. Engendering infidelity: Essentialist and social constructionist readings of a story completion task. *Feminism & Psychology*, 5, 345-372.
- KORIR, M. M. 2022. *An Investigation of Privacy and Utility Tensions in Learning Analytics*. The Open University.
- LAWSON, S., KIRMAN, B., LINEHAN, C., FELTWELL, T. & HOPKINS, L. Problematising upstream technology through speculative design: the case of quantified cats and dogs. *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015. 2663-2672.
- LINSON, A., XU, Y., ENGLISH, A. R. & FISHER, R. B. Identifying Student Struggle by Analyzing Facial Movement During Asynchronous Video Lecture Viewing: Towards an Automated Tool to Support Instructors. *International Conference on Artificial Intelligence in Education*, 2022. Springer, 53-65.
- MACFARLANE, B., ZHANG, J. & PUN, A. 2014. Academic integrity: a review of the literature. *Studies in Higher Education*, 39, 339-358.
- NAM, S. & SAMSON, P. 2019. *Integrating Students' Behavioral Signals and Academic Profiles in Early Warning System*. Springer International Publishing.
- NAMOUN, A. & ALSHANQITI, A. 2020. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, 11.
- NAWAZ, S., KENNEDY, G., BAILEY, J., MEAD, C. & HORODYSKYJ, L. Struggle town? Developing profiles of student confusion in simulation-based learning environments. *35th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education, ASCILITE*, 2018. 224-233.
- NIXON, J. 2004. Education for the good society: The integrity of academic practice. *London Review of Education*, 2.
- NUSS, E. M. 1984. Academic Integrity: Comparing Faculty and Student Attitudes. *Improving College and University Teaching*, 32, 140-144.
- QUEIROGA, E. M., BATISTA MACHADO, M. F., PARAGARINO, V. R., PRIMO, T. T. & CECHINEL, C. 2022. Early Prediction of At-Risk Students in Secondary Education: A Countrywide K-12 Learning Analytics Initiative in Uruguay. *Information*, 13, 401.
- ROBERTSON, S., NGUYEN, T. & SALEHI, N. Modeling assumptions clash with the real world: Transparency, equity, and community challenges for student assignment algorithms. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021. 1-14.
- RUAN, X., PALANSURIYA, C. & CONSTANTIN, A. Affective Dynamic Based Technique for Facial Emotion Recognition (FER) to Support Intelligent

- Tutors in Education. International Conference on Artificial Intelligence in Education, 2023. Springer, 774-779.
- SHA, L., RAKOVIC, M., DAS, A., GASEVIC, D. & CHEN, G. 2022. Leveraging Class Balancing Techniques to Alleviate Algorithmic Bias for Predictive Tasks in Education. *IEEE Transactions on Learning Technologies*, 1-12.
- SMOLAKS, M. 2019. *Chinese university deploys facial recognition tech to monitor student behavior* [Online]. AI Business. Available: <https://aibusiness.com/verticals/chinese-university-deploys-facial-recognition-tech-to-monitor-student-behavior> [Accessed 13th April 2023].
- TAUGINIENĖ, L., GAIŽAUSKAITĖ, I., GLENDINNING, I., KRAVJAR, J., OJSTERSEK, M., RIBEIRO, L., ODINECA, T., MARINO, F., COSENTINO, M., SIVASUBRAMANIAM, S. & FOLTÝNEK, T. 2018. Glossary for Academic Integrity. European Network for Academic Integrity (ENAI) Report [online].
- TROIANO, G. M., TIAB, J. & LIM, Y.-K. SCI-FI: shape-changing interfaces, future interactions. Proceedings of the 9th Nordic Conference on Human-Computer Interaction, 2016. 1-10.
- WONG, R. Y. & KHOVANSKAYA, V. 2018. Speculative design in HCI: from corporate imaginations to critical orientations. *New Directions in Third Wave Human-Computer Interaction: Volume 2-Methodologies*. Springer.
- WOOD, M., WOOD, G. & BALAAM, M. "They're Just Tixel Pits, Man" Disputing the'Reality'of Virtual Reality Pornography through the Story Completion Method. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017. 5439-5451.
- XING, W., CHEN, X., STEIN, J. & MARCINKOWSKI, M. 2016. Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in human behavior*, 58, 119-129.
- YANG, C. C. Y. & OGATA, H. 2022. Personalized learning analytics intervention approach for enhancing student learning achievement and behavioral engagement in blended learning. *Education and Information Technologies*.