



Minerva Access is the Institutional Repository of The University of Melbourne

**Author/s:**

D'Alfonso, S;Coghlan, S;Schmidt, S;Mangelsdorf, S

**Title:**

Ethical Dimensions of Digital Phenotyping Within the Context of Mental Healthcare

**Date:**

2025-03-01

**Citation:**

D'Alfonso, S., Coghlan, S., Schmidt, S. & Mangelsdorf, S. (2025). Ethical Dimensions of Digital Phenotyping Within the Context of Mental Healthcare. *Journal of Technology in Behavioral Science*, 10 (1), pp.132-147. <https://doi.org/10.1007/s41347-024-00423-9>.

**Persistent Link:**

<https://hdl.handle.net/11343/352588>

**License:**

[cc-by](#)



# Ethical Dimensions of Digital Phenotyping Within the Context of Mental Healthcare

Simon D'Alfonso<sup>1</sup> · Simon Coghlan<sup>1</sup> · Simone Schmidt<sup>1</sup> · Shaminka Mangelsdorf<sup>2,3</sup>

Received: 18 December 2023 / Revised: 6 June 2024 / Accepted: 10 June 2024  
© The Author(s) 2024

## Abstract

Digital phenotyping (DP) refers to the emerging field within digital (mental) health that involves the collection of data from individual's digital devices (smartphones, wearable, Internet usage, etc.) to monitor and analyse their behaviours, activities and health-related patterns. That such 'digital footprint' data can be mined for behavioural insights is an intriguing idea, which has motivated an increasing amount of research activity, particularly in the field of digital mental health. Whilst of potentially revolutionary utility in (mental) healthcare, the idea of DP also raises a set of rich sociotechnical, ethical and philosophical considerations, and a cottage industry of sociotechnical and ethical critiques of DP has emerged, particularly within the humanities. Within this scene of early empirical investigation in the health/behavioural sciences on the one hand and general conceptual critiques from the humanities on the other, in this paper we instead explore a case for the potential utility of DP in circumstances of clinical mental health practice and examine its ethical dimensions in this context. After providing an explicatory framework for DP and making a case for it within this framework, we subsequently examine the ethical pros and cons of three topics pertaining to DP in clinical practice, namely (1) issues in psychometric assessment/testing, (2) the role and responsibilities of mental health practitioners with regard to DP technology, and (3) the value DP can afford clients in terms of self-awareness/empowerment and strengthening the therapeutic alliance with their clinician.

**Keywords** Digital phenotyping · Clinical psychology · Mental health · Psychiatry · Therapeutic alliance

## Introduction

Our increasing usage of smartphones, wearables, and the Internet (particularly social media/networking) has increased our 'digital footprint' or 'data exhaust'. An individual's digital footprint consists of the data records created from their interactions with such digital technologies. These interactions sum up to form a unique digital history of the individual, which can be mined to infer information about them. This possibility has been utilised, somewhat notoriously, by social media companies who use our online interactions to both automatically customise our experience of

their systems and deliver targeted advertising. The fact that an individual's digital footprint could be used to infer their behaviours, preferences, and mental states also opens the possibility of mental health practitioners (e.g. psychologists and psychiatrists) using this technology to gain insights into the mental health of their clients/patients. These insights could potentially have clinical value by enabling anticipation of mental ill-health and informing treatment.

Lying at the intersection of the computing, data and behavioural sciences, this process of learning about an individual's psychology from their digital footprint has been termed 'digital phenotyping' (DP) (Torous et al., 2016), although the more accessible term 'personal sensing' (Mohr et al., 2017) has also been suggested. 'Phenotype' is a scientific term used to denote an organism's (including humans) naturally observable physical and behavioural characteristics. Thus, 'digital phenotyping' is the process of measuring or identifying certain (behavioural) characteristics based on an individual's digital footprint.

Whilst the field of DP is incipient and tentative, results from early exploratory research suggest some associations

---

✉ Simon D'Alfonso  
dalfonso@unimelb.edu.au

<sup>1</sup> School of Computing and Information Systems, University of Melbourne, Level 4, Melbourne Connect (Building 290), Parkville, VIC 3010, Australia

<sup>2</sup> Centre for Youth Mental Health, University of Melbourne, Parkville, VIC 3010, Australia

<sup>3</sup> Orygen, Parkville, VIC 3052, Australia

between digital footprint data and mental health or behaviour more generally. For example, one of the findings to emerge from early DP studies in the 2010s were connections between movement or physical activity (as tracked by geolocation and accelerometer sensors) and mental ill-health, particularly depression, as measured by scales such as the Patient Health Questionnaire (PHQ-9) (Saeb et al., 2016). Subsequent work has added to such findings (Masud et al., 2020; Shin & Bae, 2023).

Just as the language we use provides a window into our minds, text from sources such as (online) therapy/counseling transcripts, keyboard character input and peoples' social media posts can potentially be analysed with natural language processing (NLP) to assess mental health (Garg, 2023; Zhang et al., 2022). Properties characteristic of language disturbance, such as impoverished vocabulary, semantic incoherence, and reduced syntactic complexity, could be indicators of mental illness. For example, one of the earlier research projects on NLP and mental health used measures of coherence and complexity in therapy transcripts to predict the onset of psychosis (Corcoran et al., 2018).

Apart from language analysis, there has also been work on using smartphone tactile screen interactions and keystroke dynamics to gain insights into neurocognitive functioning (Nguyen et al., 2023) and mental health. One study comparing two groups, one with depressive tendencies and one without, found that the depressive group had longer periods between pressing and releasing a key, indicating a slower motor reaction time or psychomotor retardation, which is a feature of depression (Mastoras et al., 2019). Another study showed that both average delays between keystrokes and auto-correction rates (misspellings) correlated positively with the Hamilton Depression Rating Scale (Zulueta et al., 2018).

These few examples give some indication of the insights that DP might yield. Indeed, research into DP is rapidly increasing along with excitement about its potential for clinical mental and other healthcare (Perez-Pozuelo et al., 2021), social research (Liang et al., 2019), population health screening (Huckvale et al., 2019), and personal monitoring (Sheikh et al., 2021). But what exactly is the basis for these findings, how could/would the data from digital footprints be clinically employed, and could DP be justified in practice? Although there has been significant scientific research into DP and critical analysis (including techno-social criticism) from the social sciences (Birk & Samuel, 2020, 2022) and ethics/philosophy (Loi, 2019; Mulvenna et al., 2021), there has been only a relatively small set of papers to emerge that offer a focused ethical analysis of DP with respect to its implementation in clinical mental healthcare.

One recent example can be found in Shen et al. (2024), in which a detailed ethical framework for the return of individual digital phenotyping research results to clients in mental

healthcare is developed. The framework is grounded in the core bioethics principles identified in the Belmont Report: respect for persons, beneficence and justice (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). In another paper (Martinez-Martin et al., 2021), the authors employ a Delphi study to generate a collection of consensus statements regarding the ethical application of DP to mental health. Falling under the general issues of security/privacy, transparency, consent, accountability, and fairness, the statements relate to various considerations covered and points made in the current paper. A third example (Oudin et al., 2023) discusses some ethical issues such as dehumanisation and risks to the therapeutic process, before providing an overview of recommendations for implementing DP in mental health, including the potential empowerment of patients and healthcare professionals, topics we will look at more in the final section of this paper.

Adding to this incipient set of papers that focus on ethical aspects of DP in mental healthcare, in this paper we make a case for the potential utility of DP and examine the ethical pros and cons of three topics pertaining to DP in clinical mental health practice, namely (1) issues in psychometric assessment/testing, (2) the role and responsibilities of mental health practitioners with regard to DP technology, and (3) the value DP can afford clients in terms of self-awareness/empowerment and strengthening the therapeutic alliance with their clinician. Whilst critical analysis and a moderation of hype are imperative, our cautious case for considering DP is motivated in part by what is arguably an excessive techno-scepticism towards it. Although it is possible that DP will ultimately turn out to be clinically inadequate/inapplicable or will raise insurmountable ethical issues, an excessive and pre-emptive pessimism is also problematic since it may undermine a legitimate interest in DP from researchers, clients, and clinicians. We favour ethical examination based on considerations of example clinical scenarios, methodical issues in psychology, professional clinician issues, attitudes of and benefits to clients, and the therapeutic alliance between clinician and client.

## An Explication of Digital Phenotyping

In order to investigate and scrutinise the use and ethics of DP in clinical mental health practice, we first explicate how digital footprint data could be used. Our primary distinction is between (1) the manual use of DP information to inform clinician judgements regarding client behaviour and treatment decision-making, sometimes in concert with their clients, and (2) the use of DP data to feed AI-driven psychometric determinations, which are automatically generated and provided to clinicians. For brevity, we shall call them 'manual DP' and 'AI-driven DP' respectively.

## Manual DP: Using Information to Inform Clinician Judgements and Decision-Making

Suppose that information extracted from a client's digital footprint was made available to their clinician via a dashboard. The clinician could then access this information prior to or during a session, directly interpreting this information and incorporating it into their practice, and possibly informing clinical judgements and decision-making. For example, features extracted from communication data, such as call numbers, significant communication decreases, or contact variety, could be related to social activity and could be presented to and directly interpreted by clinicians, potentially serving as one input in clinical judgement or as a discussion point with clients. Basic language information, such as frequently used keywords or sentiment analyses extracted from a client's social media postings, could be used by clinicians, as well as the apps that a client is using, their type and usage duration.

Such information could also be used to infer behaviours of clinical interest. For example, if a client has a baseline phone usage activity that indicates consistent sleep by midnight, and the client commences abnormal phone usage in the middle of the night, one might infer problematic sleeping issues or insomnia. This is an abductive inference, an inference to the best explanation from multiple possible causes. Now, this need not be an inference with 100% certainty, and it is possible that there are other, unproblematic causes. For example, unbeknownst to the clinician, the client could simply be a soccer fan watching a soccer tournament in another time zone that week. Still, if the clinician knows that the client is not a soccer fan, they could rule out that possibility and other improbable possibilities and use the inference of sleep disturbance as a discussion point in a subsequent session with the client.

Furthermore, data from multiple sensors or sources could be combined to increase the chances of an abductive inference being correct. A client uncharacteristically failing to charge their phone throughout the week could indicate low mood and lack of motivation. However, if coupled with calendar and geolocation information indicating a busy week and lack of opportunity to charge, the abductive inference to low mood and lack of motivation is disconfirmed.

These are but a couple of examples of many for a type of scenario in which digital footprint information could be indirectly used in a transitive fashion: digital footprint information  $x$  abductively implies some behaviour  $y$ , behaviour  $y$  is linked to some mental health condition  $z$ , therefore  $x$  implies  $z$ . Such inferences could possibly even be used in determining the presence of criteria for diagnosis frameworks such as the Diagnostic and Statistical Manual of Mental Disorders, where behaviour is a key part of diagnosis and treatment. For example, geolocation can be

used to gauge proximity to bars in the case of managing problematic alcohol use (Carreiro et al., 2018). Inferences may also be made about transdiagnostic mechanisms like avoidance in the case of anxiety or behavioural activation in the case of depression (Bennetts, 2021). Ultimately, these behavioural insights may have broad clinical utility given that the assessment and altering of behaviour are key components of many if not all mental health interventions. These insights may not only be relevant at the assessment and intervention stages but may form an important part of discharge planning and relapse prevention if, for example, behaviours likely to lead to relapse could trigger a reminder to use appropriate skills.

Such an approach of performing abductive inferences based on digital phenotyping information indicates that digital phenotyping is not a fully automated, silver bullet solution. Rather, whether it is realistic and how feasible it is depends on the abductive inference affordances available and the confirmation efforts of vigilant practitioners working together with their clients. In this way, digital phenotyping is a tool to inform responsible practitioner decision-making.

## AI-Driven DP: Using Automated Psychometric Evaluations and Diagnoses

Beyond using digital footprint data for direct clinician interpretations or for inferences of client behaviour, there is the idea that digital footprint data could possibly be used as input into artificial intelligence (AI) systems (mainly machine learning) that could in practice detect or predict the state of one's mental health.<sup>1</sup> It is important to clarify that whilst AI could play some part in the manual type of DP described in 'Manual DP: Using Information to Inform Clinician Judgements and Decision-Making' section, such as when AI methods are used to convert raw sensor data to useful non-diagnostic information, this differs from AI-driven DP as it is defined here. For example, a clustering algorithm applied to an individual's raw geolocation data could extract location patterns. By contrast, AI-driven DP involves algorithms that make psychometric or diagnostic inferences. The main type of machine learning method, supervised machine learning, requires a training set consisting of inputs coupled with quantitative outputs or labelled categories, which form the basis of what the machine learning model ends up

<sup>1</sup> 'Prediction' in machine learning denotes the process of calculating a quantitative dependent variable output when certain values are input into a model. 'Classification' denotes the process of determining to which category an input belongs given its features. With digital phenotyping, we distinguish between the 'detection' of actual conditions/states and 'prediction' of future ones. Throughout this paper, we will use 'prediction' as the umbrella term for all these terms, unless distinction is required.

being constructed to predict or classify. When it comes to DP machine learning, what quantities or well-defined categories are to be used as the output values (i.e. dependent variables) being predicted?

Quantitative DP research has mainly consisted of exploratory studies that attempt to find correlations or construct machine learning models that contain predictive associations between sensing input features and psychometric test scores (see Barnett et al., 2018; Jacobson et al., 2020; Saeb et al., 2016 for some study examples and Cornet & Holden, 2018; Liang et al., 2019 for some reviews). At points during the data collection phase of a study, participants are given self-report measures or psychometric testing more generally for the conditions/characteristics being investigated (e.g. the Generalised Anxiety Disorder-7 (GAD-7) scale for generalised anxiety). These scores can then be used as the dependent quantitative variables of interest in labelled machine learning methods.

Condition categories could also be used as labels for machine learning classification. For example, the Kessler Psychological Distress Scale (K10) has four categories of distress: (1) likely to be well, (2) likely to have a mild mental disorder, (3) likely to have a moderate mental disorder and (4) likely to have a severe mental disorder. Another possibility is to employ binary classification methods such as logistic regression in cases where the prediction output is a yes/no. For example, a dataset of patients including various data points about them as well as whether they have relapsed after some treatment could be used to train a classifier that predicts whether relapse will occur. Whilst interesting associations have been found in such digital phenotyping research, and machine learning models with reasonable evaluation results may appear, it is worth emphasising that a point of broadly applicable and highly accurate/reliable machine learning models for detecting mental health problems has not been reached. Throughout this paper, such models and implementable AI-driven DP are largely hypothesised for the purposes of conceptual exploration.

Psychometric testing has been traditionally conducted in a pen and paper fashion or perhaps now more commonly done via computer/Internet form (Naglieri et al., 2004). Whatever input method is used, such tests are ultimately about the client providing direct answers to questions about their own behaviours or experiences (sometimes the clinician might ask the client the questions and fill the responses in themselves). Although the prospects for employing machine learning to predict psychometric test results depend on the types of tests and the intended use, the possibility of substituting a test conducted by a client/clinician with an algorithmic prediction of a test score or category based on the client's digital footprint is no doubt an interesting one that raises rich considerations, which we will explore further shortly. It is also worth noting that

self-report measures have their own issues, such as social desirability and recall bias (Althubaiti, 2016), and that aside from model construction involving the prediction of self-report scores, DP offers the possibility of providing more objective behavioural data. Whilst information concerning thoughts and feelings will continue to be obtained via subjective self-report, questions requiring individuals to recall their behaviours or experiences could be supplemented or replaced with DP data.

The prospects of developing an adequate machine learning system that uses a certain set of digital footprint input features to predict some psychometric test with sufficient accuracy/reliability across a population remain to be seen.<sup>2</sup> Were a sufficient system developed, it could have clinical utility. For example, rather than having to seek out a test, whether it be online or provided at a clinic, an individual's digital footprint could be automatically crunched periodically to determine some test result. If warranted, the individual could be notified and advised to seek pre-emptive clinical care. Or, if they were currently in clinical care, reports could be automatically sent to their clinician, without the need to conduct the psychometric test. Even if score predictions are not always exactly what the client would have gotten had they actually taken the test, a consistently sufficient approximation would still be informative. Furthermore, instances in which categorical tests are employed would circumvent the aim of a precise numerical score. In general, AI could enable likely classification of the severity of impairment, which could inform appropriate prevention or intervention.

Beyond these considerations regarding the clinical utility of imperfect but nonetheless highly accurate/reliable machine learning models, we can also entertain the existence of a perfect hypothetical machine learning model that is always correct for a given quantitative test that has an accuracy of 100%, in order to elicit further philosophical and ethical considerations. What are the implications of having a model that for every sufficient user/client dataset instance, outputs a score  $x$  such that if the user were to have taken the test manually, they would have received a score of  $x$ ? These are two radically different ways to compute  $x$ , and the existence of such an idealised model would mean that there is some mapping between the set of an individual's

<sup>2</sup> What constitutes 'sufficient accuracy/reliability' is context-dependent and does not admit of a precise answer. Any implementable model should be such that there are safeguards in place to make up for its imperfections or minor chances of inaccurate predictions. Furthermore, it is also dependent on which evaluation metrics are prioritised. For example, if false negatives are more costly than false positives, then the system should be optimised for recall (sensitivity) maximisation.

experiences and responses that lead to their score and an expression of this in their digital footprint. As will be elaborated on further, the DP method would also have the advantage of providing more information of potential utility.

In this section, we established a distinction between what we termed manual DP and AI-driven DP. We here note that with regard to the latter notion of DP, it remains within the realm of research and to our knowledge there are no clinical implementations of digital phenotyping systems based on AI models providing fully automated mental health diagnoses or information for diagnostic determinations. Given the accumulation of data and research required, as well as the technical challenges of developing predictive machine learning models of human psychological states based on digital footprint information, the implementation of viable and effective AI-driven DP systems remains to be seen. With regard to the former notion of DP, a few examples are emerging of (research) clinics exploring the incorporation of DP into their practice via the implementation of dashboard platforms that provide clinicians with (visualised) digital footprint information from their clients. For example, the division of digital psychiatry (Digital Clinic) at Beth Israel Deaconess Medical Center leverages mind-LAMP, a digital platform consisting of a smartphone app that collects sensing data and an accompanying visualisation dashboard. Both clients and clinicians can view various pieces of information derived from this data, helping to guide reflection and inform treatment (Rodriguez-Villa et al., 2020). Furthermore, Monsenso<sup>3</sup> is one example of a commercially available app that gathers some passive sensing data alongside self-reported information, aimed at enhancing the treatment and management of mental health conditions through a combination of self-monitoring, psychoeducation, and clinician support.

The presence of manual DP systems and absence of AI-driven DP systems can basically be seen in terms of the simplicity of manual DP relative to AI-driven DP, in that the latter requires sophisticated AI models which are yet to be realised whereas the former only requires well understood and readily available technologies, namely the means to collect digital footprint information and software dashboard platforms to present this information. Beyond the technical aspects, how clinicians can meaningfully and effectively incorporate manual DP into their practice calls for more research. Furthermore, one challenge (that relates to both forms of DP) concerns the quality of the raw data collected from individuals, which depends on factors such as the digital devices being used by clients and what data clients are prepared to share.

<sup>3</sup> <https://www.monsenso.com/downloads-research-papers/>

## The Existing Critical Landscape and a Case for Digital Phenotyping

### Hype or Hope? Advocacy and Criticism of Digital Phenotyping

Digital phenotyping has generated both advocacy and criticism. Some scholars and mental health practitioners argue that DP could greatly improve or even transform mental healthcare. For example, Thomas Insel claims that smartphones and digital sensing promise more precise and objective insight into mental health, partly ameliorating excessive reliance on subjective reports from clients (Insel, 2018). As the WPA-Lancet Psychiatry Commission on the Future of Psychiatry says, current ‘definitions of mental disorders are based exclusively on subjective signs and patient reported symptoms that are prone to recall error and misinterpretation’ (Bhugra et al., 2017, p. 779). Insel even suggests that ‘the revolution in technology and information science will prove more consequential for global mental health’ than have neuroscience and genomics (Insel, 2018, p. 276). Another advocate, John Torous, believes DP has tremendous potential and may become ‘increasingly necessary for clinicians seeking to engage meaningfully with their patients’ (Torous et al., 2019, p. 196).

DP has also been criticised. Some caution against an embrace of DP that stems from uncritically ‘jumping on the artificial intelligence bandwagon’ (Tekin, 2021, p. 3). It is certainly true that whilst DP has been subject to exploratory studies that show promise, it has not yet been definitively validated (Tekin, 2021). We want to be clear that DP should not be used by practitioners until it has received sufficient empirical support, consisting of years of accumulated evidence and validating clinical studies. Premature use would risk providing limited benefit and even causing harm to clients and would thus violate core duties of mental health practitioners.

DP’s advocates do indeed recognise the need for many rigorous studies prior to DP’s implementation in practice (Huckvale et al., 2019). However, some critics raise concerns that irrespective of its exploration and potential results, DP is simply too risky. For example, Birk and Samuel argue that DP faces problems of data ownership, privacy, algorithmic bias, and sustainability (Birk & Samuel, 2022). Scepticism about DP’s clinical potential often derives from concerns about its conceptual, epistemic, and empirical grounds. For example, some are concerned that DP can crucially overlook first-person perspectives, potentially leading to the exacerbation of epistemic injustice in cases where an individual’s first-person testimony is at odds with DP results (McCadden et al., 2023; Slack & Barclay, 2023). Another potential consequence of overlooking first-person

perspectives concerns the risk of false-positive diagnoses (Birk & Samuel, 2020). Yet the opposite criticism has also been made, that DP places too much emphasis on potentially inaccurate self-tracking reports, risking false negatives (Tekin, 2021).

Another epistemic or conceptual criticism is that DP will simply neglect certain individual factors that can completely alter the interpretation of the digital footprint. An example (discussed further below) is when a person's digital footprint reveals inactivity that is not actually associated with depression but rather with, say, socio-economic factors, physical disablement, or simply forgetting to take one's smartphone outside the house (Birk & Samuel, 2020). Such inaccuracies, critics worry, may both reflect and perpetuate biases and injustices against certain groups of people, such as the disabled and socio-economically disadvantaged.

Whilst critical assessments of DP are an important part of exploring this new technology, and the emerging literature is welcome, in this paper we do not intend to specifically address or engage with this part of the critical landscape. We see some concerns as being more valid than others, though we think that none of them suffice to negate DP research and development endeavours. Of those concerns that are valid (e.g. data security, user privacy, bias and explainability issues concerning AI-driven DP), we acknowledge that they may place constraints on DP and their consideration will need to be incorporated into the design and development of DP systems. Motivated by the clinical implementation possibilities of DP, informed by some of our own research into it, and guided by our explication of DP in the '[An Explication of Digital Phenotyping](#)' section, our aim rather is to explore the ethical contours of DP in terms of how it could be carefully and reasonably applied in actual clinical mental healthcare contexts.

### Considering Ethical Dimensions Within the Context of Realistic Clinical Practice

Our exploration of the ethical dimensions of DP is based on its application in clinical mental health and on the two types of usage we have established. As alluded to earlier, this is contrary to much of the ethics discourse on DP, which often deals more broadly with problematic usage in social systems, such as using digital footprint or digitised human data (e.g. facial images) to perform activities such as surveillance, advertising, or prediction of some characteristic or outcome for an individual (e.g. likelihood of (re)committing a crime). Even in literature cases that do deal with DP as a tool in behavioural or mental health fields, examples can tend to be simplistic and lack the nuances of realistic scenarios. For example, Birk and Samuel (2020) refer to a study by Saeb and colleagues (2015), which measured:

the frequency with which a person visited different locations and the distribution of that frequency across locations. The high negative correlation that was found between this feature and the PHQ-9 [patient health questionnaire] scores indicated that people with greater depressive symptom severity visited fewer locations and were more likely to favour some locations over others. Part of this was likely due to the increased amount of time people with depressive symptoms spent at home, measured by the home stay feature (Saeb et al., 2015).

Birk and Samuel then remark that 'the assumption that depression and mobility are related speaks to a tacit assumption about what constitutes a normal life. However, such assumptions can be problematic' (Birk & Samuel, 2020). Marginalised populations, they say for example, may not have access to transport or may remain at home for safety reasons. Or individuals who are differently abled and who have restricted mobility naturally might have minimal movement, because of their limited ability, not because of depression.

But it is crucial to ask, how would associations between geolocation information and depression actually be used? If such information is being used in a clinical setup with the first usage type established in the previous section, then if the clinician observes via a dashboard report minimal geolocation activity from their client over a period of time, they will naturally consider special circumstances that might explain the data. A competent clinician, familiar with their disabled client's inability to walk, will factor this fact and its implications for limited geolocation activity into their assessment. Even for the second usage type involving AI, if a model is going to predict, say PHQ9 scores, with geolocation features as inputs, then there should ultimately be conditionals which ensure that for individuals with certain conditions (e.g. mobility disability), that feature does not apply or that the individual is not eligible to be tested using that model. Alternatively, with models that are based on longitudinal information, which depend on certain reductions in geolocation activity over time, this type of problem would not arise for individuals who are mobility impaired to begin with. The factoring in of such exceptions is probably going to be harder or more complicated in the case of the marginalised population examples given above. Whilst such exceptions could be accommodated by a clinician in manual DP or captured by models in AI-driven DP, they emphasise the challenges of developing nuanced and comprehensive applied digital phenotyping systems. At any rate, a realistic DP decision framework would not simply say that little to no geolocation activity universally equals depression.

Thus, such analyses tend to exclude nuances one would expect to find in a realistic and viable clinical application of digital phenotyping. In the next few sections, we will be

considering DP via (hypothetical) clinical scenarios that are nonetheless shaped by our familiarity with digital phenotyping research and professional experiences with clinical practice. Beyond the ethical concerns and considerations that are relevant to digital health and data technologies more generally<sup>4</sup>, our focus will be on considerations that pertain more specifically to clinical mental health.

## The Applicability and Utility of Digital Phenotyping

The ‘[An Explication of Digital Phenotyping](#)’ section established two broad ways in which DP technology could be used. In the first, simpler and more practicable way, digital footprint information can be incorporated into practice as a tool for blended care and information can be interpreted by clinicians to inform their practice. In the second, more ambitious way, AI systems could deliver diagnostic predictions based on digital footprint information. With this explication at hand, we shall now briefly discuss the case for DP in terms of its potential utility and ability to provide actionable insights and to counter certain criticisms that may be levelled against it.

Some might consider DP, particularly in the sense of automated, AI-driven inferences, a form of 21st century ‘digital phrenology’ or ‘digital physiognomy’ (Aguera y Arcas, 2017; Crawford, 2021), located in the same problematic and controversial camp as predicting an individual’s personality traits or personal orientations using machine learning analyses of facial characteristics. However, such bodily features or characteristics are very different from behavioural indicators. Digital phenotyping is in fact based on two reasonable propositions.

Firstly, there is the well-established proposition that there are associations between certain behaviours and certain mental health conditions. Secondly, there is the reasonable proposition that user contexts and behaviours can be inferred from digital traces. Whilst machine prediction of something as complex and heterogenous as human psychology will never be completely accurate in every case, it is plausible that there could be strong associations between certain sets of information features and certain psychometric outcomes or diagnostic categories. This being said, it is obviously important to consider the accuracy limitations of such predictive algorithms and their applicability.

Medical testing takes many forms and there is no such thing as a 100% perfect test. Many common medical tests, from mammograms to ECGs to thyroid function tests, can give incorrect results, and this can cause patients’ harm.

DP systems would share that risk. Furthermore, even the best DP systems will not provide diagnostic results with the level of reliability, accuracy, and precision of many physical medical tests. As just mentioned, this is in large part due to the complexities of the subject matter (i.e. human mind and behaviour)—something which traditional approaches to mental health diagnosis are no stranger to.

Nevertheless, such inherent limitations do not necessarily negate the potential utility of predictive DP systems, just as the limitations of many medical tests, including those that are very far from perfect, do not necessarily render them unhelpful or too dangerous to deploy. Whether DP systems are unhelpful or excessively dangerous is, as we say, yet to be seen. Furthermore, DP systems should at most be seen as tools to augment and not supplant clinician practice. This point cannot be overemphasized.

We have established the firm idea of *manual* cases where digital footprint information could be usefully delivered to and employed by a clinician to augment and inform their judgements and decision-making, even when there are questions about predictive accuracy. But beyond such relatively less complex manual cases, how could DP ‘tests’ based on predictive AI systems be employed if they fail to reach required standards that are met by some established testing tools in other fields?

Hypothetical scenarios of perfect prediction systems aside, any realistically implemented predictive DP system is not going to be a perfect oracle delivering entirely automated diagnostic conclusions. Rather, DP will be a part of clinical decision support systems, providing reasonably accurate and reliable outputs that must be received, assessed, and utilised by a human clinician in their diagnostic and treatment processes. Indeed, as with other areas, the perils of automation bias<sup>5</sup>, excessive dependence on AI and AI paternalism (Luxton, 2022) need to be avoided or managed. Furthermore, incorporation of client testimony and ‘a commitment to epistemic humility can help promote judicious clinical decision-making at the interface of big data and AI in psychiatry’ (McCradden et al., 2023).

Given a setup in which DP calculations such as behaviour determinations and psychometric predictions are delivered to clinicians, the employability/applicability of these outputs are particularly reliant upon two features of the DP models: *interpretability/explainability* and *accuracy*. Whilst often used interchangeably, technically interpretability and explainability are distinct concepts (Linardatos et al., 2021). Machine learning *interpretability* is the degree to which a

<sup>4</sup> As mentioned, these include considerations such as data security, user privacy, biased/unrepresentative data, corporate misappropriation and increased surveillance.

<sup>5</sup> Automation bias is the tendency for people to uncritically trust and follow the recommendations or decisions made by automated systems, even when human intuition, judgement or common sense suggests otherwise.

human can understand the basis of a machine learning model's predictions by inspecting the model. Examples of interpretable models include linear regression and decision trees. *Explainability* refers to the ability to take a machine learning model and explain its outputs in human terms using certain techniques. The results of complex machine learning models that are not interpretable, such as large neural networks, may nonetheless be amenable to some explanation. At any rate, it is fair to generally say that if a machine learning model delivers a determination, then the more understandable that determination is for a clinician the better. One pertinent phenomenon in this regard is the interpretability-accuracy trade-off: generally, the less complicated a model the more interpretable it is, and on the (not always correct) assumption that less complicated means less accurate, then the right balance between interpretability and accuracy will have to be considered in clinical scenarios.

As a technical term in machine learning, accuracy is basically the number of correct predictions divided by the number of total predictions.<sup>6</sup> With regard to the accuracy of DP predictions, even if a system was not sufficiently accurate to be automatically acted upon, it could still suffice to be useful and provide information to a clinician that enables them to focus on considering certain possibilities whilst perhaps eliminating others. Imperfect predictive models can still provide clues and certain digital footprint information can still disconfirm certain possibilities, despite being insufficient to conclusively confirm the truth. We will now illustrate these points with a couple of short examples.

The first simple example is one involving DP in terms of pieces of digital footprint information that clinicians can incorporate into their judgements and decision-making. In this scenario, we imagine employing DP to help determine the condition of an individual client who probably has a mood disorder. Whilst there is no DP system that can definitively diagnose the specific type of mood disorder an individual has, certain pieces of information from this individual's digital footprint indicative of mania, such as a sudden and significant decreased need for sleep or abnormal social media such as hyperactive Twitter sprees (Kadkhoda et al., 2022; O'Connor, 2013), could mount as evidence that the individual is experiencing a manic episode characteristic of bipolar disorder rather than a major depressive episode. With this starting point, a clinician can then further explore the symptoms and experiences of the client without relying entirely on DP. This example highlights clinical information

that may not always be picked up within a self-report-based, time-limited clinical assessment. Rather, the information derived from DP may enable detection of symptoms that the client did not think to report or the clinician did not ask about.

The second example is one involving DP and AI-driven psychometric prediction. The DASS is a set of three self-report scales developed to measure the negative emotional states of depression, anxiety, and stress (University of New South Wales, 2022). It is often used as a screening tool, rather than a diagnostic test. Suppose that a regression predictive DP model was constructed that provided equally adequate results for each of the three subscales, in that for each of the three subscales, the model had an evaluation metric of approximately  $x$  (e.g.  $x$  could be some rather good R-squared value above 0.5, but certainly less than 1). If an individual's data were input into this model prior to an initial consultation appointment with a clinician, and their predicted score for the depression subscale fell within the normal range, anxiety within the mild range and stress within the severe range, then the clinician could use that as a starting point, that they are more likely to be experiencing stress rather than depression or anxiety, or that stress is the primary issue.

Similarly, if a highly accurate machine learning classifier was to predict in some individual the presence of a transdiagnostic mechanism  $A$  rather than  $B$ , then a clinically helpful *starting point* for working with that individual would be an exploration of  $A$ . For example, in a modular, personalised approach to treatment (Sauer-Zavala et al., 2017).

Starting with a brief survey of the existing DP landscape by way of both advocacy and critiques, in this section we have focused on establishing a preliminary case for DP by broadly providing justifications and illustrations of ways in which it could be applied and utilised in clinical scenarios. We now turn to three distinct issues concerning DP in clinical practice.

## Ethical Issues in Assessment/Testing

Psychological testing is a practice with inherent ethical issues. Apart from the reliability and validity of the test itself, clinicians must be (1) competent (the psychologist has the requisite knowledge and training to determine appropriate tests and how to administer them), (2) get informed consent (the test taker has agreed to be evaluated prior to testing and after being informed about the test's purpose), and (3) maintain confidentiality (test results are not to be shared or released without client consent or legal necessity) (Miller & Evans, 2004). Test security should also be maintained; although many tests can be open to the public and used for self-assessment, there are other tests that should

<sup>6</sup> See (Google, 2022a) for a brief overview. There are also the related metrics of precision (the proportion of correct positive identifications) and recall (the proportion of positives identified correctly) (Google, 2022b).

not be disseminated<sup>7</sup> and must be conducted under strict conditions. Given the significant difference between using a traditionally administered psychometric test and using a DP model to predict a psychometric test score, there certainly seem to be differences and novel considerations regarding the ethical issues of testing when done via DP.

A first thing to point out or reemphasise relates to the material difference between a traditional psychometric test and a DP testing system. A traditional test, whether it be pen and paper or computerised, will have its results stored as a single entity. A DP system on the other hand would consist of two components, an individual's dataset and the psychometric model that is applied to the data. Hence, the separability of these two components in a DP system would provide an advantage for the security of tests results; rather than storing results that are obtained, one could just couple the dataset with the machine learning model when it comes time to calculate a score, otherwise leaving these two components stored separately. On the other hand, the existence of a model that correlates with a psychometric test means that if the model is obtained and an individual's digital footprint data is also obtained, then a calculation of that psychometric test for that individual can be performed 'outside of the individual', even without their knowledge or consent. This contrasts with the potential results of a psychometric test traditionally obtained, results which are in a sense 'protected' in the brain of an individual and must require their active input to be obtained.

If a clinician does not actually hold the psychometric test results in the traditional sense, then there are no results that need to be confidentially held. Rather, the digital footprint data needs to be securely stored and those with access to models and the data should not use them unless permitted to by client, proxy or applicable authorities. Test security is therefore translated into an issue of model security. If a model has been developed, which when applied to digital footprint data calculates the results of a certain psychometric test that comes with protections, then that model should also be treated with protections. Regarding the issue of competence, if DP models based on clinical psychometric measures are to be used on data, then their results should be used by one who is qualified to determine the appropriateness of the test in its traditional form. Similarly, there may be an imperative that only those who are qualified to use a measure are qualified to request that a model based on it be run.

Psychologists should opt for gold-standard psychometric tools, and according to the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct (Section 9.02), psychologists should 'use assessment

instruments whose validity and reliability have been established for use with members of the population tested' (American Psychological Association, 2017). In terms of AI psychometric prediction models, they should, to begin with, be trained on data collected via such gold-standard psychometric tests. Beyond this, such models should be sufficiently robust and valid in terms of machine learning evaluation metrics. Also, as a consideration at the intersection of the traditional requirement concerning test population membership and the issue of bias in AI, psychometric machine learning models should only be applied to individuals who are represented in the training set on which the model was trained. A final consideration, somewhat related to competence and gold-standard tools but particular to AI forms of psychometric assessment, concerns the maintenance of models and their retraining: models should be feasibly updated when new sets of data are received that lead to retraining and improved evaluation metrics.

These considerations also relate to the issues of informed consent and autonomy (Ioannidis, 2013; Koocher & Rey-Casserly, 2003; Ulrich et al., 2020). In the case of a hypothetical setup where automated DP models are providing psychometric score predictions, the DP system becomes a proxy for traditional psychometric testing and there would be no actual test instance for the client to consent to. Rather, clients will initially need to be informed of the possibility of psychometric evaluations based on their digital footprint data being made. This should be at the time of data collection being set up, so could form part of the general consent procedure required in the case of collecting digital footprint data. Clients may then in advance consent to periodic psychometric predictions based on their data being performed, or a system could be put in place in which the client needs to be asked for consent prior to each instance of applying a model to their data to calculate some psychometric evaluation. Thus, a notion of dynamic informed consent comes into effect here (Tauginienè et al., 2021), and a distinction arises between consenting to some data being collected and consenting to the generation of AI inferences based on that data. The origins of dynamic informed consent (DIC) can be traced to the increasing complexity of biomedical research and the ethical challenges associated with traditional, one-time consent processes. Key factors contributing to the development of DIC include advances in genomics and personalised medicine, technological developments, patient-centred care movements and challenges in longitudinal studies. In terms of digital phenotyping, whilst DIC can help to prevent people from being 'tested' without their knowledge, ultimately it cannot stop cases where an individual's data is used in violation of their consent withdrawal or request for the data to be deleted. The prevention of such unethical data usage would require further security measures, such as implementing a

<sup>7</sup> Either because they need to be administered by a professional or due to copyright.

setup whereby each instance of using an individual's data requires that individual to provide an access key.

A final consideration regarding psychometric testing, in which DP versions raise further novel issues, concerns the release and explanation of test results. Unless there are legal/ethical restrictions or there could be a detrimental impact on the client, clinicians must provide clients with requested test results: according to the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct (Section 9.10), psychologists must 'take reasonable steps to ensure that explanations of results are given to the individual or designated representative' (American Psychological Association, 2017). Whether a clinician should share test results with a client does not depend on the method used, whether it be traditional testing or DP. However, providing test results or explaining finding details to clients in clear and plain language is something that may be more difficult or will not always be possible in the case of DP testing. Traditional test scoring and interpretation largely consists of processing direct client responses to clear and intelligible questions. These questions are granular items that are often grouped into thematic subdimensions that form the whole test. With DP models that are based on predicting aggregate measure scores, it will generally not be possible to directly map information features onto test components, and there will not be the same type of item, dimensions and overall semantic structure as with a traditional test. This however does raise consideration of the possibility of developing a predictive model for each item or dimension of a psychometric measure. For example, the GAD-7 results in a score ranging from 0 to 21, yet there are numerous ways in which the seven individual item scores (0–3) can be combined to result in some final score, say 15. If DP could provide results broken down into a prediction for each item<sup>8</sup>, such granularity would in a sense provide more transparency/explicability.

This leads to another general issue that concerns complicated machine learning constructions, that of explainability/interpretability, as discussed in the 'The Applicability and Utility of Digital Phenotyping' section. If DP models were employed, then simpler, more interpretable models ought to be favoured if they suffice for the job. If more complicated models were opted for, then the introduction of associated explainability requirements becomes a factor. Are complex and accurate models that produce reliably correct results generated by processes a clinician has little to no insight into (i.e. black boxes) ever acceptable in mental health practice? Whilst deeper exploration of this question is beyond our

current scope, perhaps in certain low stakes cases or cases where AI outputs serve to just kickstart human decision-making and verification the answer could be yes. At any rate, although black boxes are an issue for AI in general, accountability and transparency are particularly important in (mental) healthcare applications given the significant medicolegal and ethical import of decision-making in this domain.

## Roles and Responsibilities of Mental Health Practitioners

Some have suggested (or worried) that digital technologies may replace traditional tasks of mental health practitioners and even supplant human professionals entirely (Blease et al., 2020; Brown et al., 2021; Innes & Morrison, 2021; Simon & Yarborough, 2020). In contrast to these radical claims, our suggestion is simply that DP may assist practitioners *without* replacing them or their expertise. In fact, their roles and responsibilities should remain essentially the same, with potential time savings.

To further develop this position, we should distinguish between digital technologies involved in assessment, diagnosis, or treatment suggestions (e.g. expert systems) and digital technologies involved in interventions. DP belongs to the former, whilst examples of the latter are mental health apps and psychotherapy chatbots (Vaidyam et al., 2019), which can be powered by DP or other forms of AI in the way of content recommender systems (Valentine et al., 2022), ecological momentary interventions (Balaskas et al., 2021) or personalisation more generally. Although these technologies may sometimes be useful standalone options or complements/supplements to face-to-face therapy, they cannot replace human therapists and their range of skills, including caring and humanistic skills, clinical judgement and ethical evaluation. Even sophisticated AI agents that mimic psychotherapeutic exchange are nowhere near a level of adequacy and will perhaps never achieve the full spectrum of required qualities. Thus, the replacement of human therapists with AI is not on the horizon. Similarly, we do not think that assessment or diagnosis could or should be fully executed by AI. Digital phenotyping is at best a tool to inform the clinician and augment their practice rather than to supersede it.

The two types of usage we established in the 'An Explication of Digital Phenotyping' section involve varying degrees of both AI and clinician involvement. In manual DP, AI might be involved in converting raw data to information concerning phone sensors (e.g. geolocation patterns), but there is little to no AI involved in making behavioural inferences, assessments, or diagnoses. At most, AI might be used to offer an inference from digital footprint data to something which that data says about the individual. For example, a Bayesian classifier that provides a probability that an

<sup>8</sup> For example, predicting a score of 2 for the item 'Feeling nervous, anxious, or on edge', a score of 3 for the item 'Not being able to stop or control worrying', etc.

individual has engaged in some behaviour given such and such data. The clinician would then receive this result and decide what to do with it. Hence, clinical decisions would not be automated, but rather informed by digital techniques.

AI-driven DP, as we have explained, is more complex. Yet even here it is reasonable to posit that a practicable setup should still involve a clinician making the final judgement and determining the value of outputs provided by the AI. Like all machine learning models that capture complex systems, realistic DP models that predict psychometric outcomes would sometimes get it wrong. As clinicians themselves insist about standard tools, whilst psychological tests can aid in diagnosis, test results alone should not form the basis of a diagnosis (Arslan, 2018). Such tests can involve inaccuracies in recollection, judgement, interpretation, etc. Moreover, in healthcare generally few tests are definitive (just as few symptoms are pathognomonic); rather, diagnoses typically require clinical judgements based on a variety of evidence. This means that even if a DP process could exactly replicate a psychometric test, it should not be accepted uncritically.

Others have quite rightly argued (Birk & Samuel, 2020; Coghlan & D'Alfonso, 2021) that psychological assessment often involves richer qualitative components that cannot be captured by pure data collection and processing. This can be because of limits to quantitative methods and because of fundamental limitations in the collected data. One significant example for DP is its temporal limitations. Mental health practice often requires recollection of detailed personal information and childhood beliefs that cannot be obtained via digital footprints (Royer, 2021). Unless a new generation of people are (highly questionably!) the subjects of digital footprint collection or ecological momentary assessments (EMA)<sup>9</sup> beginning in early childhood, DP will be limited by its inability to capture salient information from earlier periods in individuals' lives.

Another potential limitation is related to the fact that some first-person claims are partly constitutive of certain conditions (Coghlan & D'Alfonso, 2021). Consider the constitutive role that a person's expression of despair and hopelessness plays in depression. Such a confession in a professional consultation may be missing from the digital footprint—and may even be in tension with it—and yet it might be important in arriving at a confident and nuanced diagnosis. Thus, a person's self-interpreted social situation must sometimes be taken into account by the clinician who employs DP (Birk & Samuel, 2020, p. 8) At the same time, we should appreciate that personal interpretations are

not necessarily definitive either, and the 'objective' data provided by smartphones—such as from geolocation and swiping (Birk & Samuel, 2022, p. 4)—may be important in consolidating diagnoses and even in questioning first-person claims or offering an alternative view.

Thus, DP, even if powerful, should not supplant clinician assessment but should at most facilitate or augment practice, and mental health practitioners who employ DP can and should bring their experience and empathy, along with other human skills and ethical qualities, to their practice.

Nevertheless, if DP systems become able to deliver valuable and timely behavioural or diagnostic information in enough cases, then an ethical obligation to employ them could emerge. Increased convenience for clients and time and cost savings might further support such a responsibility. For example, rapid application of a psychometric model to an individual's digital footprint data might yield timely and useful information without the need for active test participation. Such testing might also serve as an early warning sign of the relapse or commencement of mental ill-health and prompt the individual to see a clinician sooner rather than later.

The Universal Declaration of Ethical Principles for Psychologists states that both maximizing benefits and minimizing potential harm to clients are an ethical responsibility (Assembly of the International Union of Psychological Science, 2008). It is an open question whether DP will in the future create such benefits without imposing unacceptable harms for clients. As we stress, a sufficiently strong evidence base concerning DP's harms and benefits would be required before it is used, let alone professionally required, in clinical practice.

Digital technology advocates have argued that 'by outsourcing some aspects of medical care to machine learning, physicians will be freed up to invest more time in higher quality face-to-face doctor-patient interactions' [53, p. 2]. Technology can, however, sometimes increase a practitioner's burden rather than diminish it. As one article puts it, 'if providers receive large volumes of unsolicited data and/or data that do not directly inform their clinical work, they may perceive such a model to be burdensome and unhelpful' (Marsch, 2021, p. 194). Indeed, we have heard directly from clinicians we have worked with that a great influx of client data and the responsibility for interpreting it could potentially be overwhelming for practitioners. Any future implementable DP systems would therefore need an implementation and information architecture that accommodates the practicing clinician and facilitates their negotiation of the DP information.

It is also an open yet important question whether DP will enable mental healthcare to reach greater numbers of people who cannot immediately visit a practitioner or who might miss appointments. Some critics of DP rightly raise

<sup>9</sup> EMAs involve collecting active data from individuals in real time and in their natural environments, such as questionnaires delivered via smartphone.

the possibility that digital technologies could reduce necessary client-practitioner contact (Tekin, 2021). These issues (which we lack space and scope to further discuss here) would all need to be addressed if or when DP achieves a useful maturity.

Finally, if DP is clinically warranted and the various ethical issues concerning it are suitably addressed, then mental healthcare professionals will need to be educated about its operation and its pros and cons. One qualitative study on psychiatrists' views about AI innovations reported that:

scarce reflection on the concept of digital phenotyping and the use of diagnostic and triaging apps among respondents contrasts with the predictions of biomedical informaticians who argue that apps and mobile technologies will play an increasing role in accumulating salient personal health information (Blease, 2020, p. 14).

Another study (Boeldt et al., 2015) found that although care professionals are generally supportive of digital health technologies for information collection and diagnostics, clients are comparatively more enthusiastic. A high level of interest (Torous et al., 2014) among the public for using digital footprint data for mental health monitoring would be one impetus for its future uptake by mental health professions and the education that would need to accompany it.

### Awareness, Empowerment, and Strengthening the Therapeutic Alliance

In this final section, we focus on the manual DP (introduced in the 'An Explication of Digital Phenotyping' section) to consider the ethical benefits of employing DP where it has potential to empower the client and support their therapeutic relationship with their clinician. In this consideration, we draw from a qualitative study conducted by some of the current authors which looked at clinician perspectives on how DP can inform client treatment (Schmidt & D'Alfonso, 2023). This study explored whether, and if so how, an individual's digital footprint information could be presented on a dashboard system for clinicians to inform their practice. Our participants stated that such information could be of value and explored ways this more immediate form of DP could augment therapeutic practice.

A concern generated by the study was that if DP were to inform therapeutic treatment the client should have control in this socio-technical system. In the discourse on digital (mental) health, *control* brings up the related term *empowerment* and the perspective that technology which facilitates an information sharing/collecting process empowers the client in their care. This perspective has been criticised for placing excessive responsibility on the individual to take care

of themselves, whilst disenfranchising them from the support of professionals and institutions who in turn relinquish responsibility for the individual (Burr et al., 2020; Lovatt & Holmes, 2017). In the case proposed here, a person-centred DP system's effectiveness depends not only upon client control, but also on the careful guidance of the clinician and the forging of a feedback loop between the two—a strong *therapeutic alliance*.

The therapeutic alliance (TA), an established common factor in successful psychotherapy outcomes, is understood as integral to the client's treatment and therapeutic transformation. Bordin conceptualised the TA by way of tasks, bond, and goals (Bordin, 1979). In this conceptualisation, the client's trust in their therapist enables them to forge a bond with their therapist and to work in collaboration with them towards agreed-upon therapeutic goals and tasks to achieve these goals. Horvath and Luborsky demarcated the alliance in terms of Type 1—the therapist is supportive of the client and Type 2—the therapist and client work together with *shared responsibility* on the client's treatment goals (Horvath & Luborsky, 1993).

In the discourse on digital mental health, there is concern that technologies employed to facilitate mental health treatment, including DP, could degrade the TA (Ben-Zeev, 2017). There is the fear that technology will replace the clinician, resulting in an erasure of human care. In the case proposed here, technology would *extend* rather than replace the therapist (Pickersgill, 2019). The person-centred DP system extends therapist and client, capturing client data significant to treatment and delivering ecological momentary interventions/assessments (EMI/As)<sup>10</sup> in between therapy sessions.

As mentioned, a chief concern for study participants was client agency. Participants considered that clients in sharing their data would be vulnerable to feeling monitored and judged, which could degrade the TA and the effectiveness of the therapy. Clinicians stressed the importance of clients controlling what data they share, when they share it, and being able to pause or cease this process at any point, as well as understanding its significance and benefit. The person-centred DP system mirrors the way that clients control what they share through conversation with their therapist and their agency to pause or cease therapy. The system is co-designed by client and therapist, informed by their agreed-upon tasks and goals. Central to this system is client agency, which may increase through the client's potential to develop greater self-awareness as the processes of DP integrate into therapy. Participants noted that to protect client control and support client self-awareness, the clinician would need to *steward*

<sup>10</sup> EMIs can be defined as mini digital interventions or exercises delivered via devices such as smartphones that clients/users can do in situ.

the system, carefully guiding their client in the DP process. The quality of this stewardship could determine whether the TA is strengthened or weakened. Just as in traditional therapy, if in the DP process the client feels judged, monitored, ashamed, confused or disengaged, the TA would degrade. If the client feels understood, aware, and empowered, as if they are working together with their therapist towards their goals and that they are receiving relevant care, then the TA would be strengthened. But how can DP achieve this?

Excessive smartphone usage and the negative influences of social media have been posited as contributing to the prevalence of mental health issues, particularly in youth populations (Abi-Jaoude et al., 2020; Valkenburg et al., 2022). Given that people spend so much time on their devices, these devices could be harnessed to instead support and strengthen mental health, enabling self-awareness and informed choices in support of their wellbeing. DP captures information on sleep, mood, physical activity, and device usage (e.g. social media and communication). The relationships between these captures could offer insights to explore therapeutically and inform client treatment. For example, if the client has social anxiety and the DP system signals the client had been avoiding people's calls, this information may be discussed in therapy. Phone social (in)activity may speak to a range of mental health states and patterns of relating (Reid & Reid, 2007). Through exploring these with their therapist, the client might gain more awareness regarding their relational dynamics in relation to their wellbeing. The client's data could also inform EMA/Is (Myin-Germeys et al., 2016; Shiffman et al., 2008) aligned to their treatment. For example, the client aspires to decrease social media usage and the DP system is designed accordingly. The system signals excessive hours on Instagram and triggers an EMA asking how the client feels, supporting them in becoming aware of how their digital activity impacts their mood. In this way, the EMA extends the therapist in its co-regulation of the client to enable more conscious technology use to support wellbeing (Almoallim & Sas, 2023). This example demonstrates how the EMA could strengthen the TA, reminding the client outside of the therapy session that they are working towards a shared goal with their therapist. The client may experience a sense of collaboration with their therapist whilst they go about their day-to-day life, which is where the insights of therapy must be integrated. This extension of the TA through the digital device has been more recently discussed in terms of the digital therapeutic relationship (Torous & Hsin, 2018) and the digital therapeutic alliance (D'Alfonso et al., 2020).

Participants in the study (Schmidt & D'Alfonso, 2023) underlined the importance of *holding the data lightly* when exploring client data rather than over-determining the significance of these data. This exploration guided by the clinician could facilitate collaboration with the client, supporting client agency and client-clinician awareness in the DP system,

whilst strengthening the TA. The client-clinician awareness generated in the person-centred DP system would be the outcome of the careful meeting between quantitative capture and client-clinician qualitative exploration and interpretation. In this system, there would be scope to employ AI to generate insights in relation to an individual's dataset, particularly in terms of highlighting key relationships and patterns tailored to an individual's therapy. A period of intensive data capture would be needed before the individual's patterns and baselines could be determined to inform AI models that could then generate meaningful insights from the data. It is during the intense data capture that the qualitative work of client-clinician exploration and interpretation of the data leading to greater awareness would be done. This qualitative work could then possibly inform the framework for the individualised AI models that could then perform automated processing of the data such that it could be made meaningful to the individual's treatment.

Tailoring DP to the individual returns us to where we first began this section: the critique of empowerment (via digital health technologies) in its association with individual responsibility:

By focusing on the individual [...] these technologies reduce health problems to the micro, individual level. Such approaches do little, therefore, to identify the broader social, cultural and political dimensions of ill-health (Lupton, 2012, p. 239).

A valuable point—technology often shines a light on as well as amplifies the limitations of our systems of knowledge, such as those of our (mental) healthcare system, as well as the broader societal problems of which these systems are a product. The distinction to make here is that the individual is not responsible for the arising of their depression, anxiety, social media addiction, and so on, all of which are the outcome of 'problems of living' (Birk & Samuel, 2020). However, the individual, along with their various supports—family, friends, educators, colleagues, communities, and therapists—*shares the responsibility* in navigating a way to overcome these *problems of living*. If DP can assist the individual in becoming more aware and therefore empowered in moving towards wellbeing, then it is worth considering in the context of mental health treatment.

## Conclusion

This paper has presented a cautious case for investigating the use of DP in clinical practice, justifying it as a research enterprise, and exploring its ethical dimensions within this context. We noted that some criticisms of DP contain important points. However, we also argued that DP needs to be understood not in an overly abstract way but in

consideration of facts about clinical practice and about the practical applicability of DP and how it might actually be implemented. It is true that DP, assuming it is empirically validated, will result in missed diagnoses and misdiagnoses, but that is true of all almost all tests. Issues of misjudgement are certainly real in relation to DP and must be carefully thought through. Yet there could also be ways of limiting errors which the practitioner (and client) can be aware of and engage with. Furthermore, DP could be implemented in ways that empower clients and improve the therapeutic alliance. If the technology proves to have potential, there will need to be a careful weighing up and balancing of the benefits and risks to reach a considered judgement about the possible role of this technology in mental healthcare.

In light of this paper, we conclude with a few suggestions. Firstly, whilst AI-driven DP remains exploratory and merits further research, the present concrete possibility of implementing and investigating manual DP platforms in clinical practice calls for more attention. Secondly and relatedly, work is required in terms of promoting the possibilities of digital technology in mental health and developing programs to educate mental health professionals on DP and AI in mental health more generally. Finally, further research into DP should include gaining much needed client as well as clinician perspectives to inform development and implementation and investigating how DP can enhance the therapeutic alliance.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions No funding was received to assist with the preparation of this manuscript.

**Data Availability** No datasets were analysed or generated from this work.

## Declarations

**Competing Interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abi-Jaoude, E., Treurnicht Naylor, K., & Pignatiello, A. (2020). Smart-phones, social media use and youth mental health. *Canadian Medical Association Journal*, 192(6), E136–E141.
- Aguera y Arcas B. (2017). Physiognomy's new clothes [Online]. Available: <https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a>. Accessed 10 Feb 2023.
- Almoallim, S., & Sas, C. (2023). Patterns of meaningful and meaningless smartphone use: A diary study. In *6th International BCS Human-Computer Interaction Conference, York, UK*.
- Alhubaiti, A. (2016). Information bias in health research: Definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare*, 9, 211–217.
- American Psychological Association. (2017). Ethical principles of psychologists and code of conduct [Online]. Available <https://www.apa.org/ethics/code/index>. Accessed 29 Jan 2022.
- Arslan, R. (2018). A review on ethical issues and rules in psychological assessment. *Journal of Family, Counseling, and Education*, 3(1), 17–29.
- Assembly of the International Union of Psychological Science. (2008). Universal declaration of ethical principles for psychologists [Online]. Available <https://www.iupsys.net/about/declarations/universal-declaration-of-ethical-principles-for-psychologists/>. Accessed 8 Feb 2023.
- Balaskas, A., Schueller, S. M., Cox, A. L., & Doherty, G. (2021). Ecological momentary interventions for mental health: A scoping review. *PLoS ONE*, 16(3), e0248152.
- Barnett, I., Torous, J., Staples, P., Sandoval, L., Keshavan, M., & Onnela, J.-P. (2018). Relapse prediction in schizophrenia through digital phenotyping: A pilot study. *Neuropsychopharmacology*, 43, 1660–1666.
- Ben-Zeev, D. (2017). Technology in mental health: Creating new knowledge and inventing the future of services. *Psychiatric Services*, 68(2), 107–108.
- Bennetts, A. (2021). The infinity formulation: How transdiagnostic behaviours and endeavours for behavioural change serve to maintain co-morbid mental health presentations. *The Cognitive Behaviour Therapist*, 14, e13.
- Bhugra, D., Tasman, A., Pathare, S., Priebe, S., Smith, S., Torous, J., Arbuckle, M. R., Langford, A., Alarcón, R. D., Fung Kum Chiu, H., First, M. B., Kay, J., Sunkel, C., Thapar, A., Udomratn, P., Baingana, F. K., Kestel, D., Man Kin Ng, R., Patel, A., ... Ventriglio, A. (2017). The WPA-Lancet Psychiatry Commission on the future of psychiatry. *The Lancet Psychiatry*, 4(10), 775–818.
- Birk, R., & Samuel, G. (2020). Can digital data diagnose mental health problems? A sociological exploration of 'digital phenotyping.' *Sociology of Health & Illness*, 42(8), 1873–1887.
- Birk, R. H., & Samuel, G. (2022). Digital phenotyping for mental health: Reviewing the challenges of using data to monitor and predict mental health problems. *Current Psychiatry Reports*, 24, 523–528.
- Blease, C., Locher, C., Leon-Carlyle, M., & Doraiswamy, M. (2020). Artificial intelligence and the future of psychiatry: Qualitative findings from a global physician survey. *Digital Health*, 6, 2055207620968355.
- Boeldt, D., Wineinger, N., Waalen, J., Gollamudi, S., Grossberg, A., Steinhubl, S., McCollister-Slipp, A., Rogers, M., Silvers, C., & Topol, E. (2015). How consumers and physicians view new medical technology: comparative survey. *Journal of Medical Internet Research*, 17(9), e215.

- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, 16(3), 252–260.
- Brown, C., Story, G., Mourão-Miranda, J., & Baker, J. (2021). Will artificial intelligence eventually replace psychiatrists? *The British Journal of Psychiatry*, 218(3), 131–134.
- Burr, C., & Morley, J. (2020). Empowerment or engagement? Digital health technologies for mental healthcare. In C. Burr & S. Milano (Eds.), *The 2019 yearbook of the digital ethics lab* (pp. 67–88).
- Carreiro, S., Chai, P. R., Carey, J., Lai, J., Smelson, D., & Boyer, E. W. (2018). mHealth for the detection and intervention in adolescent and young adult substance use disorder. *Current Addiction Reports*, 5, 110–119.
- Coghlan, S., & D'Alfonso, S. (2021). Digital phenotyping: An epistemic and methodological analysis. *Philosophy & Technology*, 34(4), 1905–1928.
- Corcoran, C. M., Carrillo, F., Fernández-Slezak, D., Bedi, G., Klim, C., Javitt, D. C., Bearden, C. E., & Cecchi, G. A. (2018). Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry*, 17(1), 67–75.
- Cornet, V. P., & Holden, R. J. (2018). Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of Biomedical Informatics*, 77, 120–132.
- Crawford, K. (2021). Time to regulate AI that interprets human emotions. *Nature*, 592(7853), 167.
- D'Alfonso, S., Lederman, R., Bucci, S., & Berry, K. (2020). The digital therapeutic alliance and human-computer interaction. *JMIR Mental Health*, 7(12), e21895.
- Garg, M. (2023). Mental health analysis in social media posts: A survey. *Archives of Computational Methods in Engineering*, 30, 1819–1842.
- Google. (2022a). Classification: Accuracy | Machine learning | Google developers [Online]. Available: <https://developers.google.com/machine-learning/crash-course/classification/accuracy>. Accessed 8 Feb 2023.
- Google. (2022b). Classification: Precision and recall | Machine learning | Google developers [Online]. Available <https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall>. Accessed 8 Febr 2023.
- Horvath, A. O., & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *Journal of Consulting and Clinical Psychology*, 61(4), 561–573.
- Huckvale, K., Venkatesh, S., & Christensen, H. (2019). Toward clinical digital phenotyping: A timely opportunity to consider purpose, quality, and safety. *NPJ Digital Medicine*, 2(1), 1–1.
- Innes, J. M., & Morrison B. W. (2021). Machines can do most of a psychologist's job. The industry must prepare for disruption. The Conversation [Online]. Available <https://theconversation.com/machines-can-do-most-of-a-psychologists-job-the-industry-must-prepare-for-disruption-154064>. Accessed 16 Jan 2022.
- Insel, T. R. (2018). Digital phenotyping: A global tool for psychiatry. *World Psychiatry*, 17(3), 276–277.
- Ioannidis, J. P. A. (2013). Informed consent, big data, and the oxymoron of research that is not research. *The American Journal of Bioethics*, 13(4), 40–42.
- Jacobson, N., Summers, B., & Wilhelm, S. (2020). Digital biomarkers of social anxiety severity: Digital phenotyping using passive smartphone sensors. *Journal Medical Internet Research*, 22(5), e16875.
- Kadkhoda, E., Khorasani, M., Pourgholamali, F., Kahani, M., & Rezaei Ardani, A. (2022). Bipolar disorder detection over social media. *Informatics in Medicine Unlocked*, 32, 101042.
- Koocher, G. P., & Rey-Casserly, C. M. (2003). Ethical issues in psychological assessment. In *Handbook of psychology: Assessment psychology* (Vol. 10, pp. 165–180). John Wiley & Sons Inc.
- Liang, Y., Zheng, X., & Zeng, D. D. (2019). A survey on big data-driven digital phenotyping of mental health. *Information Fusion*, 52, 290–307.
- Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable AI: A review of machine learning interpretability methods. *Entropy*, 23(1), 18.
- Loi, M. (2019). The digital phenotype: A philosophical and ethical exploration. *Philosophy & Technology*, 32, 155–171.
- Lovatt, M., & Holmes, J. (2017). Digital phenotyping and sociological perspectives in a Brave New World. *Addiction*, 112(7), 1286–1289.
- Lupton, D. (2012). M-health and health promotion: The digital cyborg and surveillance society. *Social Theory & Health*, 10, 229–244.
- Luxton, D. (2022). AI decision-support: A dystopian future of machine paternalism? *Journal of Medical Ethics*, 48, 232–233.
- Marsch, L. A. (2021). Digital health data-driven approaches to understand human behavior. *Neuropsychopharmacology*, 46, 191–196.
- Martinez-Martin, N., Greely, H. T., & Cho, M. K. (2021). Ethical development of digital phenotyping tools for mental health applications: Delphi study. *JMIR mHealth and uHealth*, 9(7), e27343.
- Mastoras, R.-E., Iakovakis, D., Hadjimiditriou, S., Charisis, V., Kassie, S., Alsaadi, T., Khandoker, A., & Hadjileontiadis, L. J. (2019). Touchscreen typing pattern analysis for remote detection of the depressive tendency. *Scientific Reports*, 9(1), 13414.
- Masud, M. T., Mamun, M. A., Thapa, K., Lee, D., Griffiths, M. D., & Yang, S.-H. (2020). Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. *Journal of Biomedical Informatics*, 103, 103371.
- McCadden, M., Hui, K., & Buchman, D. Z. (2023). Evidence, ethics and the promise of artificial intelligence in psychiatry. *Journal of Medical Ethics*, 49, 573–579.
- Miller, C., & Evans, B. B. (2004). Ethical issues in assessment. In M. Hersen (Ed.), *Psychological assessment in clinical practice: A pragmatic guide* (pp. 21–32). Brunner-Routledge.
- Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology*, 13(1), 23–47.
- Mulvenna, M. D., Bond, R., Delaney, J., Dawoodbhoy Mustansir, F., Boger, J., Potts, C., & Turkington, R. (2021). Ethical issues in democratizing digital phenotypes and machine learning in the next generation of digital health technologies. *Philosophy & Technology*, 34, 1945–1960.
- Myin-Germeyns, I., Klippel, A., Steinhart, H., & Reininghaus, U. (2016). Ecological momentary interventions in psychiatry. *Current Opinion in Psychiatry*, 29(4), 258–263.
- Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the internet: New problems, old issues. *American Psychologist*, 59(3), 150–162.
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Bethesda, MD: Office for Human Research Protections (OHRP).
- Nguyen, T. M., Leow, A. D., & Ajilore, O. (2023). A review on smartphone keystroke dynamics as a digital biomarker for understanding neurocognitive functioning. *Brain Sciences*, 13(6), 959.
- O'Connor, M. (2013). Can you diagnose a manic episode on Twitter? [Online]. Available <https://www.thecut.com/2013/08/can-you-diagnose-a-manic-episode-on-twitter.html>. Accessed 8 Feb 2023.
- Oudin, A., Maatoug, R., Bourla, A., Ferreri, F., Bonnot, O., Millet, B., Schoeller, F., Mouchabac, S., & Adrien, V. (2023). Digital

- phenotyping: Data-driven psychiatry to redefine mental health. *Journal of Medical Internet Research*, 25, e44502.
- Perez-Pozuelo, I., Spathis, D., Clifton, E. A., & Mascolo, C. (2021). Wearables, smartphones, and artificial intelligence for digital phenotyping and health. In *Digital health: Mobile and wearable devices for participatory health applications* (pp. 33–54). Elsevier.
- Pickersgill, M. (2019). Digitising psychiatry? Sociotechnical expectations, performative nominalism and biomedical virtue in (digital) psychiatric praxis. *Sociology of Health & Illness*, 41(S1), 16–30.
- Reid, D. J., & Reid, F. J. M. (2007). Text or talk? Social anxiety, loneliness, and divergent preferences for cell phone use. *CyberPsychology & Behavior*, 10(3), 424–435.
- Rodriguez-Villa, E., Rauseo-Ricupero, N., Camacho, E., Wisniewski, H., Keshavan, M., & Torous, J. (2020). The digital clinic: Implementing technology and augmenting care for mental health. *General Hospital Psychiatry*, 66, 59–66.
- Royer, A. (2021). The wellness industry's risky embrace of AI-driven mental health care [Online]. Brookings Institute. Available <https://www.brookings.edu/techstream/the-wellness-industrys-risky-embrace-of-ai-driven-mental-health-care/>. Accessed 16 Jan 2022.
- Saeb, S., Lattie, E., Schueller, S., Kording, K., & Mohr, D. (2016). The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ*, 4, e2537.
- Saeb, S., Zhang, M., Karr, C., Schueller, S., Corden, M., Kording, K., & Mohr, D. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. *Journal of Medical Internet Research*, 17(7), e4273.
- Sauer-Zavala, S., Gutner, C. A., Farchione, T. J., Boettcher, H. T., Bullis, J. R., & Barlow, D. H. (2017). Current definitions of transdiagnostic in treatment development: A search for consensus. *Behavior Therapy*, 48(1), 128–138.
- Schmidt, S., & D'Alfonso, S. (2023). Clinician perspectives on how digital phenotyping can inform client treatment. *Acta Psychologica*, 235.
- Sheikh, M., Qassem, M., & Kyriacou, P. A. (2021). Wearable, environmental, and smartphone-based passive sensing for mental health monitoring. *Frontiers in Digital Health*, 3, 662811.
- Shen, F. X., Baum, M. L., Martinez-Martin, N., Miner, A. S., Abraham, M., Brownstein, C. A., Cortez, N., Evans, B. J., Germine, L. T., Glahn, D. C., Grady, C., Holm, I. A., Hurley, E. A., Kimble, S., Lázaro-Muñoz, G., Leary, K., Marks, M., Monette, P. J., Onnela, J.-P., ... Silverman, B. C. (2024). Returning individual research results from digital phenotyping in psychiatry. *The American Journal of Bioethics*, 24(2), 69–90.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32.
- Shin, J., & Bae, S. M. (2023). A systematic review of location data for depression prediction. *International Journal of Environmental Research and Public Health*, 20(11), 5984.
- Simon, G. E., & Yarbrough, B. J. (2020). Good news: Artificial intelligence in psychiatry is actually neither. *Psychiatric Services*, 71(3), 219–220.
- Slack, S., & Barclay, L. (2023). First-person disavowals of digital phenotyping and epistemic injustice in psychiatry. *Medicine, Health Care and Philosophy*, 26(4), 605–614.
- Tauginienė, L., Hummer, P., Albert, A., Cigarini, A., & Vohland, K. (2021). Ethical challenges and dynamic informed consent, in *The science of citizen science* (pp. 397–416). Springer.
- Tekin, Ş. (2021). Is big data the new stethoscope? *Perils of Digital Phenotyping to Address Mental Illness, Philosophy & Technology*, 34, 447–461.
- Torous, J., Chan, S., Yee-Marie Tan, S., Behrens, J., Mathew, I., Conrad, E., Hinton, L., Yellowlees, P., & Keshavan, M. (2014). Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: A survey in four geographically distinct psychiatric clinics. *JMIR Mental Health*, 1(1), e4004.
- Torous, J., Gershon, A., Hays, R., Onnela, J. P., & Baker, J. T. (2019). Digital phenotyping for the busy psychiatrist: Clinical implications and relevance. *Psychiatric Annals*, 49(5), 196–201.
- Torous, J., & Hsin, H. (2018). Empowering the digital therapeutic relationship: Virtual clinics for digital health interventions. *NPJ Digital Medicine*, 1(16), e4004.
- Torous, J., Kiang, M., Lorme, J., & Onnela, J. (2016). New tools for new research in psychiatry: A scalable and customizable platform to empower data driven smartphone research. *JMIR Mental Health*, 3(2), e5165.
- Ulrich, C. M., Demiris, G., Kennedy, R., & Rothwell, E. (2020). The ethics of sensor technology use in clinical research. *Nursing Outlook*, 68(6), 720–726.
- University of New South Wales. (2022). Depression anxiety stress scales - DASS [Online]. Available <http://www2.psy.unsw.edu.au/groups/dass/>. Accessed 10 Feb 2023.
- Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, 64(7), 456–464.
- Valentine, L., D'Alfonso, S., & Lederman, R. (2022). Recommender systems for mental health apps: Advantages and ethical challenges. *AI & Society*, 38(4), 1627–1638.
- Valkenburg, P. M., Meier, A., & Beyens, I. (2022). Social media use and its impact on adolescent mental health: An umbrella review of the evidence. *Current Opinion in Psychology*, 44, 58–68.
- Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: A narrative review. *NPJ Digital Medicine*, 5(46), 1–3.
- Zulueta, J., Piscitello, A., Rasic, M., Easter, R., Babu, P., Langenecker, S. A., McInnis, M., Ajilore, O., Nelson, P. C., Ryan, K., & Leow, A. (2018). Predicting mood disturbance severity with mobile phone keystroke metadata: A biaffect digital phenotyping study. *Journal of Medical Internet Research*, 20(7), e241.

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