



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Teng, S;D'Alfonso, S;Kostakos, V

Title:

A Tool for Capturing Smartphone Screen Text

Date:

2024-05-11

Citation:

Teng, S., D'Alfonso, S. & Kostakos, V. (2024). A Tool for Capturing Smartphone Screen Text. Conference on Human Factors in Computing Systems - Proceedings, pp.1-24. ACM. <https://doi.org/10.1145/3613904.3642347>.

Persistent Link:

<https://hdl.handle.net/11343/350569>

License:

[CC BY](#)



A Tool for Capturing Smartphone Screen Text

Songyan Teng
songyan.teng@student.unimelb.edu.au
University of Melbourne
Melbourne, Australia

Simon D'Alfonso
dalfonso@unimelb.edu.au
University of Melbourne
Melbourne, Australia

Vassilis Kostakos
vassilis.kostakos@unimelb.edu.au
University of Melbourne
Melbourne, Australia

ABSTRACT

Context sensing on smartphones is often used to understand user behaviour. Amongst the many available sensors, the collection of text is crucial due to its richness. However, previous work has been limited to collecting text only from keyboard input, or intermittently collecting screen text indirectly by taking screenshots and applying optical character recognition. Here, we present a novel software sensor that unobtrusively and continuously captures all screen text on smartphones. We conducted a validation study with 21 participants over a two-week period, where they used our software on their personal smartphones. Our findings demonstrate how data from our sensor can be used to understand user behaviour and categorise mobile apps. We also show how smartphone sensing can be enhanced by using our sensor in conjunction with other sensors. We discuss the strengths and limitations of our sensor, highlighting potential areas for improvement and providing recommendations for its use.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools.**

KEYWORDS

Screen Text, Smartphone Sensing, Context-Awareness, User Behaviour

ACM Reference Format:

Songyan Teng, Simon D'Alfonso, and Vassilis Kostakos. 2024. A Tool for Capturing Smartphone Screen Text. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 24 pages. <https://doi.org/10.1145/3613904.3642347>

1 INTRODUCTION

We present a new smartphone sensor that captures all screen text. This sensor can make a valuable contribution to the already vast amounts of data generated from smartphone interactions. Understanding user context via smartphone sensing [42] has in recent years led to breakthroughs in disciplines that study human behaviour, such as health [9, 81, 105] and education [92, 102, 120].

Yet, despite the rich sensor data that smartphones make available for context sensing, most of the sensed data is an ambiguous proxy

for the actual behaviour of interest [35]. Smartphone sensors such as location, applications running, and accelerometer/gyroscope, have been shown to correlate well with human behaviour, but must always be analysed in tandem with some alternative source of "ground truth". One crucial sensor that overcomes this limitation is the collection of screen text from smartphones, which can itself act as a basis for "ground truth" in understanding human behaviour [22].

So far, there have been two main ways of collecting text data from smartphones. One approach is to collect typing & keystroke data from smartphones [42]. Such data can provide insights into health-related aspects such as emotion [95] and stress [91], as well as infer broader context, including an individual's environment, social context, or possible distractions [4]. An important drawback of this approach is that the vast majority of screen text is not typed by users, and therefore is not captured. To tackle this, an alternative approach is to capture screenshots from mobile devices with high frequency [86]. These screenshots – or Screenomes – are typically taken every few seconds, providing a detailed visual record of individuals' smartphone usage. This technique enables the analysis of broad aspects of smartphone use, but has multiple limitations due to the use of computer vision techniques to analyse and annotate the screenshots. The main drawbacks include the computational cost, scalability challenges, and transcription errors introduced by the optical character recognition analysis. Additionally, screenomes capture only the phone activity of users, omitting other contextual elements beyond the smartphone such as their location. The lack of this additional information may reduce the accuracy of contextual inference.

We overcome these limitations by developing a software sensor that continuously and unobtrusively captures all screen text on smartphones. Our approach uses Android's accessibility services to capture textual elements on-screen, format the data, and transfer this data to a database for analysis. The use of the accessibility API requires minimal data processing and energy to collect the text, does not suffer from error-prone recognition or gaps in the recordings, and offers precise data that can act as a source of ground truth in understanding human behaviour. Furthermore, the integration of our sensor within the AWARE-Light framework [106] facilitates the seamless, concurrent capturing of screen text alongside a multitude of other sensors, providing researchers with an enhanced sensing tool that allows easy access to collecting data and running studies using screen text. The capabilities of the screen text sensor and its integration in AWARE-Light introduces a novel dimension to smartphone sensing that builds on existing methods. Firstly, it captures a broader range of smartphone interactions compared to keystrokes. Keystrokes reflect only the typing activities of users and reveal little information about the context of their typed content. For instance,



This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI '24, May 11–16, 2024, Honolulu, HI, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0330-0/24/05
<https://doi.org/10.1145/3613904.3642347>

discerning a user's intention solely through keystrokes poses challenges due to the absence of information about the context of their searches or responses. This limitation is addressed by screen text, which captures entire conversations, offering a more comprehensive context for user interactions. Secondly, our screen text sensor is fully integrated with AWARE-Light, a software that captures data from diverse smartphone sensors such as the accelerometer and GPS. By collecting and analysing screen text alongside other sensor data, a broader context beyond the smartphone such as user movement and location can be inferred and correlated with their screen content. In contrast to screenomes, which provide only information about smartphone use, the integration of our screen text sensor within a broader sensor framework allows for a comprehensive analysis of both digital and non-digital activities in tandem, providing researchers with a more comprehensive understanding of user behaviour.

In this paper we present a two-week validation study with 21 participants who used our software in their daily lives. The study aims to identify the strengths and weaknesses of this new sensor in realistic scenarios and demonstrate its potential in understanding human behaviour.

Our paper makes multiple contributions:

- We present a novel smartphone sensor that captures screen text, which integrates with the AWARE-Light framework.
- We demonstrate how the sensor captures and presents textual information from smartphone interactions.
- We show how this sensor can be used in tandem with other sensors for further contextual inference.
- We highlight the strengths and weaknesses of this sensor, and provide recommendations for its use in understanding behaviour.

2 RELATED WORK

2.1 Context-aware smartphone sensing of behaviour

Smartphone sensing has been widely used to understand human context and behaviour in various domains. The combination of smartphone sensing capabilities and their widespread usage makes them effective tools for capturing human behaviour unobtrusively [47, 63]. Integrating data obtained from multiple smartphone sensors enables a comprehensive understanding of a person's surroundings, activities, and behaviours [19, 30]. This understanding can be used to inform recommendations based on the user's context.

Healthcare extensively utilises smartphone sensor data. For example, research in the field of digital phenotyping has shown that GPS could be employed to assess mental health states [16, 18, 31, 74, 78, 89, 109, 111, 125], sociability [13, 46, 51, 112], overall health status and well-being [8, 36, 57, 64, 76], and personality [58]. GPS tracking enables the study of movement as a health indicator, such as understanding that people suffering from depression tend to travel less [53], whereas people who travel more may have larger social networks [97], which in turn increases happiness [27]. Similarly, smartphone accelerometers have been studied to classify physical activity occurrence [7, 54, 65–67, 96, 114, 118] and sleep [50, 79], both informative of lifestyle. Accelerometer data captures precise movement patterns and abnormalities in physical

behaviour. Additionally, smartphone app usage and communication data can be used to inform context in health, such as monitoring student behaviour and performance [21, 34, 110], and mental health [12, 40, 83, 94]. These software sensors provide insights into people's interactions with different content types, aiding in the design of health recommendations and interventions.

Smartphone sensing strategies have also been extensively explored in education, often incorporating multiple sensors. WoBaLearn leverages the smartphone's light sensor and microphone to deliver teaching content in work-based learning, tailoring support to learners' individual educational needs, characteristics, and circumstances [127]. Another approach utilises GPS data in a personalised context-aware recommendation learning system, notifying learners about nearby learning materials based on their current location [120]. Similarly, CALMS employs ontologies that utilise GPS data and additional contextual information, such as academic profiles and time, to offer context-relevant academic content. This approach has shown positive effects on student grades and user satisfaction [38]. In a flipped classroom environment, Louhab et al. [73] collected information from learners' mobile devices, including available software, screen size, battery level, and Internet connectivity, to determine the optimal format for delivering course content. Students using the context-aware application expressed high satisfaction with the content delivery.

Despite the availability of numerous smartphone sensors and the data they provide, combining these sensor data yields only an approximate measure of an individual's smartphone usage. This limitation arises from the inherent constraints in capturing precise content from existing smartphone sensors. Although user context can be inferred from these sensor data, the inability to accurately interpret specific user actions introduces the risk of making erroneous assumptions about their context. This limitation presents a significant challenge in achieving a comprehensive understanding of users' smartphone interactions, necessitating further research to enhance the precision and accuracy of context inference in smartphone sensing applications.

2.2 Text corpora from smartphones

In addition to utilising existing smartphone sensors, analysing textual screen content on smartphones offers another avenue for understanding user context. The proliferation of text-based communication platforms and applications has generated a vast amount of textual data, which can provide valuable insights into users' contexts [15, 33, 121]. User-generated text, such as social media posts, can provide information about the user's personality traits, such as extraversion or agreeableness [5]. Additionally, the characteristics of written text can serve as indicators of demographics, such as younger individuals using more feeling-related words when describing artwork [3]. Therefore, analysing textual information, including browsing content and communications, holds the potential to enhance our understanding of users' behaviour and preferences.

Keystroke sensing can be used to analyse how individuals type on their smartphones, aiming to infer information about their context and behaviour. Tahir et al. [103] demonstrated that employing machine learning methods on keyboard character input and keystroke data achieves high accuracy in identifying emotions, such as

happiness, sadness, and anger, from short text inputs. The number of negative words typed by a user has shown a strong correlation with increased perceived stress and reduced sleep duration [25], suggesting that the sentiment expressed in the text generated by an individual may serve as a predictor of their well-being. Furthermore, keystroke analysis has captured variances in semantic content across different texting platforms. For instance, based on the Linguistic Inquiry and Word Count (LIWC) 2022 dictionary, it was found that people prefer to share content such as books and songs, and discuss leisure activities on Facebook, whereas they were more task-oriented when communicating via SMS [71]. Analysing text generated from keyboard typing also allows for measuring mistakes made while typing and the timing between keystrokes. These metrics have been suggested to be indicators of stress, as individuals experiencing stress tend to type faster and make more errors [29]. Similarly, the accuracy and speed of typing on a smartphone keyboard can provide insights into abnormal upper extremity motor coordination, eye-hand coordination, and manual dexterity, which can be relevant in assessing certain conditions or diseases [48].

While keystroke sensing offers insights into how individuals produce textual content, it will not in general provide information about the types of content they consume or interact with. To address this, smartphone screenomes have emerged as a tool for capturing all textual interactions. Screenomes consist of sequential high-frequency screenshots taken every 5 seconds, capturing individuals' day-to-day digital experiences and facilitating analysis of their engagement with the digital environment [82]. The Screenomics framework studied screenomes extensively, employing optical character recognition (OCR) and image analysis to extract text and image content from the screenshots, respectively [82, 85]. Textual information can be extracted from each screenshot with an accuracy of 74% using OpenCV for image pre-processing and Tesseract for OCR [85]. This allows for a comprehensive understanding of human behaviour on smartphones, including content viewing patterns across various categories and platforms [85], content engagement at different times of the day [22], and task-switching behaviours [119], among other directions of exploration. ScreenLife Capture is an open-source application that reduces the burden of collecting screenome data, allowing researchers to easily run their own screenome studies [122]. It provides the flexibility for participants to start or stop screen capturing as desired, ensuring their comfort and control over data tracking.

Extracting textual content from smartphone screenshots poses challenges due to the content being displayed in multiple fonts and font sizes, resulting in resource-intensive and often inaccurate extraction processes. For instance, Chiatti et al. [28] reported that icons can be incorrectly identified as characters, such as the Bluetooth icon () being interpreted as a "\$" sign. These additional time and computational requirements limit the scalability of studying screenomes for a larger population over an extended period. These inaccuracies in text recognition can potentially compromise the precision of observed trends. Additionally, the 5-second interval between screenome screenshots may miss shorter interactions such as scrolling through text, changing a song, or checking notifications. Given the decreasing attention spans of humans leading to content consumption changing more rapidly [101], the omission of even

brief interaction windows could significantly impact the overall understanding of an individual's context.

To address these challenges, we developed a lightweight sensor to capture textual content on smartphone screens. Our sensor captures textual information directly from the screen in real-time, eliminating the need for additional machine learning processing. Data is collected whenever the sensor is activated, enabling the capture of even brief interactions. Compared to state-of-the-art methods, this approach improves both the accuracy and rate of data capture while reducing computational demands. We demonstrate the technology and capabilities of our sensor in the following sections.

3 METHODOLOGY

3.1 The Screen Text Sensor

We have developed a sensor to continuously capture screen text on Android smartphones. The sensor has been designed to be part of the AWARE-Light smartphone sensing system [106]. Built upon the AWARE framework [42], AWARE-Light enables the collection of data from various smartphone sensors, including keystrokes, screen status (locked/unlocked, on/off), and geolocation data. The application features a user interface that facilitates study sign-up, displays participant information, and provides a list of the sensors being captured once the participant joins the study. With the participant's consent, AWARE-Light passively captures data from the selected sensors without requiring further input or intervention from the participant. By default, the data is transmitted to our own MySQL database in the cloud. In keeping with AWARE-Light's open-source availability, our sensor is also openly available for other researchers to use and conduct studies¹. The sensor works in a similar way to screen-readers, but it stores the screen text instead of vocalising it. Specifically, we rely on Android's accessibility service, which notifies our software whenever the user interface updates. This can happen when users actively scroll on their phone, type, unlock their screen, or simply when the application updates its content.

When our software is notified of a screen update, we receive a data object that contains a tree representation for all on-screen UI elements. From this, we are able to extract the text of each UI element, the screen (x,y) coordinates of the UI element, and the precise timestamp when this was shown on the screen. Because our software is notified of all screen updates, it is often the case that although the screen has updated, the actual text has remained the same. This can lead to a large amount of duplicate data. Therefore, a filtering mechanism we introduced is to discard screen updates if the text labels remain unchanged.

Due to the way AWARE-Light operates, the data from our sensor is "flattened" into a single long string. Every time the screen updates, we generate and store this string in a database. Because the information originally comes in the form of a tree, we also store special markers in the string to reflect the delimitation of UI elements.

¹AWARE-Light can be downloaded from <https://github.com/awareframework/aware-light-client>, of which our screen text sensor is one of the available sensors.

3.2 Textual features

For this paper, we developed some further terminology to describe the data collected by this sensor. This is partly due to how the operating system provides us with the screen text data. To the best of our knowledge, there is a lack of established terminology to characterise such screen text. Hence, we propose a set of terms to describe screen text elements, which we refer to later in our analysis.

A *screen* refers to a single snapshot of the user's phone screen at any time. Each screen contains all text content present on the phone screen, which may include news articles, comment text, labels, or links in the navigation bar. The text for each screen is represented as one data entry, where it is concatenated into a single long string and delimited. A new screen is captured and stored only if its text content is different from the previous screen. Fig. 1 provides examples of screens and their contents. The operating system provides us with the Application that is associated with the screen (as a package name).

A *phrase* is a chunk of text within a screen. Defined within Android as a text node, this is the text that belongs to a UI element. A phrase can contain multiple words, and the length of a phrase can vary substantially since different applications use text differently. Phrases are visually distinct on screen, as demonstrated in Fig. 1. Our sensor does not attempt to make any inference regarding the relationship between phrases, and simply delimits them according to what the operating system provides us.

For each phrase, we record its screen (x,y) coordinates as provided by the operating system. Specifically, the coordinates are measured in pixels, and we record the top-left and bottom-right coordinates of the rectangular bounding box of the the phrase. Fig. 2 demonstrates how each captured phrase and its position can be used to reconstruct the layout of text content on a screen.

Finally, all phrases in a screen are concatenated into one text string along with their positions. Each phrase and its position is delimited using double pipes (| |) to facilitate individual analysis.

3.3 Text metrics

We use a number of metrics to quantify patterns in the text data we have collected. We consider the total number of screens as a measure of "information update". This number depends on how often participant's phones update the information they display.

We also use the number of phrases per screen as a measure of "information density", and we use it to quantify the volume of information. There are multiple ways to quantify information density, including word count or character count. However, we decided to use phrases per screen because it captures the underlying complexity of the UIs that participants interact with.

We also use the metric of *phrase difference* to quantify "information churn", or a measure of how much the screen has changed. The phrase difference between two sequential screens is defined as the set of unique phrases appearing in only one of the two screens. Therefore, given we have screen *A* with a set of phrases P_A and screen *B* with a set of phrases P_B , we can calculate the phrase difference using the following set difference formula:

$$difference = (P_A \cup P_B) - (P_A \cap P_B)$$

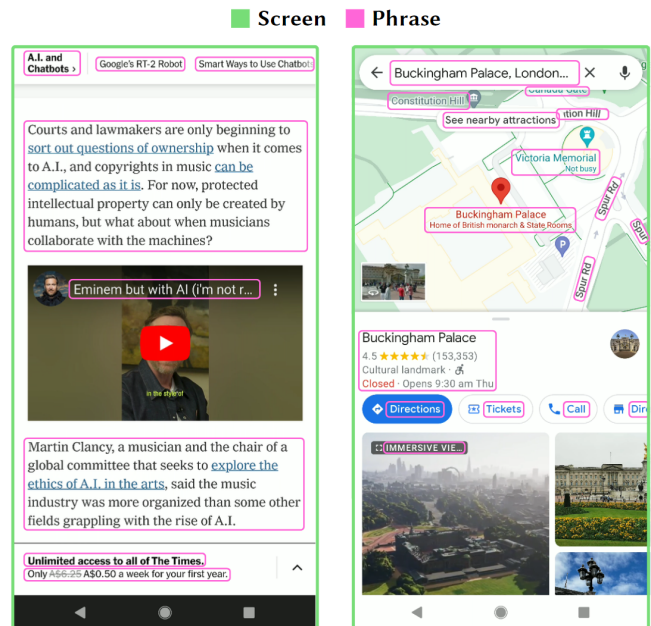


Figure 1: Examples of what a *Screen* and *Phrase* looks like on-screen. Each highlighted rectangle represents a single entity.

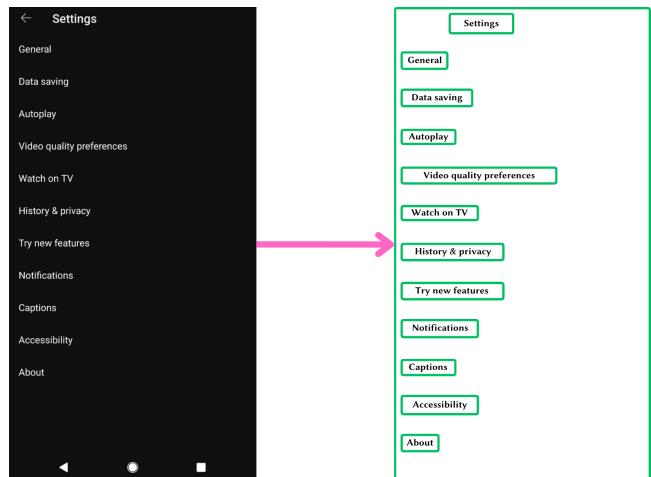


Figure 2: Reconstruction of a screen and its phrases. The original screen is shown on the left, while the reconstruction is on the right.

For example, the phrase difference between a screen with the phrases ["A", "B"] and another screen with ["B", "C", "D"] contains 3 phrases, as each phrase in the set {"A", "C", "D"} appear in only one screen. Fig. 3 compares two screens and visualises their phrase difference. We use set difference as our measure because duplicate phrases within a screen are uncommon, hence discarding duplicates will have a negligible impact on measuring overall content change between screens.

We note that the metrics we present are independent. An application may have dense or sparse screens, which may update frequently or not, and those updates can bring about small or large changes.

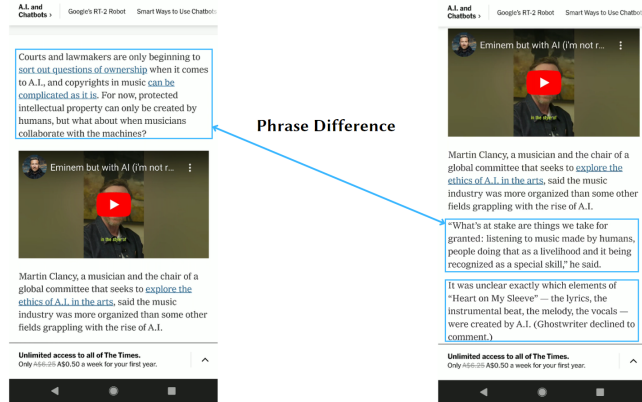


Figure 3: Illustration of how we calculate *Phrase Difference* between two screens. The highlighted phrases are the ones that differ between the two screens.

3.4 Privacy implications

Collecting data from smartphone usage comes with privacy implications given the sensitive nature of the personal user data that is being collected. The need to accommodate user privacy preferences and secure user data is particularly salient with the screen text sensor, given the highly sensitive and revealing nature of its data. To address privacy concerns, we designed our screen text sensor to be easily disabled and re-enabled at the participants' discretion. The software also displays a persistent notification that informs the user whenever data collection is taking place, ensuring that participants are aware that their phone usage is being studied. These controls give participants control over the data they choose to share and when they share it, and the constant notification of data collection reminds them to disable any sensors they do not want tracked. Additionally, we implemented Android's built-in password-detection method to identify instances where participants were entering or viewing passwords. In such cases, we do not capture the corresponding screen text data. For studies that do not require all screen text, our sensor allows researchers to specify only certain applications to collect data from and exclude others. In doing so, participants have the assurance that sensitive information from specific apps or activities will not be recorded. This selective approach to data collection not only preserves participant confidentiality but also encourages broader participation in studies, as individuals may be more inclined to contribute data when they have control over which aspects of their smartphone interactions are being captured. While these measures may result in minor data loss, they are essential for minimising participant burden and safeguarding their privacy. Beyond the privacy and security features of AWARE-Light and the screen text sensor, it is incumbent upon study administrators to implement a setup that ensures the secure transmission and storage of screen text data.

4 EVALUATION

We conducted a field study to collect a comprehensive and realistic dataset comprising smartphone screen text and sensor data. Throughout a two-week study period, we collected textual data displayed on participants' smartphone screens. We also collected a range of additional sensor data to further inform our understanding of user context, including applications, battery, Bluetooth, communication (calls and SMS), keyboard, location, network, notifications, proximity, screen state, and Wi-Fi data. Participants used their smartphones in a natural way over the two-week period. We also used Experience Sampling to understand participants' activities, and questionnaires to gauge their privacy concerns.

The study was approved by the University of Melbourne's Office of Research Ethics and Integrity. Participants were provided with a compensation of \$65 AUD upon successfully completing two weeks of data collection and a debriefing questionnaire.

4.1 Participants

We recruited 21 participants (10 male, 11 female) for our study (demographics listed in the Appendix). The participants were between the ages of 18 and 54, with a median age group of between 25 to 34 years. Participants had a highest education level of a current university student ($n = 3$) Bachelor's degree ($n = 10$), Master's degree ($n = 7$), or Doctoral degree ($n = 1$), with a median annual household income of between \$40,000 to \$49,999 AUD. Participants reported a median daily phone usage time of 3 to 4 hours. We considered only English-speaking participants who regularly use their smartphones. We recruited only participants using Android devices (given that AWARE-Light runs only on Android), with an Android version from 9 to 13. Participants used a variety of devices from different manufacturers: Samsung ($n = 10$), Google ($n = 5$), OnePlus ($n = 2$), Huawei ($n = 1$), Motorola ($n = 1$), OPPO ($n = 1$), and Xiaomi ($n = 1$).

4.2 Experience Sampling Method

To gain a deeper understanding of user context, we collected Experience Sampling Method (ESM) responses from participants. ESM involves measuring participants' behaviour, thoughts, and feelings during their day-to-day activities through short questionnaires answered throughout the day [107]. These questionnaires are completed by users in their actual environment, more closely replicating natural behaviour compared to controlled lab studies [80, 107]. The ESM questionnaire we employed aimed to capture user behaviour that may not be readily apparent in the sensor data, thus enabling a more comprehensive inference of context.

To understand this context, we asked two questions:

(1) *Where are you right now?*

Participants were presented with seven multiple-choice options (Fig. 4) and asked to select one that best represented their current location. The options included Home (Indoors), Home (Outdoors), School/Work (Indoors), School/Work (Outdoors), Outdoors (Not at home), Travelling in a vehicle, or Other. For participants who chose "Other," they were further prompted to provide a brief description of their specific location.

(2) *What are you doing right now?*

Participants provided free-text responses (Fig. 5) by describing the primary activity they were engaged in during the five minutes preceding the questionnaire. Participants were asked to provide a single sentence to capture this information.

Figure 4: ESM Interface - Question 1.

Figure 5: ESM Interface - Question 2.

We designed our questionnaire to be brief and easy to complete to minimise participant fatigue. We aimed to avoid burdening participants with frequent or lengthy responses, as such requirements can cause annoyance [45]. Given our focus on capturing context from participants' self-reports in real-time, it was important for them to spend minimal time on answering the questions to ensure

timely responses. The first question utilised predefined categories that represent common situations encountered in adults' daily lives [70]. This approach allowed participants to select from general location types without providing specific details. The second question asked participants to summarise their current activity in a single sentence. Categories were not provided for this question, as it is challenging to categorise the wide range of possible activities in which participants may be engaged.

4.3 Study procedure

Participants initially expressed their interest in the study by completing an expression of interest form. Following this, they received a Plain Language Statement that provided detailed information about the study, along with a consent form to be completed. Upon returning the consent form, participants were provided with an instruction sheet outlining the process of setting up the application on their smartphones, joining the study, and completing the ESM questionnaires. A three-day testing period followed to ensure compatibility of participants' smartphones with AWARE-Light and accurate capture of sensor data. Participants whose smartphones were incompatible due to accessibility setting issues ($n = 2$) or experienced force-closing of the application ($n = 1$) were excluded from the study. Following the testing period, 21 participants proceeded with the remainder of their two-week study period.

In addition to passive data collection, participants were prompted to complete ESM questionnaires five times a day, with two-hour intervals between 10 a.m. and 6 p.m. These questionnaires were delivered via notifications that remained visible for 15 minutes in the smartphones' notification menu. Considering that users tend to activate their phone screens within 15 minutes of inactivity [41], they were likely to encounter the questionnaire before the notification expired, even if it was received shortly after their screen was turned off. If the notification expired, participants were unable to access the questionnaire until the next scheduled time. The questionnaires were scheduled at the same time each day to capture potential changes in participant behaviour during consistent time periods across different days.

Upon completing the two-week study period, participants were asked to fill out a debriefing questionnaire. This questionnaire aimed to gather feedback on participants' experience during the study, including the ease of using the application, participants' comfort level with data collection from each sensor, reasons for turning off any sensors, and overall feedback on the study (see the Appendix for details). The study concluded with participants being reimbursed upon completion of this questionnaire.

5 RESULTS

In our study, we gathered a total of 7,004,867 screens, 135,414,272 phrases, and 4,264,737,110 characters. The most active participant viewed 1,076,069 screens and 13,941,219 phrases, while the least active participant had 25,200 screens and 158,555 phrases. On average, the time between screen updates was 0.87 seconds. The most "dense" screen recorded comprised 5,245 phrases and was 127,593 characters long. The screen text table has a total size of 12.69 GB, with each participant generating an average of 43.16 MB of screen text data per day.

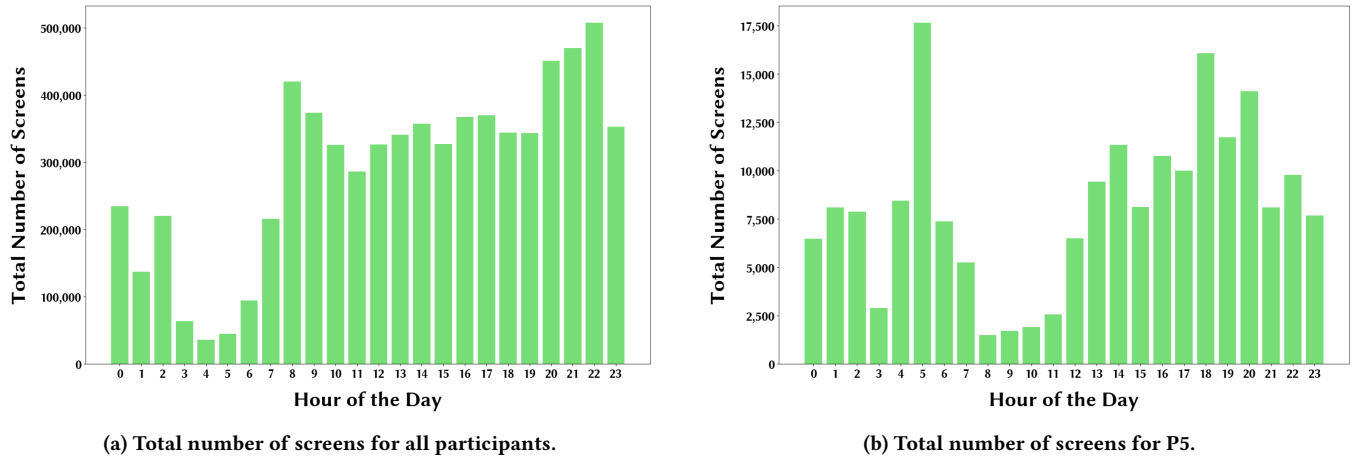


Figure 6: Total number of screens for each hour of the day.

In contrast, we collected only 32,199,892 characters in keyboard input, which represents 0.76% of the characters captured from screen text. This highlights the significantly richer interactions that can be captured within screen text compared to keystroke data.

5.1 Screen text and user behaviour

Various techniques, including NLP methods (e.g., keyword extraction, sentiment analysis, named entity recognition, topic modelling), corpus linguistics (e.g., phraseology, lexicogrammar, register, formal language usage), and statistical analyses, can be employed to analyse text. Our focus is to present the applicability of these approaches to the collected data. The presented measures are aggregated across hours and days of the week, encompassing interaction frequency (total screens), information density (phrases per screen), and text sentiment.

Firstly, examining screen text patterns over time reveals individual user characteristics and general phone usage trends. Across all participants, the lowest number of screens is captured from 3 a.m. to 5 a.m., constituting 2.06% of total screens, while the highest occurs from 8 p.m. to 10 p.m., comprising 20.37% (Fig. 6a). While this trend is generally consistent, individual behaviours may differ. For instance, P5 peaks at 5 a.m. (9.03%) and from 6 p.m. to 8 p.m. (21.46% total) but shows minimal usage during daytime from 8 a.m. to 11 a.m. (3.92% total) (Fig. 6b). Examining daily trends (Fig. 7a), Thursday consistently records a higher number of screens than other days from 1 p.m. onward, and Friday peaks at 2 a.m. Individual patterns are discernible as well; for instance, P2 exhibits alternating increases and decreases in screens viewed from morning to evening, a pattern consistent across each day (Fig. 7b).

For all participants, the average number of phrases per screen, reflecting information density, peaks at 7 a.m., averaging 27.49 phrases. This is 2.17 times higher than the lowest average at 5 a.m., which stands at 12.68 phrases (Fig. 8a). This overarching trend is further elucidated for each participant. For instance, P12 exhibits a peak at 7 a.m. with an average of 49.15 phrases, and a similar peak at 10 p.m., averaging 48.19 phrases. Interestingly, this participant

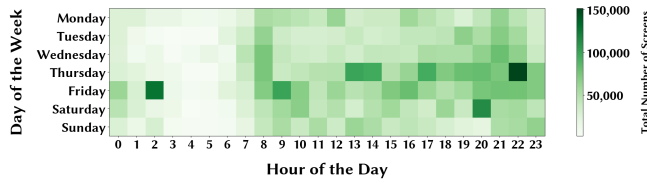
views an average of 4.87 phrases per screen from 12 a.m. to 6 a.m., constituting less than 10% of their peak at 7 a.m. (Fig. 8b). Exploring the hourly average number of phrases per screen across different days of the week reveals no noticeable similarities.

To showcase the potential for further linguistic analysis, we conduct sentiment analysis on all screen text using the VADER (Valence Aware Dictionary and sEntiment Reasoner) model [52]. After pre-processing phrases by removing stop words, we concatenate all phrases of a screen into one string, tokenize the string, and apply the VADER model. The sentiment score ranges from -1 (most extreme negative) to 1 (most extreme positive), with 0 indicating neutral sentiment. The lowest average text sentiment occurs between 12 a.m. and 5 a.m., while sentiment remains stable for the rest of the day. All average hourly sentiments are positive (Fig. 9a). Some participants deviate substantially from this overall trend; for instance, P15 engages with content of positive sentiment mostly from 3 a.m. to 10 a.m. Overall, we observe that the average sentiment generally decreases from 12 a.m. to 5 a.m. on each day, then increases at 6 a.m. and fluctuates little throughout the day (Fig. 10). This trend is consistent across each day.

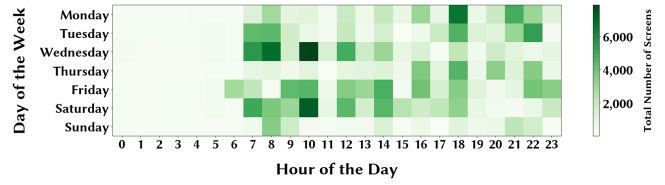
5.2 Screen text and app behaviour

Analysing screen text from various apps can help classify them based on the type and content of text they display. Using Android's package naming system, each unique package name is linked to a specific app. We employ the app categorisation dataset by Schoedel et al. [93] to group apps into categories like social media, transportation, and gaming. Apps not in this dataset are labelled as "NA." We demonstrate categorising apps based on "density" (phrases per screen) and "dynamics" (phrase difference). Additionally, we present aggregated sentiment analyses for each app, exploring commonly-viewed content and its variations between users. These analyses highlight the semantics in screen text, offering insights into app usage and user behaviour.

The average number of phrases per screen and the average phrase difference for each app show a strong positive correlation, $r(392) = .63, p < .001$ (Fig. 11a). Some apps deviate from this trend;

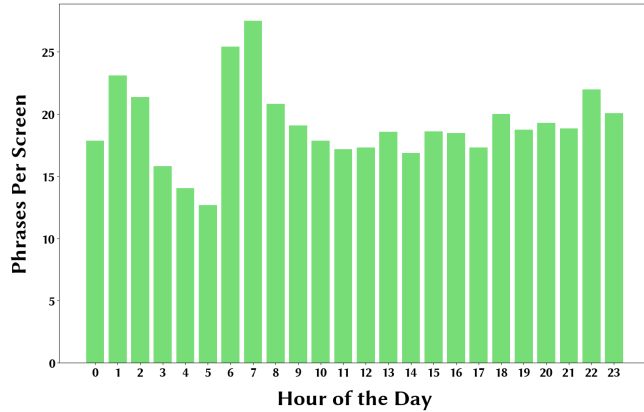


(a) Total number of screens for all participants.

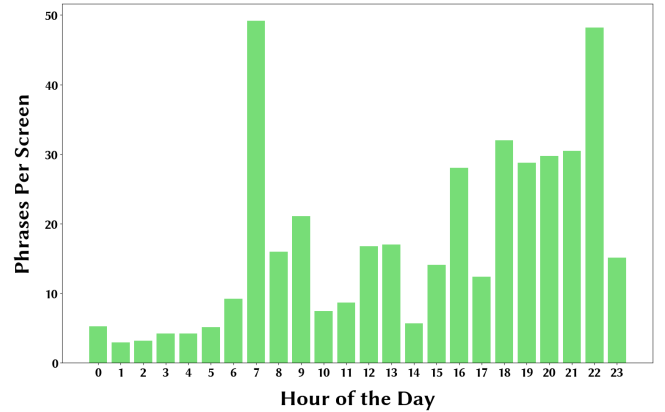


(b) Total number of screens for P2.

Figure 7: Total number of screens for each hour and day of the week.

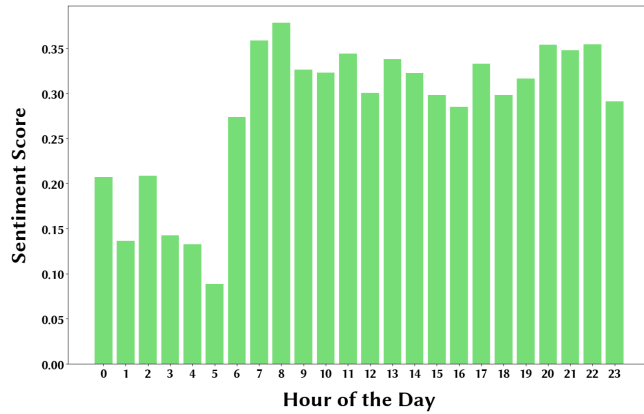


(a) Phrases per screen for each hour of the day for all participants.

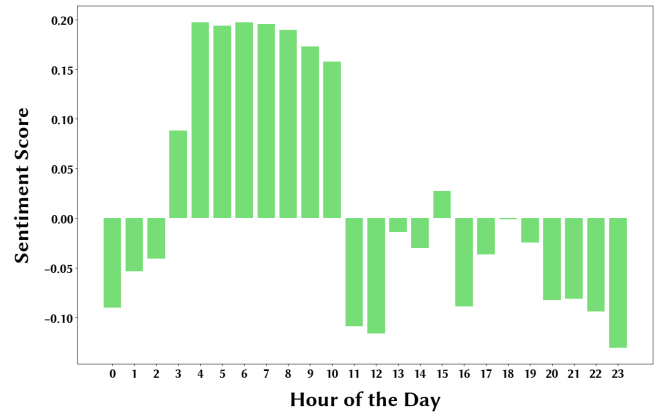


(b) Phrases per screen for P12.

Figure 8: Phrases per screen for every hour of the day.



(a) Sentiment for all participants.



(b) Sentiment for P15.

Figure 9: Sentiment for each hour of the day.

for instance, the Samsung Safety Information app has an average of 206.20 phrases per screen but only an average phrase difference of 11.68 phrases. Similarly, the Stockbit and CoinMarketCap apps, with averages of 101.11 and 75.68 phrases per screen, respectively, exhibit average phrase differences of 59.05 and 45.10 phrases. In both cases, the average phrase difference is above 50% of the average number of phrases per screen.

The average number of phrases per screen and the average phrase difference for each app category exhibit a strong positive correlation, $r(19) = .84, p < .001$ (Fig. 11b). Internet-related apps have the highest averages for both phrases per screen (92.41) and phrase difference (25.38) (not displayed in the figure for clarity). However, weather-related apps deviate from the general trend, displaying low density but highly dynamic content, with an average of 17.99 phrases per screen and a phrase difference of 17.59.

Word	Count	Word	Count	Word	Count	Word	Count
relevant	1070	cal	11348	taylor	3964	foo	3197
unit	856	km	7875	swift	3248	fighters	3197
teaching	758	distance	5473	version	1832	bon	3031
lecturer	689	set	5057	playlist	1749	jovi	3016
information	642	heart	4995	red	1628	led	2490
accounting	535	sleep	4949	loving	1174	zeppelin	2490
professional	535	hr	4811	english	980	remaster	1747
academic	535	fitbit	4793	never	927	jukebox	1594
cyber	535	rate	4162	chou	879	postmodern	1592
security	535	2023	4100	g.e.m	860	rock	1346

App: SEEK (Job-Finding)
(All Participants)

Category: Health
(All Participants)

App: Spotify (Music)
(P8)

App: Spotify (Music)
(P11)

Table 1: Top 10 words by occurrence, grouped by App and App Category.

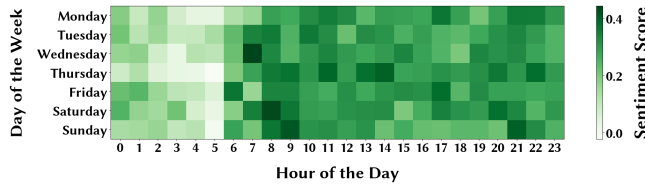


Figure 10: Sentiment for each hour and day of the week (all participants).

Based on the VADER model, a sentiment score of below -0.05 is considered negative, between -0.05 and 0.05 is considered neutral, and above 0.05 is considered positive. Out of the 394 apps that participants used, 263 apps have a positive average sentiment, 108 apps have a neutral average sentiment, and 23 apps have a negative average sentiment. The SEEK app (job finding) has the highest average sentiment, at 0.96, whereas the Samsung Safety Information app has the lowest average sentiment, at -0.95 (Fig. 12a).

Each app is grouped into one of 20 categories based on its primary functionality, with the remaining apps grouped as "NA". Out of the 21 app categories, 19 categories have a positive average sentiment, and 2 categories have a neutral average sentiment. Career-related apps have the highest average sentiment, at 0.75, while Time-related apps have the lowest average sentiment, at -0.05 (Fig. 12b).

Identifying the most frequently occurring words within each app or app category allows for content categorisation and understanding individual preferences. Table 1 displays the top 10 highest-occurring words across all participants for a specific app and app category, along with a comparison between two participants (P8 and P11) for the same app. For instance, the SEEK app, a job-searching application, presents content about teaching and lecturing jobs, as well as jobs related to accounting and cybersecurity. Health-related apps contain information related to exercise (cal, km, distance, fitbit) and cardiovascular health (heart, rate), along with sleep. P8 predominantly uses the Spotify app (music) to find Taylor Swift’s music, while P11 primarily views content related to rock music, including the Foo Fighters, Bon Jovi, and Led Zeppelin.

5.3 Screen text and other sensors

We further explore how our screen text data can be integrated with other sensor data to understand context from different perspectives.

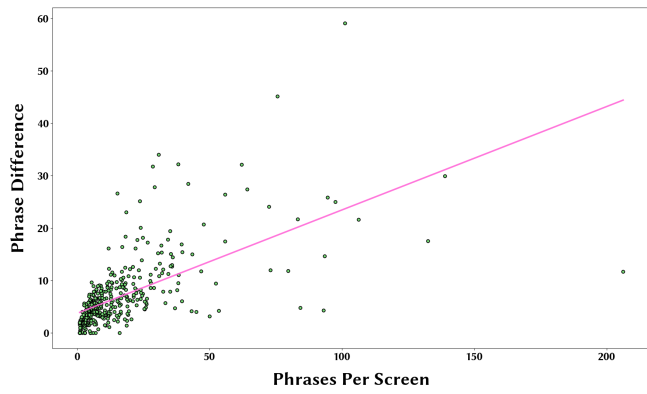
5.3.1 Screen text across geographic locations. We analyse geographic variations in screen text to uncover smartphone usage patterns in physical spaces. Using latitude and longitude data, we group locations within a 10-meter radius, reducing visual clutter while maintaining data precision. Figures (Fig. 13a, 13b, 13c) illustrate our findings.

Comparing a park (middle-bottom of the map) and a business area (left of the map), we note fewer recorded locations in the park, indicating higher activity in the business area. In the park, participants generally view neutral content (Fig. 13a). Conversely, the business area exhibits diverse sentiments, with positive clusters near the cinema and bar. Despite a higher screen count (Fig. 13b), the park sees fewer phrases per screen (Fig. 13c), suggesting frequent phone use during park traversal with minimal content consumption. In the business area, patterns are less distinct, but notable screen counts occur at the bar, restaurant, and road intersections.

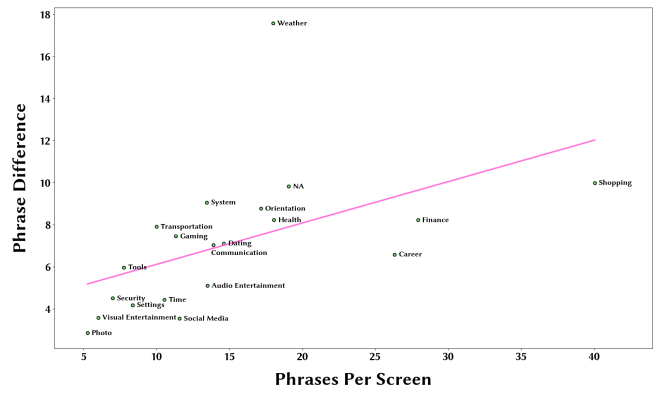
5.3.2 Screen text and connectivity. We examine the impact of social context, measured through Bluetooth and Wi-Fi connectivity, on smartphone screen content [6, 61]. Bluetooth RSSI values below -70 dBm, indicating lower social interaction [61], lead to an increase in screens viewed (111,695 screens per bin), phrases per screen (24.20), and phrase difference (12.65) compared to higher social interaction (Fig. 14a, 14b, 14c). However, no discernible differences in sentiment are observed (Fig. 14d). Similarly, weak Wi-Fi RSSI values below -70 dBm [77] result in increased screens viewed (136,240 screens per bin) and phrase difference (8.14) (Fig. 15a, 15c), but no differences in phrases per screen or sentiment (Fig. 15b, 15d). Battery levels below 15% lead to a reduction in screens viewed (39,976 screens per bin) (Fig. 16a) compared to levels above 15%, without affecting other metrics (Fig. 16b, 16c, 16d). Participants tend to minimise phone use as battery levels decrease.

5.3.3 Screen text and ESM. We analyse responses from the second question of our ESM questionnaires, categorising them into 11 categories detailed in the Appendix. For each ESM response, we retrieve the participant’s screen text from the five minutes prior to receiving the questionnaire, allowing us to study the relationship between their activity and screen content.

The highest average number of screens viewed occurs during shopping, contrasting with the lowest during online searching (Fig. 17a). In terms of the average number of phrases per screen and average phrase difference, online searching and housework exhibit

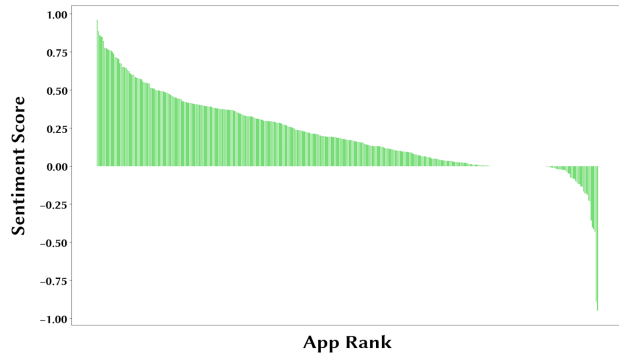


(a) All apps. Each data point represents one app.

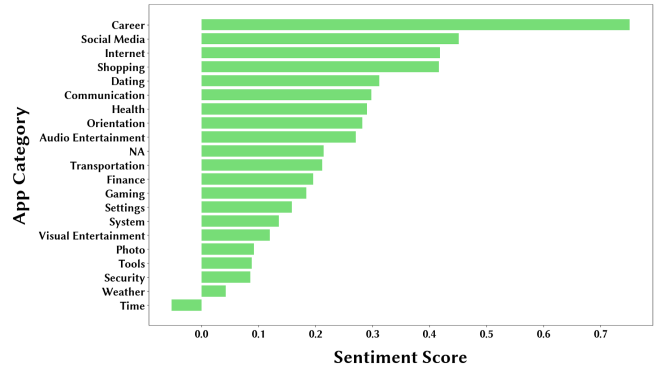


(b) All app categories. Each data point represents one app category.

Figure 11: Average number of phrases per screen vs. Average phrase difference.

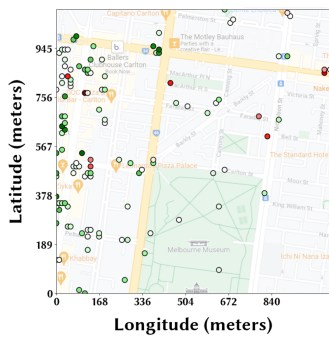


(a) All apps. Each app is represented by a single vertical bar. Apps are ranked from highest sentiment (left) to lowest sentiment (right). App names are omitted to improve legibility.

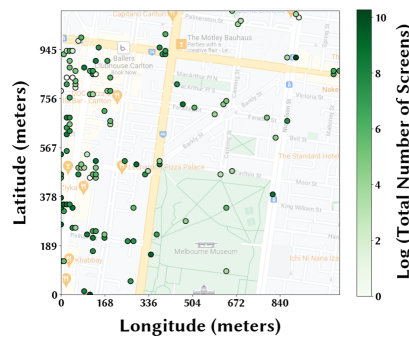


(b) All app categories. Each app is represented by a single horizontal bar

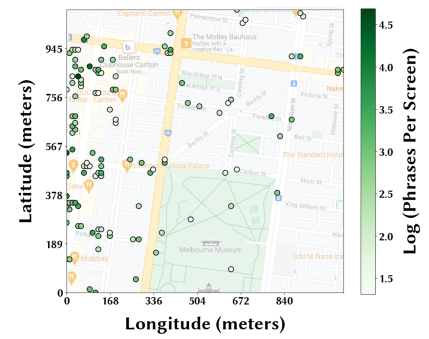
Figure 12: Average sentiment.



(a) Average sentiment.



(b) Total number of screens.



(c) Average number of phrases per screen.

Figure 13: Statistics per location in the study (selected subset).

the highest and lowest values, respectively (Fig. 17b, 17c). Regarding sentiment, it peaks during housework and exercise, while it is lowest during socialising and rest (Fig. 17d).

5.4 Self-report data

We asked participants to provide self-report feedback on how they perceived smartphone sensing using the AWARE-Light application.

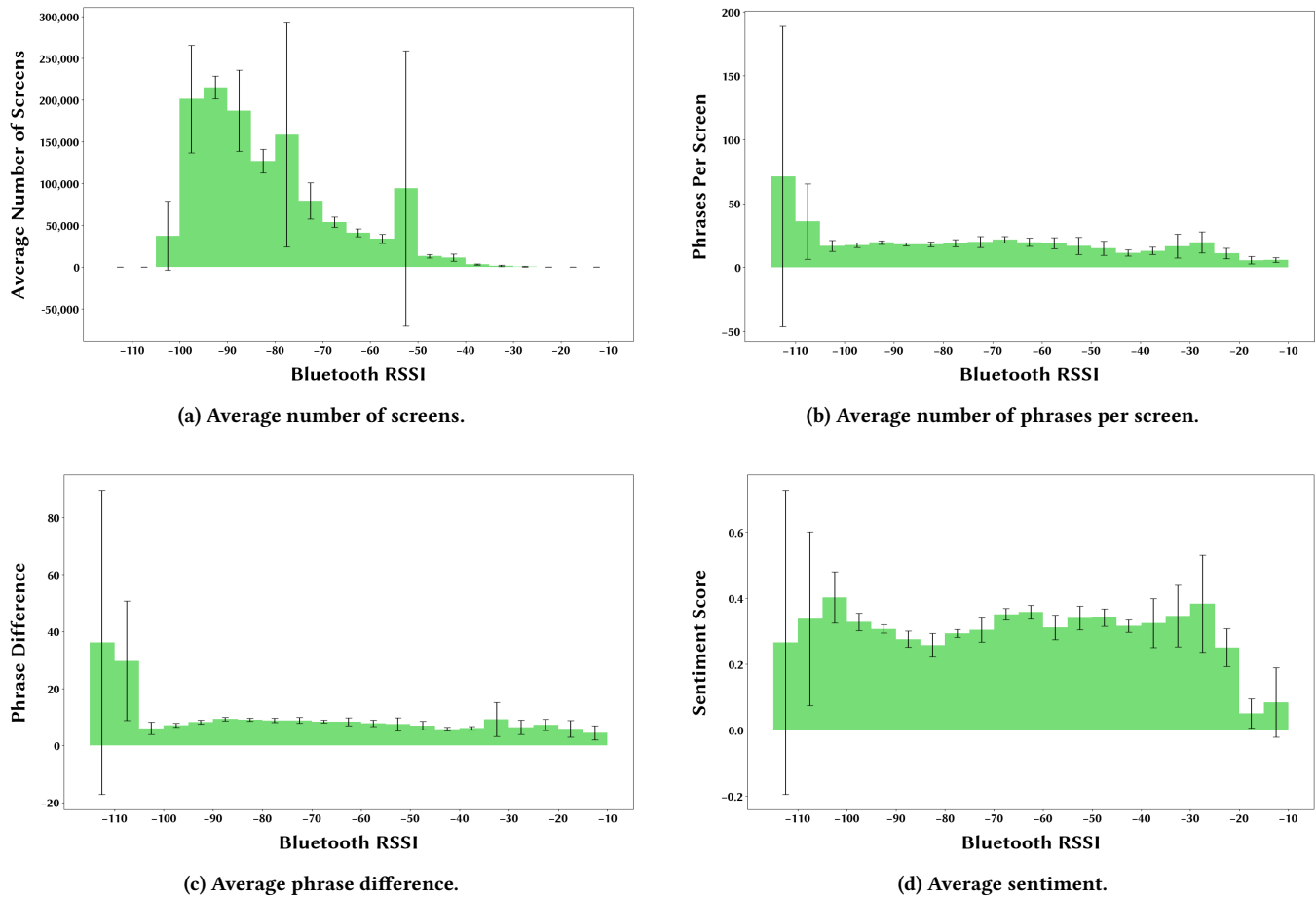


Figure 14: Average statistics per Bluetooth RSSI. Error bars denote standard error (SE).

Each question was answered on a scale from 1 (strongly disagree) to 7 (strongly agree); the details are provided in the Appendix. We find that it is generally easy for participants to use the app and they are comfortable using it. Participants also indicate that they tended to be more conscious of their smartphone use and were somewhat concerned about their privacy while sensing occurred. Given an opportunity in future studies, they generally would be interested in receiving daily feedback about their phone usage.

Table 2: Self-report of perceptions when using the AWARE-Light application.

Perception	Mean	SD
Ease of Use	5.81	1.12
Comfort of Use	5.14	1.39
Increased Consciousness of Smartphone Use	4.81	1.75
Concerns About Privacy	5.10	1.45
Daily Feedback About Smartphone Use	5.71	1.38

We also aimed to gain an understanding of how participants felt towards data collection from each sensor to see how our screen text sensor compares with existing sensors. Each question was answered on a scale from 1 (extremely uncomfortable) to 7 (extremely

comfortable); the details are provided in the Appendix. Participants were generally comfortable with the battery, screen state, Bluetooth, proximity, Wi-Fi, network, and applications sensors, as these sensors reveal little information and cannot be used to identify users. The notifications, communication, and location sensors were perceived neutrally, as they could be used in conjunction with other sensors to gather information specific to a user. Participants felt slightly uncomfortable towards the screen text and keyboard sensors, likely due to the ability for these sensors to capture direct user interactions with their phone, which may include sensitive information.

During the study, 7 participants intermittently disabled specific sensors, while the remaining 14 participants chose to retain the default settings over the entire study period. The screen text sensor was most-commonly disabled, being disabled by 4 participants for an average of 1.27 days during the study. The application and location sensors were disabled by 3 participants each, for an average of 5.08 hours and 9.33 hours, respectively. Participants were clearly told that they would not be penalised in any way if they chose to disable sensors.

Participants who disabled sensors generally did so due to privacy concerns when viewing sensitive content:

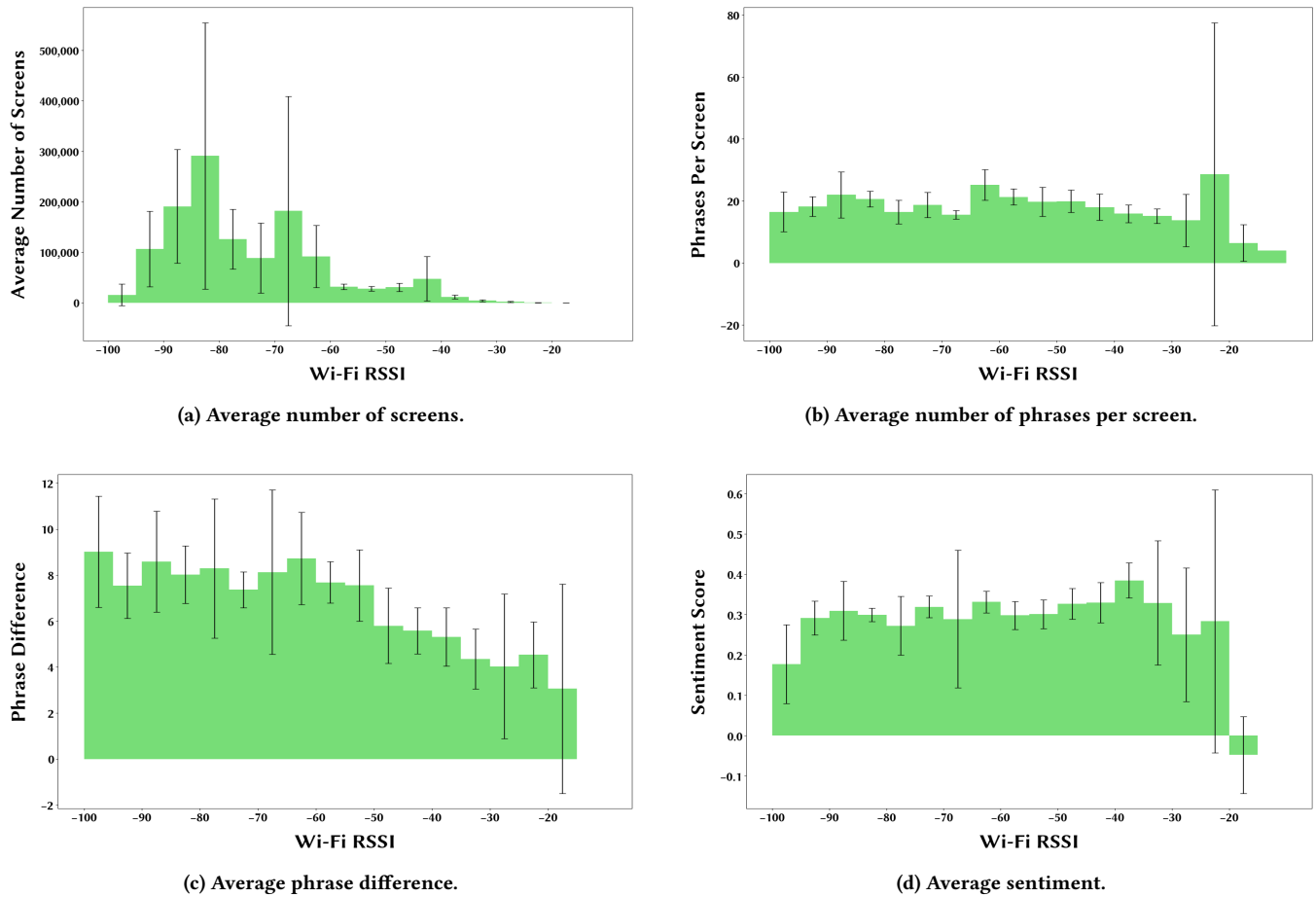


Figure 15: Average statistics per Wi-Fi RSSI. Error bars denote standard error (SE).

Table 3: Self-report of comfort for each sensor activation.

Sensor	Mean	SD
Battery	6.14	1.24
Screen State	5.95	1.07
Bluetooth	5.43	1.16
Proximity	5.38	1.53
Wi-Fi	5.24	1.70
Network	4.86	1.98
Applications	4.57	1.69
Notifications	4.19	1.75
Communication	4.14	1.68
Location	4.10	1.87
Screen Text	3.05	1.96
Keyboard	2.86	1.74

“[I disabled the] screen text [sensor] on occasion as there were sensitive information related to work and personal life that I didn’t feel comfortable sharing.” - P8

“During the time of launching certain private applications, I turned off all the sensors for a short time.” - P5

Some participants who did not disable any sensors see potential benefits of smartphone sensing, or are less concerned about privacy due to also regularly using other devices:

“I did not turn off any sensors. Initially I was a bit hesitant about the sensing, and was conscious about it when I was using my phone. However, as time went on, the sensing stood out less to me and I became more comfortable with it. I guess if there was a meaningful end result to the sensing, such as detailed analyses of my app usage or phone habits, I would not mind having my data collected in this way.” - P19

“I mostly use my phone for reading stuff on the internet and some messaging, but I also use my laptop a lot, so I didn’t feel the need to turn off phone sensors, as I could always use my laptop for communication.” - P10

We also gather general feedback from participants about the study, where they express being more aware of their smartphone use while taking part:

“In my opinion, it is a good study and encouraged me to be more aware of [my] smartphone use in everyday life.” - P15

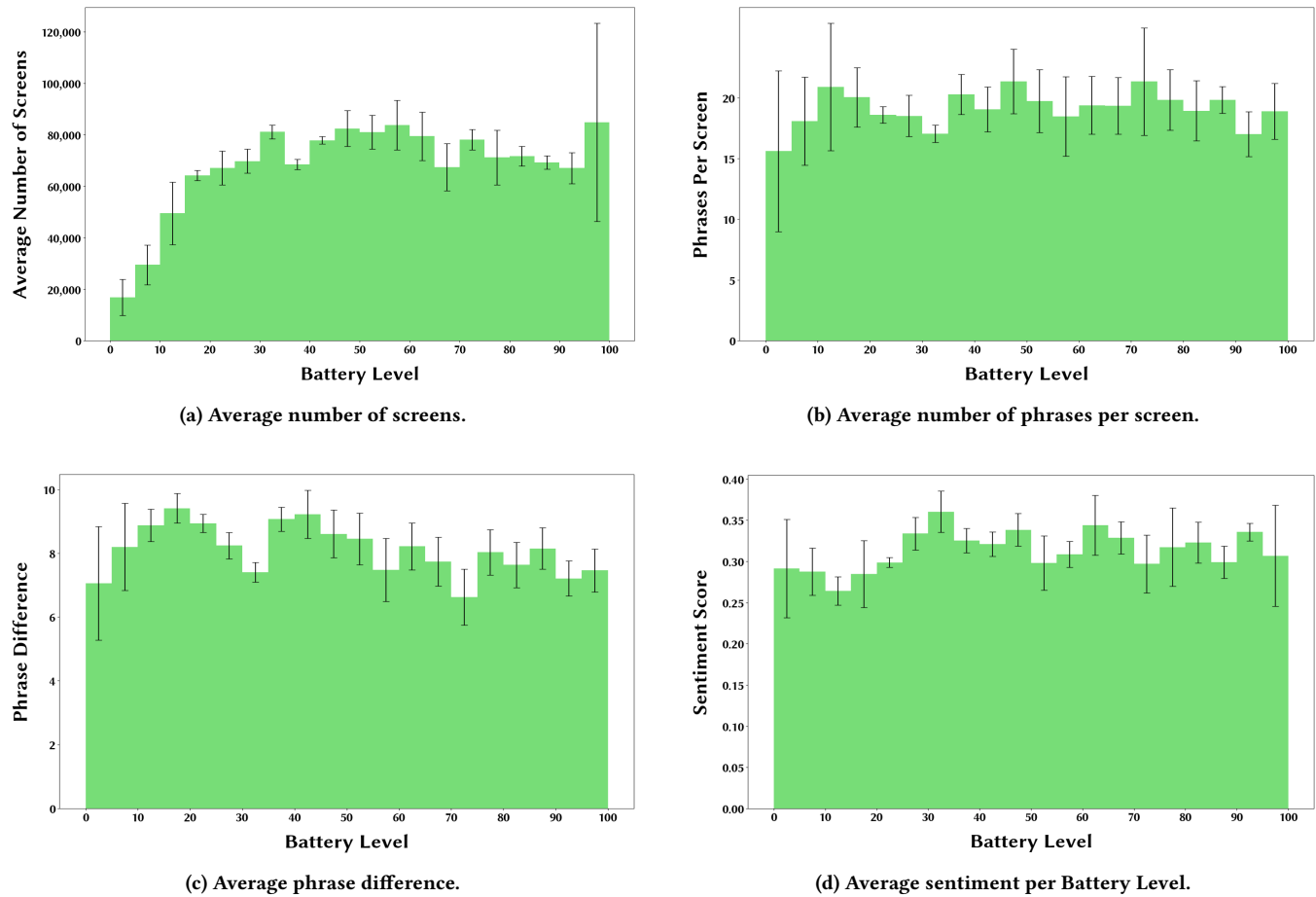


Figure 16: Average statistics per Battery level. Error bars denote standard error (SE).

"I liked it. The questions about what I was doing were interesting – made me more aware of my phone usage habits." - P19

6 DISCUSSION

We ran this study to validate our screen text sensor and understand its behaviour "in-the-wild". We collected diverse screen text data from websites and mobile apps, which we believe is accurate and representative of actual smartphone use. Our study indicates that the data collected by the sensor is timely, precise, and complete. The data appears to be collected in real-time without any observable delays. We have visualised and analysed the collected data in a number of ways. The data appear to follow the daily, weekly, and semantic patterns that we expected without any notable anomalies. Furthermore, we find that data can be reliably associated with the applications that generate it, and our analyses show that the collected text match our expectations.

In addition, we have shown that the sensor data can be reliably cross-referenced against any of the other sensor data that can be collected by AWARE-Light, including with ESM questionnaires. This opens up a fascinating range of possibilities for new types of

experiments and field studies. At the same time, we also identify a number of challenges and limitations in deploying the sensor. We fully unpack these in our discussion.

6.1 Data richness

Unlike traditional time-based phone usage or screen time trackers, which offer only a broad indication of time spent on each app, analysing screen text provides richer insights into phone usage behaviour. Time-based trackers merely measure how long the phone screen is on and which apps are displayed [44], without considering actual phone use or idle periods. Moreover, they don't capture the level of user interaction, distinguishing between passive content consumption and active information generation. Our sensor measures total phone usage and interactions with greater precision. For instance, the time difference between screens serves as a metric for understanding the frequency of phone use. The total number of screens captured reflects the interactivity between a user and their phone, as a new screen is recorded only when the screen content changes.

Existing studies primarily focus on app usage to understand user behaviour [59, 68, 128, 129]. However, user interactions within an

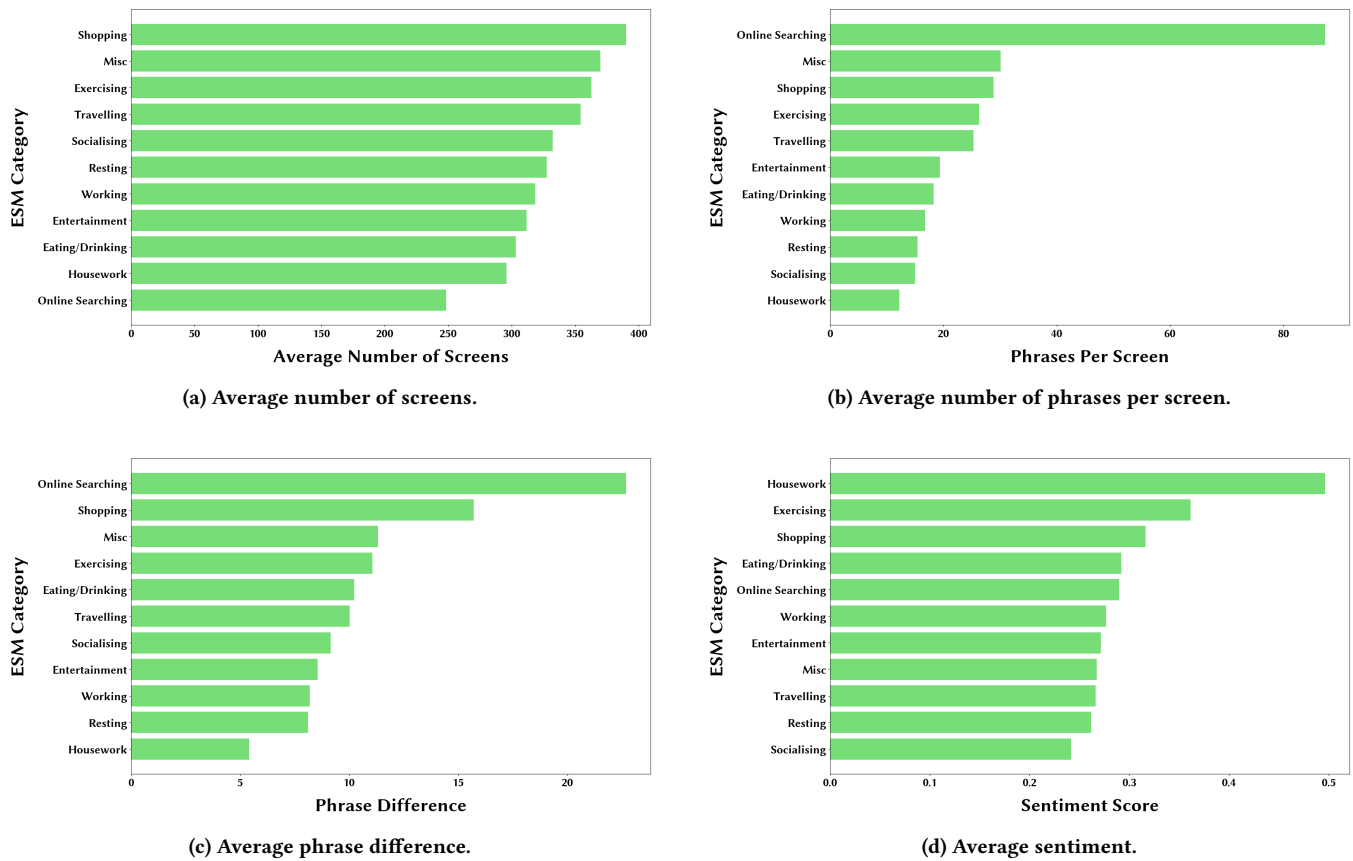


Figure 17: Average statistics per ESM category.

app can vary significantly across categories, times of the day, and users [22], making it challenging to generalise behaviour from app usage alone. Analysing phrases allows for studying semantic meaning, understanding content engagement, and identifying patterns related to information consumption. Natural language processing can extract text sentiment to summarise the overall mood of user-interacted content. The number of phrases on a screen serves as an information density measure, indicating richer content. Examining total phrases viewed provides insights into user preferences for reading longer or shorter text.

Task switching on mobile phones is often studied by analysing app switches [37, 69]. However, the purpose of each app can vary, making it challenging to understand content changes as users move between screens. Examining phrase differences between screens helps discern newly generated or retained content, offering insights into user engagement with specific topics or activities that involve multiple applications. Phrase differences can also measure content diversity and change, with larger differences suggesting a broader range of content and potentially making the app more engaging for users.

Screenomes have been used to capture a rich array of smartphone interactions [85]. By capturing a screenshot every few seconds and applying OCR, screenomes can record text and image content viewed by users. Our screen text sensor captures all text within the

screen’s UI using Android’s accessibility API, meaning that it may pick up noise due to encountering invisible text. Although OCR will not pick up invisible elements, its lower accuracy with recognising text inherently generates noise that can be difficult to identify and correct within a large dataset. Additionally, the need for screenomes to capture and perform OCR on screenshots imposes a limit on the rate of data collection and analysis, which commonly occurs in five-second intervals [82]. While the screen text sensor collects only text and does not capture images and videos, the efficiency of analysing plain text data enables real-time data collection and analysis, which can be applied to generate recommendations and interventions *in situ*. The screen text sensor can be effective in facilitating data collection and storage for studies on smartphone usage where text is the primary focus or there is a need for real-time analysis. On the other hand, studies that require capturing image and video content can be conducted using screenomes. Therefore, researchers should consider their study objectives when deciding on the most appropriate method.

6.2 Behaviour modelling

In our study, we showcase techniques for comprehending user and app behaviour on smartphones using screen text data.

Understanding user behaviour from smartphone use characteristics becomes feasible through screen text analysis. It sheds light

on digital context, capturing sentiments and content frequency throughout the day. When integrated with other movement-based sensors or wearables, users can be profiled based on both digital interactions and physical context. Passive sensing, prevalent in digital health fields [24, 98, 115, 124], allows unobtrusive data collection. Although existing technologies track physical behaviour, assessing mental well-being *in situ* remains challenging. The ubiquity of smartphone use allows screen text analysis to delve deeper into people's feelings in their natural setting. By combining screen text with other sensor data, we enhance context inference by considering multiple dimensions of behaviour. Existing sensors, such as accelerometer/gyroscope movement [84] and connectivity [26], capture physical surroundings and social context. The addition of smartphone and mobile context via screen text prompts critical questions:

- How do people's viewing habits vary across different physical locations?
- What types of content do people tend to view when commuting?
- How does proximity to others affect smartphone information consumption?

Answering these questions can advance research across fields. For instance, medical practitioners can identify patients' smartphone viewing habits and mobility information to detect anomalies [11, 90]. Understanding consumer interests is vital for marketing, facilitated by screen text. Market research can analyse locations generating searches relevant to products and services [43]. Screen text analysis in education can profile students' study habits, enabling personalised learning [39, 88].

Screen text attributes also allow inference of app behaviour. Tracing content across different apps reveals if certain apps are used together for specific tasks. For app developers, this insight can improve functionality by incorporating technologies from other apps or streamlining app flow for easier switching. Apps can be profiled based on displayed information and its rate of change. Social media apps may display digestible content with frequent updates, while government apps may offer informative content with less frequent changes. Comparing intended purposes with user interactions informs on-screen content delivery design.

Additionally, the sentiment of content within each app can inform usage patterns and user emotions. App usage between users can be examined to understand how the sentiment of content within one app can vary across individuals, representing differences in the types of social media or news viewed. Patterns in app-switching can be analysed in greater detail, including identifying how the sentiment of content viewed changes as users switch between apps. This could be beneficial in classifying whether certain apps are "complementary" in sentiment, where they are used consecutively to retain (e.g. using a "positive" app followed by another "positive" app) or alter mood (e.g. using a "negative" app followed by a "positive" app).

6.3 Further approaches to Screen Text analysis

Our analysis focuses primarily on validating our screen text sensor through understanding the properties of collected data. These

properties and screen text metrics can inform a wide range of behaviours when examined with its corresponding semantic content. For example, analysing people's reading habits involves considering how much time and how often they interact with different types of content. More engagement, measured by the number of screens viewed, suggests a higher level of interest. Additionally, the thoroughness of a person's reading can be inferred based on the information volume and variability of the text, measured by the number of phrases per screen and phrase difference, respectively, as well as how long they spend on each screen. For instance, if a user rapidly browses through numerous screens featuring a substantial volume of content and frequent changes, it may suggest a lack of interest in the topic or a more cursory reading approach. Specifically, analysing screen text allows for studying the influence of natural reading habits on factors such as academic achievement [2, 10] and social behaviour [87, 113], as opposed to within a lab setting. App developers can also leverage this insight to enhance their app design. For example, they could analyse which content types maximise user retention and encourage prolonged reading. Additionally, understanding how to strategically position content to optimise scrolling and engagement can help craft an effective and user-friendly app interface.

Nevertheless, there are many possibilities for screen text analysis that extend beyond the methods we explore. Examining screen text metadata allows us to study on-screen textual features, encompassing aspects like UI structure and text sources. These features can be extended to broader contexts, facilitating exploration of phenomena such as infinite scrolling and differentiation between passively consumed and actively typed text. The semantics of screen text can be analysed for studying how people communicate within different scenarios. This may include studying workplace relations to see how people communicate with their managers in contrast to people they manage, which could be reflected in how their tone of speech changes [55]. Properties of formal versus informal communication can also be investigated, such as analysing the differences in average text length and common topics between email and text messaging. Furthermore, people's use of language in relation to particular aspects of their work, such as their perceptions of cybersecurity, can be assessed [56].

Additionally, bi-lingual or multi-lingual analysis can also be conducted, aiming to understand if people who interact with smartphone content from multiple languages view different information or perform different tasks based on the language they are using, such as product reviews [32] or using personal assistants [117]. These findings can help to distinguish nuances in smartphone use, allowing for greater accuracy in inferring context of smartphone users. Broader questions about our use of language can also be investigated. For example, it is possible to study how the use of language varies across different age demographics [20], how language itself evolves (such as with the introduction of emojis [116]), and how ChatGPT-like services can be used in conjunction with our sensor to provide recommendations, summaries, and explanations [1].

While we have highlighted screen text metrics like phrase difference and sentiment, it is important to recognise the rich variety of metrics available. Each study requires an analysis of metrics tailored to the unique aspects of the studied phenomenon. For instance, in

studies of workplace relations, researchers may focus on the tone of speech in texting using tools like Linguistic Inquiry and Word Count (LIWC) [104]. This approach involves designing metrics that precisely capture and differentiate between the nuances of various tones, which can inform behavioural inference. Conversely, explorations of how various age groups use language may narrow down to specific elements like slang [49] or emoji analysis [60], which could reveal distinctions between generations [23]. The broadness of our screen text sensor allows for the application of a wide range of analysis techniques and metrics, providing researchers with the flexibility to align their analyses with the needs of their studies.

6.4 User perceptions

Previous work has shown that people are willing to share data captured on their smartphones with scientists who are engaged in a worthwhile cause, and often this behaviour can be considered as a donation of data [72]. Struminskaya et al. [100] found that the most common reason for unwillingness in sharing data is due to privacy and anonymity concerns. To increase transparency and privacy in mobile sensing, platforms should provide control to users over what data they choose to share [17]. Our study sheds some light on participants' perceptions of the various sensors on their smartphones, including the screen text sensor that we have developed. On average, participants felt slightly uncomfortable with their screen text being captured. Although we did not collect passwords, some participants naturally felt more alert when viewing private content, which is consistent with previous studies [99].

However, participants appreciated the ability to disable sensors when they did not feel comfortable sharing their data and re-enable them once they had completed their current task. However, we note that only one-third of the participants disabled any sensors during the study, demonstrating that they did not fully utilise this control [62]. We find that even though keyboard inputs are captured as part of screen text, people felt less comfortable on average with keyboard sensing. This could be due to the modality of interaction with each of these sensors influencing perception. For the keyboard sensor, "tracking" the action of typing on the keyboard may induce more caution for users as they are actively producing content. In contrast, viewing text on the phone screen may not necessarily carry this connotation, as users are generally engaging in a passive state of interaction.

The open-source nature of our sensor allows researchers to customise the tool based on their needs. We have demonstrated password-detection as one method for enhancing privacy, though there are many more that could be implemented using similar techniques based on the research context and the precision of data required. We emphasise that the range of collected data can vary based on the studied phenomenon, and often does not need to encompass all screen text. We have earlier mentioned the ability to configure AWARE-Light studies such that screen text data collection can be confined to nominated apps (or alternatively certain apps can be excluded). Beyond this configuration option, one could modify the AWARE-Light code such that filters are added, whereby only certain screen text content is collected and saved to the database. For example, studies related to fitness may capture only screen text containing keywords associated with fitness [108], which could

indicate the frequency of which users engage in physical activity. Research on understanding user behaviour on Twitter could limit data collection to only the Twitter app. Despite this restriction, the screen text sensor enables the capture of various interactions, including creating, viewing, or re-posting posts. This capability enhances existing social media studies, enabling a more thorough investigation of user activity to understand individual behaviour, rather than solely analysing behaviour across all social media users [126]. Additionally, studies that do not require participant identification can further remove information from screen text data such as email addresses, phone numbers, and people names as an additional security measure that can alleviate participant concerns [14].

6.5 Lessons learned

Our study has provided valuable insights into the implementation and considerations associated with our novel sensor, and we aim to share key learnings for researchers intending to employ this sensor in their studies. Many of these insights were gained through extensive trial-and-error and technical troubleshooting, highlighting the need for careful consideration to ensure smooth and uninterrupted data collection.

Firstly, researchers should be mindful of the text variability displayed on smartphones, which can be quite diverse. One consideration is the database table configuration for storing screen text. We recommend using the MEDIUMTEXT or LONGTEXT data types for the screen text column due to potential encounters with large volumes of text, exceeding the default TEXT data type limit. This issue is particularly relevant when dealing with hidden text on web pages, where the 65,535-character limit may be exceeded, leading to data transfer blockages and preventing further data uploading to the database until this data is discarded. Utilising the MEDIUMTEXT (16,777,215-character limit) or LONGTEXT (4,294,967,295-character limit) types can address this limitation.

Additionally, the database table for storing screen text data should be configured using the utf8mb4 character set (4-Byte UTF-8 Unicode Encoding) to enable the capture of emojis, which is not possible with utf8mb3. For longitudinal studies or those with large sample sizes, researchers must consider database storage capacity. Our study, spanning two weeks and 21 participants, accumulated 12.69 GB in screen text data, underscoring the potential for substantial storage requirements in larger and longer-term sensing scenarios. It is advisable to calculate the total storage needed based on the study size and duration, allocating additional storage to accommodate potential outliers.

Given that our sensor captures all screen text, researchers should be aware of the presence of non-printable ASCII characters, which may be retained using their ASCII codes (e.g. the null character is often represented as \x00). These characters, if unidentified, can cause errors during data analysis or when imported into a code editor. Precautions should be taken during the data analysis phase to screen for such anomalies and sanitise the data accordingly.

To address participant discomfort regarding screen text collection, AWARE-Light provides researchers with the flexibility to configure which applications are included or excluded when capturing screen text. For instance, a study focusing on web browsing

behaviour may choose to collect data solely from web browsers, excluding sensitive apps like messaging and banking apps. This selective approach can enhance participant comfort during the study.

Lastly, we acknowledge that our sensor may not be universally compatible with every Android version or smartphone make. The reliance on Android's accessibility services necessitates participant permission, but certain phones may automatically disable these services after a brief period. Although we instructed participants to disable battery optimisation, which is a primary cause for system-disabling of accessibility services, some phones do not allow for prolonged use of accessibility by third-party applications. To address this, we implemented a "trial" phase for potential participants, allowing us to monitor data uploads over a few days to confirm compatibility. This helps identify participants who may not upload data or upload incomplete data due to compatibility issues.

6.6 Limitations and Future Work

Our sensor only collects screen text provided it is available to the operating system as raw text. Unlike previous work, our sensor does not detect text within images or graphics. This potentially limits our data collection, but at the same time ensures that our sensor can be easily deployed to large numbers of participants without computational consequences.

Furthermore, our screen text sensor uses Android's accessibility services to capture text from phone screens. Therefore, we may capture hidden text within certain apps, mostly Internet browsing. For example, websites sometimes use hidden text either for accessibility purposes, or as a search engine optimisation strategy. Common methods for storing hidden text include positioning text beyond the screen limits of the page, matching the text colour with the background colour, or using a text font size of zero. Although they are invisible to users, the Android system still detects and stores these texts as they exist as on-screen text nodes. This may be misleading for analysis if a screen has large volumes of hidden text. Therefore, further work can be done to build on our screen text sensor to distinguish between visible and invisible text on each screen, such as inspecting each text node for properties influencing visibility. Similarly, we collect the rectangular boundary of each text node as its position. However, in some cases, it may be unclear where the text is located within the boundary box. Although all text is located within its boundaries, various styling such as centering, font sizes, and padding means that the text may not cover the entirety of its boundary box. Given that the boundaries of text nodes may overlap, the positions of some text may not be precisely inferred.

Additionally, some participants expressed that they changed their behaviour during the study [75]. We note that participants may have regulated their behaviour in a manner that appeared more socially desirable [123], meaning that we may not have captured the true behaviour of all participants in a natural setting. To increase participant privacy, we provided participants with control over enabling and disabling sensors. In doing so, participants who alter their behaviour or frequently disable sensors inherently reduce the ecological validity of their data, which our tool cannot prevent. To facilitate understanding of the reasons behind sensor

deactivation, AWARE-Light logs instances when users enable or disable a sensor. This not only adds another valuable data point for analysis but also allows for exploring associations with other data. For instance, it can help identify patterns of behaviour preceding the users' decision to disable the screen text sensor or correlate such actions with changes observed in other sensors. Gaining further insights into why the screen text sensor is disabled can contribute to enhancing its design and better addressing privacy concerns. To mitigate behaviours that reduce ecological validity, studies employing our sensor should tailor the scope of data collection to the specific phenomena under investigation. By collecting only relevant and necessary information, clearly communicating the study's objectives to participants, and ensuring transparency in data collection methods, participants can feel more comfortable in their involvement, contributing to a more accurate representation of participants' natural behaviours.

Whilst AWARE-Light has been developed with security practices in mind, ultimately safeguards for the screen text data collected are the responsibility of those using this privacy-sensitive tool, both those who configure the data collection setups and those smartphone users with AWARE-Light installed on their phone. For those responsible for configuring AWARE-Light instances, it is essential to set up databases so that they receive data from AWARE-Light over SSL-secured transmissions. Furthermore, at-rest encryption of the stored data is another security choice. For smartphone users with AWARE-Light installed and configured to collect screen text data, they must be clearly informed of the nature of this sensor (including which apps screen text is being collected from), and clearly informed of the option to disable the sensor if/when they do not want such data collected from their phone.

Overall, our study had a small sample size of 21 participants, with a majority being students and staff of our university. Therefore, the screen text we collected may largely contain themes that are representative of people working in a tertiary education environment. As we are validating our screen text sensor, this may have reduced the scope of our analysis ideas. Collecting data from a larger and more diverse cohort can enable more reliable comparisons between participant groups and hypothesis testing. However, we believe that the size and demographic of our sample has little effect on how we validate our sensor's technological capabilities.

7 CONCLUSION

We have presented a study to validate our screen text sensor that gathers text from smartphone interactions. Previous work has collected text from smartphones by either logging keyboard interactions, or taking intermittent screenshots and performing optical character recognition. Our work overcomes many of these limitations by collecting screen text continuously, unobtrusively, and without the need for computer vision processing.

In this paper, we have validated the technology and capabilities of our screen text sensor in a field study with 21 participants over two weeks. We presented examples of analyses that can be conducted using screen text data, and we suggested multiple possible directions for further exploration. Given the ubiquity of screen text data, it can be effectively applied across various disciplines to investigate specific behaviours according to researchers' needs.

Hence, a wealth of opportunities exist for exploring the potential applications of screen text across different domains.

ACKNOWLEDGMENTS

Development of the screen text sensor was supported by a grant from the Cisco University Research Program Fund and the Silicon Valley Community Foundation (ID: 2021-232746 3696), and with help from Tianyi Zhang. Additionally, this work is partially funded by NHMRC grants 1170937 and 2004316 and AUSMURI grant 13203896.

REFERENCES

- Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligent Systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. <https://doi.org/10.1145/3173574.3174156>
- Nisar Abid, Sarfraz Aslam, Abdulrahman A. Alghamdi, and Tribhuwan Kumar. 2023. Relationships among students' reading habits, study skills, and academic achievement in English at the secondary level. *Frontiers in Psychology* 14 (Jan. 2023). <https://doi.org/10.3389/fpsyg.2023.1020269>
- Hyung J. Ahn. 2023. The Influence of Demographic Factors on Word Usage in Texts Written About Artworks. *Empirical Studies of the Arts* 41, 2 (Feb. 2023), 392–410. <https://doi.org/10.1177/02762374231159999>
- Elgin Akpinar, Yeliz Yeşilada, and Pinar Karagöz. 2023. Effect of Context on Smartphone Users' Typing Performance in the Wild. *ACM Transactions on Computer-Human Interaction* 30, 3 (June 2023), 1–44. <https://doi.org/10.1145/3577013>
- Andry Alamsyah, Sri Widiyanesti, Rizqy Dwi Putra, and Puspita Kencana Sari. 2020. Personality Measurement Design for Ontology Based Platform using Social Media Text. *Advances in Science, Technology and Engineering Systems Journal* 5, 3 (2020), 100–107. <https://doi.org/10.25046/aj050313>
- Yair Amichai-Hamburger and Zack Hayat. 2011. The impact of the Internet on the social lives of users: A representative sample from 13 countries. *Computers in Human Behavior* 27, 1 (Jan. 2011), 585–589. <https://doi.org/10.1016/j.chb.2010.10.009>
- Muhammad Arif, Mohsin Bilal, Ahmed Kattan, and S. Iqbal Ahmed. 2014. Better Physical Activity Classification using Smartphone Acceleration Sensor. *Journal of Medical Systems* 38, 9 (July 2014). <https://doi.org/10.1007/s10916-014-0095-0>
- Hania Aslam, Hamid Mukhtar, Farhana Seemi, and Djamel Belaid. 2016. Harnessing Smartphones as a Personal Informatics Tool towards Self-Awareness and Behavior Improvement. In *2016 IEEE 14th Intl Conf on Dependable, Autonomous and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. IEEE. <https://doi.org/10.1109/dase-picom-datacom-cyberscitech.2016.92>
- Joost Asselbergs, Jeroen Ruwaard, Michal Ejdy, Niels Schrader, Marit Sijbrandij, and Heleen Riper. 2016. Mobile Phone-Based Nonobtrusive Ecological Momentary Assessment of Day-to-Day Mood: An Exploratory Study. *Journal of Medical Internet Research* 18, 3 (March 2016), e72. <https://doi.org/10.2196/jmir.5505>
- Sasikala Balan, Josephine Esther Katenga, and Amanda Simon. 2019. Reading Habits and Their Influence on Academic Achievement Among Students at Asia Pacific International University. *Abstract Proceedings International Scholars Conference* 7, 1 (Dec. 2019), 1490–1516. <https://doi.org/10.35974/isc.v7i1.928>
- Ian Barnett, John Torous, Patrick Staples, Luis Sandoval, Matcheri Keshavan, and Jukka-Pekka Onnela. 2018. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology* 43, 8 (Feb. 2018), 1660–1666. <https://doi.org/10.1038/s41386-018-0030-z>
- Ghassan F. Bati and Vivek K. Singh. 2017. Are you altruistic? Your mobile phone could tell. In *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE. <https://doi.org/10.1109/uic-atc.2017.8397659>
- Ghassan F. Bati and Vivek K. Singh. 2018. "Trust Us". In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. <https://doi.org/10.1145/3173574.3173904>
- Felix Beierle, Vinh Thuy Tran, Mathias Allemand, Patrick Neff, Winfried Schlee, Thomas Probst, Johannes Zimmermann, and Rüdiger Pryss. 2019. What data are smartphone users willing to share with researchers? *Journal of Ambient Intelligence and Humanized Computing* 11, 6 (June 2019), 2277–2289. <https://doi.org/10.1007/s12652-019-01355-6>
- Ghazaleh Beigi, Kai Shu, Ruocheng Guo, Suhang Wang, and Huan Liu. 2019. Privacy Preserving Text Representation Learning. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media*. ACM. <https://doi.org/10.1145/3342220.3344925>
- Till Beiwinkel, Sally Kindermann, Andreas Maier, Christopher Kerl, Jörn Mook, Guido Barbian, and Wulf Rössler. 2016. Using Smartphones to Monitor Bipolar Disorder Symptoms: A Pilot Study. *JMIR Mental Health* 3, 1 (Jan. 2016), e2. <https://doi.org/10.2196/mental.4560>
- Florian Bemmman, Maximiliane Windl, Jonas Erbe, Sven Mayer, and Heinrich Hussmann. 2022. The Influence of Transparency and Control on the Willingness of Data Sharing in Adaptive Mobile Apps. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (Sept. 2022), 1–26. <https://doi.org/10.1145/3546724>
- Dror Ben-Zeev, Emily A. Scherer, Rui Wang, Haiyi Xie, and Andrew T. Campbell. 2015. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric Rehabilitation Journal* 38, 3 (Sept. 2015), 218–226. <https://doi.org/10.1037/prj0000130>
- Sofian Berrouguet, David Ramirez, Maria Luisa Barrigón, Pablo Moreno-Muñoz, Rodrigo Carmona Camacho, Enrique Baca-García, and Antonio Artés-Rodríguez. 2018. Combining Continuous Smartphone Native Sensors Data Capture and Unsupervised Data Mining Techniques for Behavioral Changes Detection: A Case Series of the Evidence-Based Behavior (eB2) Study. *JMIR mHealth and uHealth* 6, 12 (Dec. 2018), e197. <https://doi.org/10.2196/mhealth.9472>
- Su Lin Blodgett, Q. Vera Liao, Alexandra Olteanu, Rada Mihalcea, Michael Muller, Morgan Klaus Scheuerman, Chenhao Tan, and Qian Yang. 2022. Responsible Language Technologies: Foreseeing and Mitigating Harms. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM. <https://doi.org/10.1145/3491101.3516502>
- Mehdi Boukhechba, Alexander R. Daros, Karl Fua, Philip I. Chow, Bethany A. Teachman, and Laura E. Barnes. 2018. DemonicSalmon: Monitoring mental health and social interactions of college students using smartphones. *Smart Health* 9–10 (Dec. 2018), 192–203. <https://doi.org/10.1016/j.smhl.2018.07.005>
- Miriam Brinberg, Nilam Ram, Xiao Yang, Mu-Jung Cho, S. Shyam Sundar, Thomas N. Robinson, and Byron Reeves. 2021. The idiosyncrasies of everyday digital lives: Using the Human Screenome Project to study user behavior on smartphones. *Computers in Human Behavior* 114 (Jan. 2021), 106570. <https://doi.org/10.1016/j.chb.2020.106570>
- Ashley Broadwater. 2023. *Do You Text Your Grandkids? Read This Before Accidentally Making a Big Mistake*. Retrieved Dec, 2023 from https://www.huffpost.com/entry/texting-generational-differences_1_63f8db81e4b0cab1fa2eaa16
- Prabin Byanjankar, Anubhuti Poudyal, Brandon A Kohrt, Sujen Man Maharjan, Ashley Hagaman, and Alastair van Heerden. 2020. Utilizing passive sensing data to provide personalized psychological care in low-resource settings. *Gates Open Research* 4 (Aug. 2020), 118. <https://doi.org/10.12688/gatesopenres.13117.1>
- Michelle L Byrne, Monika N Lind, Sarah R Horn, Kathryn L Mills, Benjamin W Nelson, Melissa L Barnes, George M Slavich, and Nicholas B Allen. 2021. Using mobile sensing data to assess stress: Associations with perceived and lifetime stress, mental health, sleep, and inflammation. *DIGITAL HEALTH* 7 (Jan. 2021), 205520762110372. <https://doi.org/10.1177/20552076211037227>
- Mattia Giovanni Campana and Franca Delmastro. 2022. On-device modeling of user's social context and familiar places from smartphone-embedded sensor data. (2022). <https://doi.org/10.48550/ARXIV.2205.08790>
- Y. K. Chan and Rance P. L. Lee. 2006. Network Size, Social Support and Happiness in Later Life: A Comparative Study of Beijing and Hong Kong. *Journal of Happiness Studies* 7, 1 (March 2006), 87–112. <https://doi.org/10.1007/s10902-005-1915-1>
- Agnese Chiatti, Mu Jung Cho, Anupriya Gagneja, Xiao Yang, Miriam Brinberg, Katie Roehrick, Sagnik Ray Choudhury, Nilam Ram, Byron Reeves, and C. Lee Giles. 2018. Text extraction and retrieval from smartphone screenshots. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. ACM. <https://doi.org/10.1145/3167132.3167236>
- Matteo Ciman, Katarzyna Wac, and Ombretta Gaggi. 2015. iSenseStress: Assessing stress through human-smartphone interaction analysis. In *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*. ICST. <https://doi.org/10.4108/icst.pervasivehealth.2015.259280>
- Desirée Colombo, Javier Fernández-Álvarez, Carlos Suso-Ribera, Pietro Cipresso, Hristo Valev, Tim Leufkens, Corina Sas, Azucena Garcia-Palacios, Giuseppe Riva, and Cristina Botella. 2020. The need for change: Understanding emotion regulation antecedents and consequences using ecological momentary assessment. *Emotion* 20, 1 (Feb. 2020), 30–36. <https://doi.org/10.1037/emo0000671>
- Marios Constantinides, Jonas Busk, Aleksandar Matic, Maria Faurholt-Jepsen, Lars Vedel Kessing, and Jakob E. Bardram. 2018. Personalized versus Generic Mood Prediction Models in Bipolar Disorder. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp '18)*. ACM. <https://doi.org/10.1145/3267305.3267536>
- Diego Moreira da Rosa and Lucas Pinheiro Pons. 2017. Evaluating the presentation of user reviews in multiple languages. In *Proceedings of the 8th Latin American Conference on Human-Computer Interaction*. ACM. <https://doi.org/10.1145/3151470.3151472>

- [33] David D'Acunto, Annamaria Tuan, Daniele Dalli, Giampaolo Viglia, and Fevzi Okumus. 2020. Do consumers care about CSR in their online reviews? An empirical analysis. *International Journal of Hospitality Management* 85 (Feb. 2020), 102342. <https://doi.org/10.1016/j.ijhm.2019.102342>
- [34] Alex W DaSilva, Jeremy F Huckins, Rui Wang, Weichen Wang, Dylan D Wagner, and Andrew T Campbell. 2019. Correlates of Stress in the College Environment Uncovered by the Application of Penalized Generalized Estimating Equations to Mobile Sensing Data. *JMIR mHealth and uHealth* 7, 3 (March 2019), e12084. <https://doi.org/10.2196/12084>
- [35] Michael Del Rosario, Stephen Redmond, and Nigel Lovell. 2015. Tracking the Evolution of Smartphone Sensing for Monitoring Human Movement. *Sensors* 15, 8 (July 2015), 18901–18933. <https://doi.org/10.3390/s150818901>
- [36] Orianna DeMasi and Benjamin Recht. 2017. A step towards quantifying when an algorithm can and cannot predict an individual's wellbeing. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM. <https://doi.org/10.1145/3123024.3125609>
- [37] Tao Deng, Shaheen Kanthawala, Jingbo Meng, Wei Peng, Anastasia Kononova, Qi Hao, Qinshao Zhang, and Prabu David. 2018. Measuring smartphone usage and task switching with log tracking and self-reports. *Mobile Media & Communication* 7, 1 (April 2018), 3–23. <https://doi.org/10.1177/2050157918761491>
- [38] Lenin Erazo-Garzon, Andres Patino, Priscila Cedillo, and Alexandra Bermeo. 2019. CALMS: A Context-Aware Learning Mobile System Based on Ontologies. In *2019 Sixth International Conference on eDemocracy and eGovernment (ICEDEG)*. IEEE. <https://doi.org/10.1109/icedeg.2019.8734423>
- [39] Syeda Shazmina Fayyaz and Adeel Akram. 2022. Students' perception of smartphones as a new modality to enhance health sciences knowledge. *Journal of Social & Health Sciences* (Dec. 2022), 13–21. <https://doi.org/10.58398/0001.000003>
- [40] Raihana Ferdous, Venet Osmani, and Oscar Mayora. 2015. Smartphone app usage as a predictor of perceived stress levels at workplace. In *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*. ICST. <https://doi.org/10.4108/icst.pervasivehealth.2015.260192>
- [41] Denzil Ferreira, Jorge Goncalves, Vassilis Kostakos, Louise Barkhuus, and Anind K. Dey. 2014. Contextual experience sampling of mobile application micro-usage. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM. <https://doi.org/10.1145/2628363.2628367>
- [42] Denzil Ferreira, Vassilis Kostakos, and Anind K. Dey. 2015. AWARE: Mobile Context Instrumentation Framework. *Frontiers in ICT* 2 (April 2015). <https://doi.org/10.3389/fict.2015.00006>
- [43] Anindya Ghose and Vilma Todri. 2015. Towards a Digital Attribution Model: Measuring the Impact of Display Advertising on Online Consumer Behavior. *SSRN Electronic Journal* (2015). <https://doi.org/10.2139/ssrn.2672090>
- [44] Google. 2023. *Manage how you spend time on your Android phone with Digital Wellbeing*. Retrieved Aug, 2023 from <https://support.google.com/android/answer/9346420>
- [45] GOV.UK. 2020. *Guidance Ecological Momentary Assessment Using an Ecological Momentary Assessment to Evaluate Your Digital Health Product*. Retrieved Jul, 2023 from <https://www.gov.uk/guidance/ecological-momentary-assessment>
- [46] Fei Gu, Jianwei Niu, Zhenxue He, and Xin Jin. 2017. FamilyPal: An effective system for detecting family activities based on smartphone. In *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*. IEEE. <https://doi.org/10.1109/indin.2017.8104763>
- [47] Samuli Hemminki, Petteri Nurmi, and Sasu Tarkoma. 2013. Accelerometer-based transportation mode detection on smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM. <https://doi.org/10.1145/2517351.2517367>
- [48] Aleide Hoesjmakers, Giovanni Licitra, Kim Meijer, Ka-Hoo Lam, Pam Molenaar, Eva Strijbis, and Joep Killestein. 2023. Disease severity classification using passively collected smartphone-based keystroke dynamics within multiple sclerosis. *Scientific Reports* 13, 1 (Feb. 2023). <https://doi.org/10.1038/s41598-023-28990-6>
- [49] Thomas Holtgraves. 2011. Text messaging, personality, and the social context. *Journal of Research in Personality* 45, 1 (Feb. 2011), 92–99. <https://doi.org/10.1016/j.jrp.2010.11.015>
- [50] Ke Huang, Xiang Ding, Jing Xu, Guanling Chen, and Wei Ding. 2015. Monitoring Sleep and Detecting Irregular Nights through Unconstrained Smartphone Sensing. In *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Automatic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*. IEEE. <https://doi.org/10.1109/uic-atc-scalcom-cbdcom-iop.2015.30>
- [51] Yu Huang, Haoyi Xiong, Kevin Leach, Yuyan Zhang, Philip Chow, Karl Fua, Bethany A. Teachman, and Laura E. Barnes. 2016. Assessing social anxiety using gps trajectories and point-of-interest data. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM. <https://doi.org/10.1145/2971648.2971761>
- [52] C. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media* 8, 1 (May 2014), 216–225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- [53] Seungjae Hyun, Yeonjin Lee, and Sangshin Park. 2022. No travel worsens depression: reciprocal relationship between travel and depression among older adults. *Annals of General Psychiatry* 21, 1 (Aug. 2022). <https://doi.org/10.1186/s12991-022-00405-2>
- [54] Ozlem Durmaz Incel and Atay Ozgovde. 2018. ARService: A Smartphone based Crowd-Sourced Data Collection and Activity Recognition Framework. *Procedia Computer Science* 130 (2018), 1019–1024. <https://doi.org/10.1016/j.procs.2018.04.142>
- [55] Simon Jones, Rachel Cotterill, Nigel Dewdney, Kate Muir, and Adam Joinson. 2014. Finding Zelig in Text: A Measure for Normalising Linguistic Accommodation. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin City University and Association for Computational Linguistics, Dublin, Ireland, 455–465. <https://aclanthology.org/C14-1044>
- [56] Simon L. Jones, Emily I. M. Collins, Ana Levordashka, Kate Muir, and Adam Joinson. 2019. What is 'Cyber Security'? In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. <https://doi.org/10.1145/3290607.3312786>
- [57] Daniel Kelly, Seamas Donnelly, and Brian Caulfield. 2015. Smartphone derived movement profiles to detect changes in health status in COPD patients - A preliminary investigation. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. <https://doi.org/10.1109/embc.2015.7318399>
- [58] Mohammed Khwaja, Sumer S. Vaid, Sara Zannone, Gabriella M. Harari, A. Aldo Faisal, and Aleksandar Matic. 2019. Modeling Personality vs. Modeling Personalidad: In-the-wild Mobile Data Analysis in Five Countries Suggests Cultural Impact on Personality Models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (Sept. 2019), 1–24. <https://doi.org/10.1145/3351246>
- [59] Dohyun Kim, Joshua Gluck, Malcolm Hall, and Yuvraj Agarwal. 2019. Real World Longitudinal iOS App Usage Study at Scale. <https://doi.org/10.48550/ARXIV.1912.12526>
- [60] Philippe Kimura-Thollander and Neha Kumar. 2019. Examining the "Global" Language of Emojis: Designing for Cultural Representation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM. <https://doi.org/10.1145/3290605.3300725>
- [61] Kostas Konsolakis, Hermie Hermens, and Oresti Banos. 2019. A Novel Framework for the Holistic Monitoring and Analysis of Human Behaviour. In *13th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI 2019*. MDPI. <https://doi.org/10.3390/proceedings2019031043>
- [62] Frauke Kreuter, Georg-Christoph Haas, Florian Keusch, Sebastian Bähr, and Mark Trappmann. 2018. Collecting Survey and Smartphone Sensor Data With an App: Opportunities and Challenges Around Privacy and Informed Consent. *Social Science Computer Review* 38, 5 (Dec. 2018), 533–549. <https://doi.org/10.1177/0894439318816389>
- [63] Pranav Kulkarni, Reuben Kirkham, and Roisin McNaney. 2022. Opportunities for Smartphone Sensing in E-Health Research: A Narrative Review. *Sensors* 22, 10 (May 2022), 3893. <https://doi.org/10.3390/s22103893>
- [64] Nicholas Lane, Mashfiqui Mohammad, Mu Lin, Xiaochao Yang, Hong Lu, Shahid Ali, Afsaneh Doryab, Ethan Berke, Tanzeem Choudhury, and Andrew Campbell. 2011. BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing. In *Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare*. IEEE. <https://doi.org/10.4108/icst.pervasivehealth.2011.246161>
- [65] Neal Lathia, Gillian M. Sandstrom, Cecilia Mascolo, and Peter J. Rentfrow. 2017. Happier People Live More Active Lives: Using Smartphones to Link Happiness and Physical Activity. *PLOS ONE* 12, 1 (Jan. 2017), e0160589. <https://doi.org/10.1371/journal.pone.0160589>
- [66] Kangjae Lee and Mei-Po Kwan. 2018. Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results. *Computers, Environment and Urban Systems* 67 (Jan. 2018), 124–131. <https://doi.org/10.1016/j.compenvurbsys.2017.09.012>
- [67] Pengfei Li, Yu Wang, Yu Tian, Tian shu Zhou, and Jing song Li. 2016. An Automatic User-adapted Physical Activity Classification Method Using Smartphones. *IEEE Transactions on Biomedical Engineering* (2016), 1–1. <https://doi.org/10.1109/tbme.2016.2573045>
- [68] Shaobo Liang. 2021. Characterizing and predicting the cross-app behavior in mobile search. *Aslib Journal of Information Management* 74, 1 (Oct. 2021), 78–93. <https://doi.org/10.1108/ajim-08-2021-0220>
- [69] Shaobo Liang, Ziyi Wei, and Ning Zou. 2022. A Preliminary Study on App-Switching during the Mobile Search. *Proceedings of the Association for Information Science and Technology* 59, 1 (Oct. 2022), 741–743. <https://doi.org/10.1002/prae.2.711>
- [70] Yue Liao, Stephen S. Intille, and Genevieve F. Dunton. 2014. Using Ecological Momentary Assessment to Understand Where and With Whom Adults' Physical and Sedentary Activity Occur. *International Journal of Behavioral Medicine* 22, 1

- (March 2014), 51–61. <https://doi.org/10.1007/s12529-014-9400-z>
- [71] Tingting Liu, Salvatore Giorgi, Xiangyu Tao, Sharath Chandra Guntuku, Douglas Bellew, Brenda Curtis, and Lyle Ungar. 2022. Different Affordances on Facebook and SMS Text Messaging Do Not Impede Generalization of Language-Based Predictive Models. <https://doi.org/10.48550/ARXIV.2202.01802>
- [72] Yong Liu, Denzil Ferreira, Jorge Goncalves, Simo Hosio, Pratyush Pandab, and Vassilis Kostakos. 2016. Donating Context Data to Science: The Effects of Social Signals and Perceptions on Action-Taking. *Interacting with Computers* (May 2016). <https://doi.org/10.1093/iwc/iww013>
- [73] Fatima Ezzahraa Louhab, Yousaf Khiaat, Faycal Bensalah, Azeddine Khiaat, Ayoub Bahasse, and Mohamed Talea. 2019. Adapted Course Content Format for Flipped Mobile Learning. In *2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*. IEEE. <https://doi.org/10.1109/ihmsc.2019.10109>
- [74] Mark Matthews, Saeed Abdullah, Elizabeth Murnane, Stephen Volda, Tanzeem Choudhury, Geri Gay, and Ellen Frank. 2016. Development and Evaluation of a Smartphone-Based Measure of Social Rhythms for Bipolar Disorder. *Assessment* 23, 4 (June 2016), 472–483. <https://doi.org/10.1177/1073191116656794>
- [75] Jim McCambridge, John Witton, and Diana R. Elbourne. 2014. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology* 67, 3 (March 2014), 267–277. <https://doi.org/10.1016/j.jclinepi.2013.08.015>
- [76] Lakmal Meegahapola, Salvador Ruiz-Correa, Viridiana del Carmen Robledo-Valero, Emilio Ernesto Hernandez-Huerfano, Leonardo Alvarez-Rivera, Ronald Chenu-Abente, and Daniel Gatica-Perez. 2021. One More Bite? Inferring Food Consumption Level of College Students Using Smartphone Sensing and Self-Reports. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (March 2021), 1–28. <https://doi.org/10.1145/3448120>
- [77] MetaGeek. 2023. *Understanding RSSI*. Retrieved Aug, 2023 from <https://www.metageek.com/training/resources/understanding-rssi/>
- [78] Jonah Meyerhoff, Tony Liu, Konrad P Kording, Lyle H Ungar, Susan M Kaiser, Chris J Karr, and David C Mohr. 2021. Evaluation of Changes in Depression, Anxiety, and Social Anxiety Using Smartphone Sensor Features: Longitudinal Cohort Study. *Journal of Medical Internet Research* 23, 9 (Sept. 2021), e22844. <https://doi.org/10.2196/22844>
- [79] Laura Montanini, Nicola Sabino, Susanna Spinsante, and Ennio Gambi. 2018. Smartphone as unobtrusive sensor for real-time sleep recognition. In *2018 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE. <https://doi.org/10.1109/icce.2018.8326220>
- [80] Veljko Pejovic, Neal Lathia, Cecilia Mascolo, and Mirco Musolesi. 2015. Mobile-Based Experience Sampling for Behaviour Research. <https://doi.org/10.48550/ARXIV.1508.03725>
- [81] Mashfiqul Rabbi, Shahid Ali, Tanzeem Choudhury, and Ethan Berke. 2011. Passive and In-Situ assessment of mental and physical well-being using mobile sensors. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM. <https://doi.org/10.1145/2030112.2030164>
- [82] Nilam Ram, Xiao Yang, Mu-Jung Cho, Miriam Brinberg, Fiona Muirhead, Byron Reeves, and Thomas N. Robinson. 2019. Screenomics: A New Approach for Observing and Studying Individuals' Digital Lives. *Journal of Adolescent Research* 35, 1 (Nov. 2019), 16–50. <https://doi.org/10.1177/0743558419883362>
- [83] Thoniwut Rapeepisarn, Supasin Tatiyanupanwong, Bussarakorn Kornvisitvat, and Songsri Tangsripairoj. 2016. iRelief: An Android application for smartphone syndrome prevention and treatment. In *2016 Fifth ICT International Student Project Conference (ICT-ISPC)*. IEEE. <https://doi.org/10.1109/ict-ispc.2016.7519251>
- [84] Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark Hansen, and Mani Srivastava. 2010. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks* 6, 2 (Feb. 2010), 1–27. <https://doi.org/10.1145/1689239.1689243>
- [85] Byron Reeves, Nilam Ram, Thomas N. Robinson, James J. Cummings, C. Lee Giles, Jennifer Pan, Agnese Chiatti, Mj Cho, Katie Roehrick, Xiao Yang, Anupriya Gagneja, Miriam Brinberg, Daniel Muise, Yingdan Lu, Mufan Luo, Andrew Fitzgerald, and Leo Yeykelis. 2019. Screenomics: A Framework to Capture and Analyze Personal Life Experiences and the Ways that Technology Shapes Them. *Human-Computer Interaction* 36, 2 (March 2019), 150–201. <https://doi.org/10.1080/07370024.2019.1578652>
- [86] Byron Reeves, Thomas Robinson, and Nilam Ram. 2020. Time for the Human Screenome Project. *Nature* 577, 7790 (Jan. 2020), 314–317. <https://doi.org/10.1038/d41586-020-00032-5>
- [87] Garrett J. Roberts, Michael Solis, Stephen Ciullo, John W. McKenna, and Sharon Vaughn. 2014. Reading Interventions With Behavioral and Social Skill Outcomes: A Synthesis of Research. *Behavior Modification* 39, 1 (Dec. 2014), 8–42. <https://doi.org/10.1177/0145445514561318>
- [88] Andrea Rung, Frauke Warnke, and Nikos Mattheos. 2014. Investigating the Use of Smartphones for Learning Purposes by Australian Dental Students. *JMIR mHealth and uHealth* 2, 2 (April 2014), e20. <https://doi.org/10.2196/mhealth.3120>
- [89] Sohrab Saeb, Emily G Lattie, Konrad P Kording, and David C Mohr. 2017. Mobile Phone Detection of Semantic Location and Its Relationship to Depression and Anxiety. *JMIR mHealth and uHealth* 5, 8 (Aug. 2017), e112. <https://doi.org/10.2196/mhealth.7297>
- [90] Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, and David C Mohr. 2015. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *Journal of Medical Internet Research* 17, 7 (July 2015), e175. <https://doi.org/10.2196/jmir.4273>
- [91] Ensar Arif Sağbaşı, Serdar Korukoglu, and Serkan Ballı. 2020. Stress Detection via Keyboard Typing Behaviors by Using Smartphone Sensors and Machine Learning Techniques. *Journal of Medical Systems* 44, 4 (Feb. 2020). <https://doi.org/10.1007/s10916-020-1530-z>
- [92] Shivam Saryar, Sucheta V. Kolekar, Radhika M. Pai, and M. M. Manohara Pai. 2019. Mobile Learning Recommender System Based on Learning Styles. In *Advances in Intelligent Systems and Computing*. Springer Singapore, 299–312. https://doi.org/10.1007/978-981-13-3600-3_29
- [93] Ramona Schoedel, Michelle Oldemeier, Léonie Bonauer, and Larissa Sust. 2022. Dataset for: Systematic categorisation of 3091 smartphone applications from a large-scale smartphone sensing dataset. <https://doi.org/10.23668/PSYCHARCHIVES.5680>
- [94] Sandra Servia-Rodríguez, Kiran K. Rachuri, Cecilia Mascolo, Peter J. Rentfrow, Neal Lathia, and Gillian M. Sandstrom. 2017. Mobile Sensing at the Service of Mental Well-being. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee. <https://doi.org/10.1145/3038912.3052618>
- [95] Shams Shapsough, Ahmed Hesham, Youssef Elkhazraty, Imran A. Zulkernan, and Fadi Aloul. 2016. Emotion recognition using mobile phones. In *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE. <https://doi.org/10.1109/healthcom.2016.7749470>
- [96] Muhammad Shoaib, Stephan Bosch, Ozlem Incel, Hans Scholten, and Paul Havinga. 2014. Fusion of Smartphone Motion Sensors for Physical Activity Recognition. *Sensors* 14, 6 (June 2014), 10146–10176. <https://doi.org/10.3390/s140610146>
- [97] Julia Silvis, Deb Niemeier, and Raissa D'Souza. 2006. Social networks and travel behavior: Report from an integrated travel diary. In *11th International Conference on Travel Behaviour Research, Kyoto*.
- [98] Pier Spinazze, Alex Bottle, and Jospic Car. 2019. Digital Health Sensing for Personalized Dermatology. *Sensors* 19, 15 (Aug. 2019), 3426. <https://doi.org/10.3390/s19153426>
- [99] Bella Struminskaya, Peter Lugtig, Vera Toepoel, Barry Schouten, Deirdre Giesen, and Ralph Dolmans. 2021. Sharing Data Collected with Smartphone Sensors. *Public Opinion Quarterly* 85, S1 (Sept. 2021), 423–462. <https://doi.org/10.1093/poq/nfab025>
- [100] Bella Struminskaya, Vera Toepoel, Peter Lugtig, Marieke Haan, Annemieke Luiten, and Barry Schouten. 2020. Understanding Willingness to Share Smartphone-Sensor Data. *Public Opinion Quarterly* 84, 3 (Oct. 2020), 725–759. <https://doi.org/10.1093/poq/nfaa044>
- [101] Kalpathy Ramaiyer Subramanian. 2018. Myth and mystery of shrinking attention span. *International Journal of Trend in Research and Development* 5, 1 (2018), 1–6.
- [102] Jerry Chih-Yuan Sun and Kai-Yu Chang. 2014. Design and development of a location-based mobile learning system to facilitate English learning. *Universal Access in the Information Society* 15, 3 (Nov. 2014), 345–357. <https://doi.org/10.1007/s10209-014-0392-x>
- [103] Madiha Tahir, Zahid Halim, Atta Ur Rahman, Muhammad Waqas, Shanshan Tu, Sheng Chen, and Zhu Han. 2022. Non-Acted Text and Keystrokes Database and Learning Methods to Recognize Emotions. *ACM Transactions on Multimedia Computing, Communications, and Applications* 18, 2 (Feb. 2022), 1–24. <https://doi.org/10.1145/3480968>
- [104] Yla R. Tausczik and James W. Pennebaker. 2009. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29, 1 (Dec. 2009), 24–54. <https://doi.org/10.1177/0261927x09351676>
- [105] Lee Valentine, Simon D'Alfonso, and Reeva Lederman. 2022. Recommender systems for mental health apps: advantages and ethical challenges. *AI & SOCIETY* (Jan. 2022). <https://doi.org/10.1007/s00146-021-01322-w>
- [106] Niels van Berkel, Simon D'Alfonso, Rio Kurnia Susanto, Denzil Ferreira, and Vassilis Kostakos. 2022. AWARE-Light: a smartphone tool for experience sampling and digital phenotyping. *Personal and Ubiquitous Computing* 27, 2 (Nov. 2022), 435–445. <https://doi.org/10.1007/s00779-022-01697-7>
- [107] Niels van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *Comput. Surveys* 50, 6 (Dec. 2017), 1–40. <https://doi.org/10.1145/3123988>
- [108] Anna Vinnikova, Liangdong Lu, Jiuchang Wei, Guangbao Fang, and Jing Yan. 2020. The Use of Smartphone Fitness Applications: The Role of Self-Efficacy and Self-Regulation. *International Journal of Environmental Research and Public Health* 17, 20 (Oct. 2020), 7639. <https://doi.org/10.3390/ijerph17207639>
- [109] Fabian Wahle, Tobias Kowatsch, Elgar Fleisch, Michael Rufer, and Steffi Weidt. 2016. Mobile Sensing and Support for People With Depression: A Pilot Trial in

- the Wild. *JMIR mHealth and uHealth* 4, 3 (Sept. 2016), e111. <https://doi.org/10.2196/mhealth.5960>
- [110] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. 2014. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students using Smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM. <https://doi.org/10.1145/2632048.2632054>
- [111] Rui Wang, Weichen Wang, Min S. H. Aung, Dror Ben-Zeev, Rachel Brian, Andrew T. Campbell, Tanzeem Choudhury, Marta Hauser, John Kane, Emily A. Scherer, and Megan Walsh. 2017. Predicting Symptom Trajectories of Schizophrenia using Mobile Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (Sept. 2017), 1–24. <https://doi.org/10.1145/3130976>
- [112] Yameng Wang, Xiaotong Ren, Xiaoqian Liu, and Tingshao Zhu. 2021. Examining the Correlation Between Depression and Social Behavior on Smartphones Through Usage Metadata: Empirical Study. *JMIR mHealth and uHealth* 9, 1 (Jan. 2021), e19046. <https://doi.org/10.2196/19046>
- [113] Joseph H. Wehby, Katherine B. Falk, Sally Barton-Arwood, Kathleen L. Lane, and Cristy Cooley. 2003. The Impact of Comprehensive Reading Instruction on the Academic and Social Behavior of Students with Emotional and Behavioral Disorders. *Journal of Emotional and Behavioral Disorders* 11, 4 (Oct. 2003), 225–238. <https://doi.org/10.1177/10634266030110040401>
- [114] Gary M. Weiss, Jeffrey W. Lockhart, Tony T. Pulickal, Paul T. McHugh, Isaac H. Ronan, and Jessica L. Timko. 2016. Actitracker: A Smartphone-Based Activity Recognition System for Improving Health and Well-Being. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE. <https://doi.org/10.1109/dsaa.2016.89>
- [115] Hongyi Wen, Michael Sobolev, Rachel Vitale, James Kizer, J P Pollak, Frederick Muench, and Deborah Estrin. 2021. mPulse Mobile Sensing Model for Passive Detection of Impulsive Behavior: Exploratory Prediction Study. *JMIR Mental Health* 8, 1 (Jan. 2021), e25019. <https://doi.org/10.2196/25019>
- [116] Sarah Wiseman and Sandy J. J. Gould. 2018. Repurposing Emoji for Personalised Communication. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. <https://doi.org/10.1145/3173574.3173726>
- [117] Yunhan Wu, Daniel Rough, Anna Bleakley, Justin Edwards, Orla Cooney, Philip R. Doyle, Leigh Clark, and Benjamin R. Cowan. 2020. See What I'm Saying? Comparing Intelligent Personal Assistant Use for Native and Non-Native Language Speakers. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM. <https://doi.org/10.1145/3379503.3403563>
- [118] Hua-Cong Yang, Yi-Chao Li, Zhi-Yu Liu, and Jie Qiu. 2014. HARLib: A human activity recognition library on Android. In *2014 11th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. IEEE. <https://doi.org/10.1109/iccwamtip.2014.7073416>
- [119] Xiao Yang, Nilam Ram, Thomas Robinson, and Byron Reeves. 2019. Using Screenshots to Predict Task Switching on Smartphones. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. <https://doi.org/10.1145/3290607.3313089>
- [120] Ching-Bang Yao. 2017. Constructing a User-Friendly and Smart Ubiquitous Personalized Learning Environment by Using a Context-Aware Mechanism. *IEEE Transactions on Learning Technologies* 10, 1 (Jan. 2017), 104–114. <https://doi.org/10.1109/tlt.2015.2487977>
- [121] Ali Hameed Yassir, Ali A. Mohammed, Adel Abdul-Jabbar Alkhazraji, Mustafa Emad Hameed, Mohammed Saad Talib, and Mohanad Faeq Ali. 2020. Sentimental classification analysis of polarity multi-view textual data using data mining techniques. *International Journal of Electrical and Computer Engineering (IJECE)* 10, 5 (Oct. 2020), 5526. <https://doi.org/10.11591/ijece.v10i5.pp5526-5534>
- [122] Andrew Z. H. Yee, Ryan Yu, Sun Sun Lim, Kwan Hui Lim, Tien Tuan Anh Dinh, Lionell Loh, Andre Hadiano, and Miguel Quizon. 2022. ScreenLife Capture: An open-source and user-friendly framework for collecting screenomes from Android smartphones. *Behavior Research Methods* (Oct. 2022). <https://doi.org/10.3758/s13428-022-02006-z>
- [123] Leo Yeykelis, James J. Cummings, and Byron Reeves. 2018. The Fragmentation of Work, Entertainment, E-Mail, and News on a Personal Computer: Motivational Predictors of Switching Between Media Content. *Media Psychology* 21, 3 (Dec. 2018), 377–402. <https://doi.org/10.1080/15213269.2017.1406805>
- [124] Alexander S Young, Abigail Choi, Shay Cannedy, Lauren Hoffmann, Lionel Levine, Li-Jung Liang, Melissa Medich, Rebecca Oberman, and Tanya T Olmos-Ochoa. 2022. Passive Mobile Self-tracking of Mental Health by Veterans With Serious Mental Illness: Protocol for a User-Centered Design and Prospective Cohort Study. *JMIR Research Protocols* 11, 8 (Aug. 2022), e39010. <https://doi.org/10.2196/39010>
- [125] Chaoqun Yue, Shweta Ware, Reynaldo Morillo, Jin Lu, Chao Shang, Jinbo Bi, Jayesh Kamath, Alexander Russell, Athanasios Bamis, and Bing Wang. 2021. Fusing Location Data for Depression Prediction. *IEEE Transactions on Big Data* 7, 2 (June 2021), 355–370. <https://doi.org/10.1109/tbdata.2018.2872569>
- [126] Daniel Zarate, Michelle Ball, Maria Prokofieva, Vassilis Kostakos, and Vasileios Stavropoulos. 2023. Identifying self-disclosed anxiety on Twitter: A natural language processing approach. *Psychiatry Research* 330 (Dec. 2023), 115579. <https://doi.org/10.1016/j.psychres.2023.115579>
- [127] Bingxue Zhang, Chuantao Yin, Bertrand David, René Chalon, and Zhang Xiong. 2015. A context-aware mobile system for work-based learning. *Computer Applications in Engineering Education* 24, 2 (Oct. 2015), 263–276. <https://doi.org/10.1002/cae.21704>
- [128] Junqi Zhang, Bing Bai, Ye Lin, Jian Liang, Kun Bai, and Fei Wang. 2020. General-Purpose User Embeddings based on Mobile App Usage. <https://doi.org/10.48550/ARXIV.2005.13303>
- [129] Sha Zhao, Feng Xu, Yizhi Xu, Xiaojuan Ma, Zhiling Luo, Shijian Li, Anind Dey, and Gang Pan. 2019. Investigating smartphone user differences in their application usage behaviors: an empirical study. *CCF Transactions on Pervasive Computing and Interaction* 1, 2 (Aug. 2019), 140–161. <https://doi.org/10.1007/s42486-019-00011-4>

A PARTICIPANT DEMOGRAPHICS

In Table 4 we provide a summary of the participant demographics.

Table 4: Participant demographics

ID	Gender	Age (Years)	Education	Income (\$)	Daily Phone Use (Self-Reported Hours)	Smartphone Manufacturer
P1	Female	25 - 34	Bachelor's degree	20,000 - 29,999	2 - 3	Samsung
P2	Female	25 - 34	Master's degree	100,000 - 149,999	>5	Huawei
P3	Male	18 - 24	Bachelor's degree	30,000 - 39,999	>5	Samsung
P4	Female	45 - 54	Master's degree	40,000 - 49,999	2 - 3	Samsung
P5	Male	25 - 34	Bachelor's degree	10,000 - 19,999	>5	Samsung
P6	Female	25 - 34	Master's degree	>150,000	>5	OnePlus
P7	Female	35 - 44	Master's degree	100,000 - 149,999	3 - 4	Samsung
P8	Male	25 - 34	Bachelor's degree	90,000 - 99,999	3 - 4	Google
P9	Male	18 - 24	Some university, but no degree	>150,000	>5	Motorola
P10	Male	25 - 34	Doctoral degree	>150,000	3 - 4	Google
P11	Male	25 - 34	Bachelor's degree	10,000 - 19,999	2 - 3	Google
P12	Female	25 - 34	Bachelor's degree	>150,000	3 - 4	Google
P13	Male	18 - 24	Master's degree	40,000 - 49,999	4 - 5	Samsung
P14	Female	18 - 24	Bachelor's degree	90,000 - 99,999	4 - 5	Samsung
P15	Female	25 - 34	Bachelor's degree	20,000 - 29,999	2 - 3	Xiaomi
P16	Male	25 - 34	Master's degree	40,000 - 49,999	4 - 5	OPPO
P17	Male	25 - 34	Bachelor's degree	30,000 - 39,999	1 - 2	Google
P18	Female	18 - 24	High school graduate	40,000 - 49,999	2 - 3	Samsung
P19	Female	18 - 24	Master's degree	50,000 - 59,999	2 - 3	OnePlus
P20	Female	18 - 24	Bachelor's degree	70,000 - 79,999	4 - 5	Samsung
P21	Male	18 - 24	Some university, but no degree	<10,000	1 - 2	Samsung

B ESM RESPONSE CATEGORISATION

In Table 5 we provide a summary of the ESM response categories.

Table 5: ESM response categories.

Category	Examples
Eating/Drinking	Having Lunch, Drinking Tea, Eating Out
Entertainment	Watching TV, Gaming, Reading
Exercising	Walking, Gym, Rock Climbing
Housework	Cooking, Cleaning, Laundry
Misc	Other Activities
Online Searching	Searching Real Estate, Viewing Online Maps
Resting	Relaxing, In Bed, Napping
Shopping	Getting Groceries, Online Shopping
Socialising	Calling, Visiting Friends, Attending Convention
Travelling	Catching the Train, Driving, Biking
Working	Work Meeting, Studying For Test, Conducting Experiments

C DEBRIEFING QUESTIONNAIRE

C.1 Please rate how you felt towards using the AWARE-Light app

All self-reported ratings used Likert scales from 1 (Strong Disagree) to 7 (Strongly Agree).

It was easy to learn to use the app.

I felt comfortable using the app.

Using the app made me more conscious of the way I use my smartphone.

I was concerned about my privacy while using the app.

If future studies could give me daily feedback about my phone usage, I would be interested in receiving this information.

C.2 For each of the following sensors, please rate how comfortable you felt with it being activated on your smartphone

All self-reported ratings used Likert scales from 1 (Extremely uncomfortable) to 7 (Extremely comfortable).

Applications - captures which applications are being used, and which are installed and removed.

Battery - captures battery levels.

Communication - captures the number of incoming/outgoing/unanswered calls and SMSs.

Keyboard - captures characters that are typed using the phone keyboard (excluding passwords).

Notifications - captures information about when notifications are received and its content, if available.

Screen State - captures whether the phone is on, off, locked, or unlocked.

Screen Text - captures the text content that appears on the phone screen (excluding passwords).

Bluetooth - captures information of Bluetooth-enabled devices detected by the phone.

Location - captures the phone's location.

Network - captures information about the phone's network.

Proximity - captures the distance between the phone and the object in front of it.

Wi-Fi - captures information about Wi-Fi networks identified by the phone.

C.3 General feedback

Did you turn off any sensors during the study? Yes/No

If you turned off any sensors, please indicate which ones you turned off and why you did so. If you did not turn off any sensors, please explain what informed your decision. <open ended text entry>

Please provide any additional feedback about your experience with the study. <open ended text entry>