AWARE Narrator and the Utilization of Large Language Models to Extract Behavioral Insights from Smartphone Sensing Data

TIANYI ZHANG, University of Melbourne, Australia
MIU KOJIMA, Institute of Science Tokyo, Japan
SIMON D'ALFONSO, University of Melbourne, Australia

Smartphones, equipped with an array of sensors, have become valuable tools for personal sensing. Particularly in digital health, smartphones facilitate the tracking of health-related behaviors and contexts, contributing significantly to digital phenotyping, a process where data from digital interactions is analyzed to infer behaviors and assess mental health. Traditional methods process raw sensor data into information features for statistical and machine learning analyses. In this paper, we introduce a novel approach that systematically converts smartphone-collected data into structured, chronological narratives. The AWARE Narrator translates quantitative smartphone sensing data into English language descriptions, forming comprehensive narratives of an individual's activities. We apply the framework to the data collected from university students over a week, demonstrating the potential of utilizing the narratives to summarize individual behavior, and analyzing psychological states by leveraging large language models.

Additional Key Words and Phrases: large language models, digital phenotyping, mental health, ubiquitous computing, smartphone sensing

ACM Reference Format:
Tianyi Zhang, Miu Kojima, and Simon D'Alfonso. 2024. AWARE Narrator and the Utilization of Large Language Models to Extract Behavioral Insights from Smartphone Sensing Data. 1, 1 (November 2024), 14 pages.

1 Introduction

The modern smartphone contains an array of sensors that enable the sensing and tracking of various phone states, uses and properties. These sensors include accelerometer, GPS/geolocation, Bluetooth, communication logs (phone and SMS), application usage and keyboard activity. Given their various sensors and the opportunities to utilise them, smartphones, the Swiss army knives of digital technology, have proven to be valuable personal sensing devices, with applications in domains such as health, education and leisure.

Given their potential to track various health-related behaviours and user contexts, as well as the emergence of health apps, smartphone sensing has become a pivotal topic in digital health. This is particularly the case in digital mental health, where the concept of digital phenotyping has emerged in recent years. In short, digital phenotyping espouses the idea that the data created from our use of and interaction with digital technologies, such as smartphones, can be mined or analysed to infer behaviours and, ultimately assess mental health [1, 2]. The focus of our work in this paper is on leveraging smartphone sensing as a tool in psychology and mental health.

Once raw sensor data is collected, it is typically processed into information features that can be used in statistical analyses and machine learning model construction. For instance, from raw geolocation data one, features such as total distance travelled or time spent at the most visited location can be derived. In this paper, however, we propose a novel approach to analyze smartphone sensing data. The core idea is to translate quantitative smartphone sensing

Authors' Contact Information: Tianyi Zhang, t.zhang59@student.unimelb.edu.au, University of Melbourne, Melbourne, Australia; Miu Kojima, kojima.m.ap@m.titech.ac.jp, Institute of Science Tokyo, Tokyo, Japan; Simon D'Alfonso, dalfonso@unimelb.edu.au, University of Melbourne, Melbourne, Australia.

2024. Manuscript submitted to ACM

data records into corresponding descriptions in English (or other natural language), which could ultimately be used to construct a narrative summary that describes an individual's day (or other specified period).

The generation of sensing statements based on data from digital sensors, let alone modern smartphone sensors, has received limited attention in the literature [6]. Beyond general interest, from our perspective the idea of translating smartphone data records into English descriptions has a further motivation in recent times given the contemporary influence of large language models (LLM) and the availability of systems such as OpenAI's GPT models and Google's Gemini. Whilst one approach involves applying LLMs to tabular data for tasks such as description, prediction and general quantitative reasoning [4], another promising avenue is to convert tabular smartphone data records into sets of English statements. These statements could then be fed into LLMs to extract descriptive summaries, generate pattern insights, or even explore the potential for LLMs to make inferences about mental health.

AWARE Narrator systematically and chronologically organizes smartphone-collected data into structured narratives. In this study, we demonstrate the potential of consolidating multi-sensor smartphone data into sensing statements that can be effectively utilized with LLMs. This approach shifts the behavioral analysis problem into the realm of natural language processing. Compared to traditional analysis methodologies (e.g. basic data feature calculations), this approach offers several benefits. First, by integrating data from a wide array of sensors, the sensing statements reveal more information that might not be readily apparent through conventional analysis. Compared to raw quantitative and categorical data, the sensing statements offer finer granularity and richer, human-readable information. Second, this approach provides multi-dimensional information. Each entry point of the sensing statement encapsulates various dimensions such as timeline, sensor type, and detailed target information collected (e.g. message/call contact names or Bluetooth device names). Third, the sensing statements are designed to be interpretable by both humans and machines, thereby enhancing transparency in data analysis. This dual interpretability may better facilitate communication and collaboration between human analysts and automated systems.

2 AWARE Smartphone Sensing Data

To illustrate the core ideas of this paper and establish a framework for translating smartphone sensing data into English sensing statements, we utilize the AWARE smartphone sensing platform, in particular, the AWARE-Light app variant [9].

The AWARE platform is a versatile and powerful tool designed to harness the sensor capabilities of modern smartphones for both research and practical applications. By integrating a variety of sensors, such as accelerometers, gyroscopes, GPS, and light sensors, AWARE enables the continuous and real-time collection of data related to physical activities, environmental conditions, and user interactions. The data gathered from these sensors is stored in a MySQL database, with each sensor's data being stored into its own table. This structured data can then be analyzed and interpreted to yield deep insights into user behavior and environmental contexts.

AWARE's open-source nature, along with its robust API and plugin architecture, offers researchers and developers significant flexibility to customize and extend its functionalities to suit specific needs. This adaptability makes AWARE an invaluable resource across a wide range of disciplines, including health monitoring, smart environments, and urban computing, thereby fostering innovation and collaboration within the global research community.

We will now list all the AWARE sensors included in our conversion framework and provide a description of the fields associated with each sensor. It is important to note that, in addition to the specific fields detailed for each sensor, every sensor table includes three standard columns: row id, timestamp, and device ID.

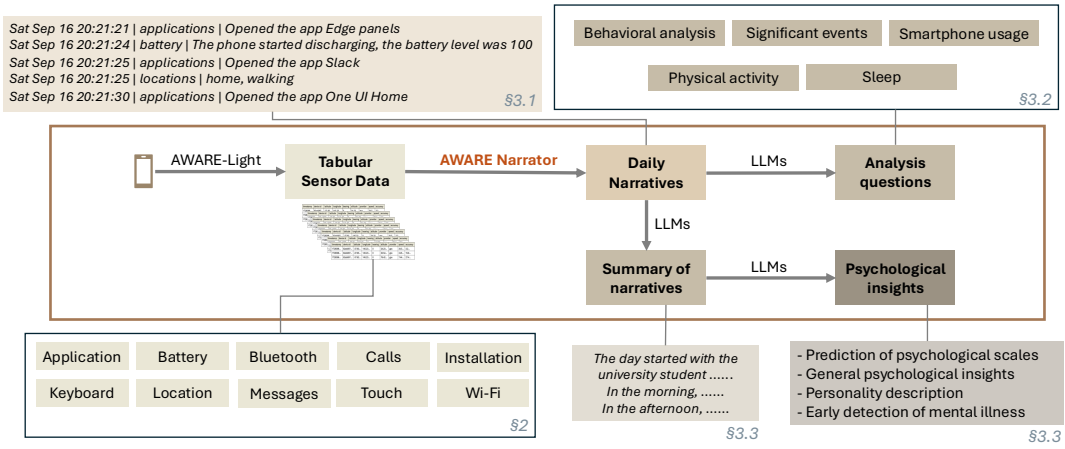

Fig. 1. AWARE Narrator workflow.

- Application (foreground) - contains the log of applications the user has interacted with. Each record contains the application name, its Android package name and whether it is an Android system app.
- Application (notifications) - contains the log of application notifications. As well as application name, this table also records the text content of the notification, although this can be masked if that setting is activated.
- Battery - contains the battery level of the device
- Bluetooth - contains the log of nearby scanned devices and connected devices.
- Calls - contains the log of incoming, outgoing and missing calls, along with masked identifiers
- Keyboard - contains the input content from the keyboard, which can be masked if set by participants
- Location - contains the log of longitude and latitude coordinates of the device's location.
- Messages - contains the log of sent and received messages with masked identifiers
- Screen - monitors the screen statuses, which can be one of four values: on, off, locked and unlocked.
- Wifi - contains the log of nearby scanned networks and connected networks.

3 AWARE Narrator: Converting AWARE Tabular Data to Narratives

In this section, we outline the process of converting tabular data into descriptive sensing statements and utilizing these sensing statements with large language models (LLMs). We begin by collecting smartphone sensor data through AWARE-Light, which is then transformed into daily sensing statements using the AWARE Narrator framework[1]. These daily sensing statements are subsequently summarized into weekly abstractions to gain psychological insights or analyze behavioral patterns based on daily activities, as depicted in Figure 1.

3.1 Converting AWARE tabular data to descriptive rows

We have developed a Python script that extracts records from each of the tables mentioned in the previous section, converts each record into an English description, and sorts all the descriptions in chronological order, from earliest to latest. Each line of output is in the following format:

[1]https://www.aware-light.org/aware-narrator/

<datetime> | <sensor> | <description of sensor record>

A comprehensive list of possible sensor descriptions follows:

- <datetime> | applications | Opened the app <application name>
- <datetime> | notifications | Received a notification from the <application name>. The content of the notification was <notification content> (or <datetime> | notifications | Received a notification from the <application name>)
- <datetime> | battery | <battery status>, the battery level was <battery level>
- <datetime> | bluetooth | Detected the nearby bluetooth device <bluetooth name> (or <datetime>| bluetooth | Detected a nearby bluetooth device)
- <datetime> | bluetooth | Connected to the bluetooth device <bluetooth name> (or <datetime> | bluetooth | Connected to a bluetooth device)
- <datetime> | calls | <call description> The call lasted <duration> seconds.
- <datetime> | installations | <application name> was <removed/added/updated>
- <datetime> | keyboard | Entered the following text into the phone keyboard: <keyboard input>
- <datetime> | locations | <current place>, <distance>m from home, <stopping/running/walking/riding vehicle> (or | locations | home, <stopping/running/walking/riding vehicle>)
- <datetime> | messages | Received a message from person <anonymous ID>
- <datetime> | messages | Sent a message to person <anonymous ID>
- <datetime> | screen status | Phone screen <turned off/turned on/locked/unlocked>
- <datetime> | touch | <touch action> <touched content> in the app <application name>
- <datetime> | wifi | Detected the nearby wifi network <wifi ID> (or <datetime> | wifi | Detected a nearby wifi network)
- <datetime> | wifi | Connected to the wifi network <wifi ID> (or <datetime> | wifi | Detected a nearby wifi network)

The battery status is described as follows:

- The phone rebooted
- The phone shutdown
- The phone started charging
- The phone started discharging
- The phone was not charging
- The phone battery became fully charged

The call description is described as follows:

- Received a phone call from person [anonymous ID]
- Made a phone call to person [anonymous ID]
- Missed a call from person [anonymous ID]

The touch actions are described as follows:

- Clicked
- Clicked longer
- Scrolled down within a view
- Scrolled up within a view

Note that for Bluetooth and Wi-Fi data, both detected and connected devices are recorded. Detected devices can provide insights into how occupied a location is, while connected devices reflect the user's active connection activity.

For battery data, rather than recording every change, we capture only the local minimum and maximum values. A similar approach is applied to the keyboard sensor, where instead of logging each character typed, we focus on the core content produced by the user, such as complete sentences, individual words, or paragraphs within a session.

Location data is processed using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, specifically employing agglomerative clustering with a 50-meter cluster diameter. The centroid of each cluster is determined using either a manually developed map or the Google Maps API to provide specific location information. Due to the reliance on cluster centroids, this method may result in inaccuracies when retrieving specific location points. Users of the AWARE Narrator framework have the flexibility to adjust the clustering threshold as needed.

The home location is identified based on the cluster containing the most nighttime data points, specifically filtered to include only data between 20:00 and 04:00. Additionally, we calculate the distance from home and determine the corresponding movement status based on recorded speed, defining four types of activity events: stopping (0 m/s), walking (0-1 m/s), running (1-3 m/s), and riding vehicles (over 3 m/s).

For applications running in the foreground, the "System UI" application is excluded as it frequently appears but does not contribute valuable information for analysis. Regarding application notifications, if the content is captured, it is included in the description; otherwise, the description will simply state "Received a notification from <application>." The same approach applies to Bluetooth and Wi-Fi data, where names are omitted if they are not detected.

For messages and calls, AWARE uniquely encrypts each contact's phone number, ensuring privacy while still allowing for the analysis of the number of different contacts and the frequency of communication with each.

```
The following data is a chronological list that describes the smartphone sensor events collected over a day from the smartphone of a university student.
The form of each data record is: timestamp | sensor | description. Answer the following question based on this data: {question}.
{data}
```

Fig. 2. Prompt structure of daily questions for descriptive records

AWARE Narrator processes sensitive information collected from raw data, such as application notifications, location data, and keyboard inputs. While this private information is preserved and integrated within the AWARE Narrator framework, it can be selectively excluded from the narratives in specific settings to address privacy concerns. This flexibility allows users to balance the richness of the data with the need for privacy protection.

3.2 Analysis of the sensing statements

Given a complete set of smartphone sensing data statements for a specific period, we investigate how large language models (LLMs) could be used to summarize the data and extract meaningful insights. For this purpose, we employed the Gemini 1.5-flash, GPT-4o, and Claude Haiku models to perform analysis tasks.

Our first task involves prompting the LLMs with the data and asking them to generate a narrative summary that captures a set of sensing statements for a given day. This summary can serve as a daily report of the individual's activities, which can be valuable for assessing daily events and monitoring mood. When extended over a longer timeline, these narratives can be collected to analyze weekly or monthly changes in behavior and well-being.

We also posed specific questions to the LLMs to extract answers from the sets of sensing statements. These questions can be divided into two categories: factual and analytical. Factual questions focus on descriptive information that

can be validated through numeric measures or simple calculations, while analytical questions involve speculation or rely on the LLM's prior knowledge, such as identifying correlational or causal relationships between elements. The questions targeted areas such as smartphone usage, physical activity, sleep, and significant events. Rather than focusing on questions that can be answered through simple calculations, like averages or standard deviations, we emphasized non-numerical, open-ended information that is challenging to infer solely from sensor data. An example is "Are there any patterns in the types of messages (e.g., work-related, social, informational)?" which can vary among different smartphone users. Each question was sent as a separate query to the LLMs. The structure of the prompts is illustrated in Figure 2. For the reader's benefits, we provide the following list of sample questions that could be applied to such data:

Smartphone Usage

- What are the peak usage times for the person using smartphones throughout the day?
- Did the person spend more time interacting with others (e.g. sending messages or making calls), or on their own (e.g. watching videos, playing games, etc.) throughout the day?
- What activities does the person use the smartphone for? What kind of role did the smartphone play in the person's life throughout the day (e.g., working tool, communication tool, or game device; work-related, social, informational)?
- Did the user have any representative behavior, such as frequently changing applications, unlocking/locking their phone, checking notifications, or frequently scrolling the application menu?
- What is the time distribution of the person using different applications throughout the day?
- What actions did the person do in their social media engagement throughout the day (e.g. scrolling, posting posts, or giving likes)?

Physical Activity

- How many and what places did the person visit, and what kind of activities did the person presumably conduct at these locations?
- What are the frequently visited places for the person?
- Was the visited place crowded or not?
- How long did the person spend at home?

Sleep

- Provide an estimation for how long the person slept.
- What patterns emerged regarding the first and last activities before sleep?

Significant Events

- Generate a narrative of the day for the person in chronological order.
- What are the minor behaviors that you may notice as a large language model that may not be evident or obvious when represented with numeric data?

Analysis

- What can be revealed from the provided data (e.g. the keyboard input the person typed, the content they browse and preference of browsed topics)? For example, the personality of the person from their tone when sending

messages, or if they initiate or respond to most calls, the schedules/plans of the person, or the opinion/background of the person?
- What would be the highlighted events for the person for the day?
- What psychological insights into the person can be provided based on their data for the day?

4 Case Study

In this section, we apply AWARE Narrator and LLM queries to some data that was obtained from one of the digital phenotyping studies we have run using AWARE-Light, namely StudentSense [3]. The StudentSense study collected data from university students for 17 weeks during an Australian summer semester at the University of Melbourne. Data was collected for the following sensors: application usage, battery, Bluetooth, calls, keyboard usage, geolocation, SMS messages, screen status and WiFi. This study was approved by the University of Melbourne Ethics Committee (approval number: [2024-25051-51517-6]).

For demonstration purposes, we have selected one of the participants with particularly comprehensive smartphone sensing data and focused on one week of the data during which they were tracked. At the end of the week, psychological measures were administered including the Depression, Anxiety and Stress Scale 21 (DASS-21) [5] and the International Positive and Negative Affect Schedule (I-PANAS-SF) [8]. The sequence of prompts we experimented with to generate summaries and derive insights from the data at both daily and weekly levels are presented.

```
Thu Sep 14 09:29:10 | applications | Opened the app One UI Home
Thu Sep 14 09:29:11 | applications | Opened the app Phone
Thu Sep 14 09:29:14 | calls | Made a phone call to person 6. The call lasted 0 seconds
Thu Sep 14 09:29:15 | applications | Opened the app Call
Thu Sep 14 09:29:15 | notifications | Received a notification from the Call
Thu Sep 14 09:29:28 | calls | Made a phone call to person 6. The call lasted 0 seconds
Thu Sep 14 09:29:32 | wifi | Connected to the wifi network <unknown ssid>
Thu Sep 14 09:29:32 | wifi | Detected the nearby wifi network “WiFi-A3E2"
Thu Sep 14 09:29:32 | applications | Opened the app Phone
Thu Sep 14 09:29:46 | calls | Made a phone call to person 6. The call lasted 610 seconds
Thu Sep 14 09:29:47 | applications | Opened the app Call
Thu Sep 14 09:29:47 | notifications | Received a notification from the Call
Thu Sep 14 09:30:12 | wifi | Detected a nearby wifi network
Thu Sep 14 09:31:30 | locations | X Sydney Rd, Coburg VIC 3058, 3099.9m from home, stopping
Thu Sep 14 09:31:59 | locations | X Sydney Rd, Coburg VIC 3058, 3099.9m from home, stopping
```

Fig. 3. AWARE Narrator example

We convert the raw data into a narrative using the AWARE Narrator framework. Figure 3 presents an example segment of the descriptive data generated by AWARE Narrator from the participant’s data, with some critical numbers and street names altered for privacy protection. Compared to traditional tabular sensor data, the narrative provides more straightforward context. For instance, we can observe that at around 9:30 AM, the participant made a 10-minute call after several attempts to reach the same person on X Sydney Road. Additional context is provided by the detected WiFi connections; since only two connections were detected, it suggests that the participant was possibly in a less crowded place. This level of abstraction and contextual information is difficult to obtain from numeric data alone but provides more meaningful insights to both clinicians and the users themselves.

From Figure 3, it is evident that multiple events from different sensors can occur simultaneously, and some information may be repeated over time (e.g., the last two lines of location information). This simultaneous occurrence and repetition highlight the richness of the narrative data, which can capture overlapping events from different sensors. The detailed

log of the person's day includes every details of them interacting with their smartphone and the information and environment they are exposed to. More in-depth analysis can be developed from this detailed plain descriptive text of behaviors.

4.1 Querying the Sets of Narrative Sensing Statements

To briefly demonstrate how answers can be extracted from the narrative information, we provide a few simple examples using GPT-4o, queried from one day's data of the participant. The first query addresses phone usage (Figure 4), the second query explores the behavior of the smartphone user (Figure 5), and the third query examines the psychology of the smartphone user (Figure 6). These three examples illustrate how general questions applied to such data can yield meaningful insights.

Example I (Figure 4) explores the relationship between the participant and the smartphone. This example can serve as a summary for smartphone users, offering daily or periodic reports on their smartphone usage and recommendations for healthier usage habits. Leveraging the LLMs, the smartphone usage data can reveal problematic behaviors such as excessive use, smartphone addiction, and nomophobia ("NoMobilePhobia"). In addition, the example captures and classifies application usage, a task that would be challenging to perform manually with traditional analysis methods. Translating the tabular form of application usage into textual descriptions makes it clearer for LLMs to analyze, with application names and functionalities. Such a summary of application usage provides comprehensive insights into the user's smartphone behaviors and how they interact with various information sources.

Example II (Figure 5) demonstrates how the collected smartphone sensor data can be used in narrative form to extract behavioural insights. Specifically, using evidence from various perspectives such as movement patterns, application usage, communication patterns, activity patterns, and connectivity via WiFi and networks, the large language models speculate on the personality and behavior based on descriptive data. We confirm the validity of information including frequent visits, used applications, and activity timelines. Although the analysis does not speculate on personality traits like the Big Five Personality with the open-ended question asked, it can reveal how daily smartphone interactions may reflect personality and psychological states. This understanding can help individuals comprehend how their smartphone usage impacts their psychological well-being.

Example III (Figure 6) presents the output from LLMs analyzing psychological insights into the participant. The summary suggests that the participant had a "highly connected, academically focused" routine and "relied heavily on technology with a structured yet flexible routine." This conclusion is based on the participant's academic activity engagement, social interactions, and location data from their smartphone. The LLMs also observe minor details, such as staying up late, time management habits, and constant app switching. These insights, which are difficult to obtain from raw numeric data, are valuable for analyzing psychological well-being and mental health. Previous studies indicate that psychological states can be reflected in smartphone usage patterns, such as anxiety shown by frequent app switching or depression indicated by keystroke patterns. The AWARE Narrator framework captures and analyzes these minor behavioral details more effectively than traditional feature generation methods.

4.2 Applying LLMs to the Narrative Data for the Prediction of Psychometric Results

Regarding the psychological insights question, whether the information provided by the LLM ultimately provides truthful/valid responses is an open question beyond the scope of this paper and would require validation in terms of work that establishes associations between digital footprint data and psychological states. Furthermore, how such psychological summaries could be used by relevant practitioners remains another interesting consideration.

What activities does the person use the smartphone for? What kind of role did the smartphone play in the person's life throughout the day (e.g., working tool, communication tool, or game device; work-related, social, informational)?

Analysis of Smartphone Usage for the University Student

Based on the provided data, we can infer several activities and roles that the smartphone played in the university student's life throughout the day. Here is a detailed analysis:

Activities the Person Uses the Smartphone For:

(1) **Charging and Battery Monitoring:**
- The phone was frequently charged, and the battery status was monitored multiple times throughout the day.

(2) **Screen Usage:**
- The phone screen was turned on and off multiple times, indicating active usage.

(3) **Wi-Fi Connectivity:**
- The phone connected to various Wi-Fi networks throughout the day, suggesting internet usage.

(4) **Location Tracking:**
- The phone's location was tracked continuously, showing the student's movements from home to various locations such as the university campus, shops, and other places in Melbourne.

(5) **Application Usage:**
- The student opened the "Assistive Ball" app at 03:46:20.

(6) **Communication:**
- There was a missed call from "person 2" at 16:40:46.

Roles the Smartphone Played in the Person's Life:

(1) **Communication Tool:**
- The missed call indicates that the phone was used for communication purposes.

(2) **Informational Tool:**
- The frequent Wi-Fi connections suggest that the phone was used to access information online, possibly for academic purposes or general browsing.

(3) **Navigation and Location Tracking:**
- The continuous location updates indicate that the phone was used for navigation or location tracking, helping the student move between different places.

(4) **Utility Tool:**
- The use of the "Assistive Ball" app suggests that the phone was used for utility purposes, possibly to assist with accessibility or ease of use.

(5) **Entertainment or Social Tool:**
- Although not explicitly mentioned, the frequent screen usage and Wi-Fi connections could imply that the phone was also used for social media, entertainment, or other leisure activities.

Summary of the Smartphone's Role Throughout the Day:

- **Morning to Early Afternoon (00:00 - 12:00):**
  - The student was mostly at home, with the phone being charged and used intermittently. The phone connected to Wi-Fi and tracked the student's location at home.
- **Midday to Early Evening (12:00 - 18:00):**
  - The student moved from home to various locations, including the university campus and nearby shops. The phone continued to track locations and connect to different Wi-Fi networks.
- **Evening to Night (18:00 - 22:00):**
  - The student returned home and continued to use the phone, with frequent Wi-Fi connections and location updates. The phone was charged multiple times, and there was a missed call.

Conclusion:

The smartphone played a multifaceted role in the student's life, serving as a communication tool, informational resource, navigation aid, utility device, and possibly an entertainment or social tool. The data indicates that the student relied on the smartphone for various activities throughout the day, both at home and while moving around the city.

Fig. 4. Example I. Smartphone usage question answered by GPT-4o: what are the activities observed from the smartphone and what role the smartphone play?

Beyond applying a general query to a phone sensing dataset, our dataset was collected during a study in which psychometric assessments were also collected, including the DASS-21 and I-PANAS-SF as specified earlier. The combination of this weekly sensing data and accompanying psychometric results provides an opportunity to experiment with asking LLMs to predict what questionnaire responses the smartphone user participant provided.

To test this idea, we started by taking the original narrative lists for each of the seven days in the given week and asked the LLM to generate a narrative summary of the day in chronological order. We then combined these seven daily

What user behaviors can be revealed from the provided data (e.g., the keyboard input the person typed, the content they browse, and preference of browsed topics)? For example, the personality of the person from their tone when sending messages or if they initiate or respond to most calls, the schedules/plans of the person, or the opinion/background of the person?)

User Behaviors and Patterns

Based on the provided data, several user behaviors and patterns can be inferred about the university student:

Home and Movement Patterns

- **Home Location:** The student spends a significant amount of time at home, as indicated by frequent "home, stopping" entries.
- **Frequent Short Movements:** The student often moves short distances from home, typically around 11.0m to 58.0m, suggesting frequent visits to nearby locations or neighbors.
- **Longer Movements:** Occasionally, the student travels further distances, such as to the main campus (995.0m to 1544.6m from home) and other locations up to 3300.7m from home.

Daily Schedule

- **Late Night Activity:** The student is active late into the night, with movements and phone usage recorded past midnight until early morning (e.g., 00:01:27 to 05:10:55).
- **Morning Activity:** The student starts moving around 07:01:43 and continues to be active throughout the day.
- **Campus Visits:** The student visits the main campus and other university-related locations during typical school hours (11:35:06 to 13:21:29).

Phone Usage Patterns

- **Frequent Screen Unlocks:** The student frequently turns on and off their phone screen, indicating regular phone usage.
- **App Usage:** The student uses specific apps like "Assistive Ball" and "System Launcher," suggesting frequent customization of the phone or access to system settings.
- **WiFi Connectivity:** The student connects to various WiFi networks throughout the day, indicating movement between different locations with available WiFi.

Charging Habits

- **Regular Charging:** The phone is frequently charged, often reaching 100% battery, indicating the student is diligent about keeping their phone charged.
- **Charging at Home:** Most charging events occur at home, suggesting the student ensures their phone is charged before leaving.

Social and Academic Life

- **Campus and Nearby Locations:** Frequent visits to the main campus and nearby streets suggest the student is actively attending classes or university-related activities.
- **Social Interactions:** Movements to various street numbers and establishments indicate the student might be meeting friends or engaging in social activities.

Personality and Preferences

- **Night Owl:** The late-night activity suggests the student might be a night owl, preferring to stay up late and possibly study or socialize during these hours.
- **Tech-Savvy:** Frequent use of system-related apps and settings indicates the student might be tech-savvy or interested in customizing their phone.

Potential Study or Work Locations

- **Frequent Locations:** The student frequently visits specific locations like "street_number202, 202 A'Beckett St" and "main campus," which could be study spots, work locations, or places of interest.

Physical Activity

- **Walking and Running:** Occasional entries of "walking" and "running" suggest the student engages in physical activity, possibly walking or running to different locations.

Connectivity and Network Usage

- **Multiple WiFi Networks:** The student connects to various WiFi networks, indicating movement between different locations with available internet access, possibly for academic or social purposes.

Conclusion

The data reveals that the student has a dynamic daily routine with frequent short movements around their home and longer trips to the university campus. They are active late at night and early in the morning, suggesting a flexible or non-traditional schedule. The student is tech-savvy, regularly uses their phone, and ensures it is well-charged. Their movements indicate a balance between academic responsibilities and social activities.

Fig. 5. Example II. Smartphone-inferred information provided by GPT-4o: What behaviors can be inferred from the provided data?

What psychological insights into the person can be provided based on their data for the day?

Psychological Insights from Smartphone Data

Based on the provided data, several psychological insights can be inferred about the university student's behavior and lifestyle:

(1) **Nocturnal Activity:**
- The student was active late into the night and early morning hours, as indicated by phone usage and location changes between 00:14 and 04:03. This suggests a possible nocturnal lifestyle or irregular sleep patterns, which could be due to studying, social activities, or other personal habits.

(2) **Frequent Movement:**
- The student moved around various locations in Melbourne throughout the day, including multiple stops at different street numbers and establishments. This indicates a high level of mobility and possibly a busy schedule involving various activities or social engagements.

(3) **Battery Management:**
- The phone's battery level was critically low at 0% around 13:21, leading to a shutdown. The student then charged the phone intermittently throughout the evening, with frequent disconnections and reconnections to the charger. This behavior might indicate a lack of planning or awareness regarding battery management, or it could be due to the student being in different locations without consistent access to charging facilities.

(4) **Social and Public Spaces:**
- The student spent significant time in public and social spaces such as Riverfront Plaza, Unity Plaza, and various streets in Melbourne. This suggests a preference for being in social environments or public areas, which could indicate extroverted tendencies or a need for social interaction.

(5) **Frequent Phone Screen Activity:**
- The phone screen was turned on and off frequently, especially in the evening. This could indicate a high level of phone usage, possibly for communication, social media, or other applications. The frequent screen activity might also suggest a habit of regularly checking the phone, which could be linked to anxiety or a need for constant connectivity.

(6) **Irregular Charging Patterns:**
- The phone was charged in short bursts, often disconnecting and reconnecting within minutes. This irregular charging pattern might indicate a lack of stable access to power sources or a habit of charging the phone only when necessary.

(7) **Use of Assistive Apps:**
- The student opened the app "Assistive Ball" at 02:47, which is often used for accessibility purposes or to enhance phone navigation. This could suggest that the student has specific needs or preferences for phone usage that require assistive technology.

(8) **Connectivity to Unknown Wi-Fi Networks:**
- The phone connected to various unknown Wi-Fi networks throughout the day. This could indicate that the student frequently moves between different locations with available Wi-Fi, possibly indicating a need for internet access for academic or personal purposes.

(9) **Physical Activity:**
- The data shows instances of walking, running, and riding a vehicle. This variety in movement types suggests that the student engages in different forms of physical activity, which could be part of their daily routine or related to their transportation methods.

(10) **Potential Stress or Anxiety:**
- The frequent phone screen activity, irregular sleep patterns, and constant movement might indicate underlying stress or anxiety. The student's behavior could be a response to academic pressures, social obligations, or personal issues.

Overall Conclusion

The data suggests that the student leads a dynamic and possibly hectic lifestyle characterized by irregular sleep patterns, frequent movement, and high phone usage. These behaviors could be influenced by academic responsibilities, social activities, or personal habits. Understanding these patterns can help in providing better support and resources for managing their time, stress, and overall well-being.

Fig. 6. Example III. One of the analysis question answered by GPT-4o: What psychological insights can be inferred?

narrative summaries together in order and queried this weekly summary information in terms of questions related to the I-PANAS-SF and DASS-21. The prompt for DASS-21 and I-PANAS-SF are shown in Figure 7 and 8 respectively:

We experimented with three LLMs (GPT-4o, Gemini 1.5 Pro, and Claude-Haiku) to summarize the descriptions generated by the AWARE Narrator for a weekly psychological report. These LLMs were also tasked with predicting scores for the DASS-21 and I-PANAS-SF based on the daily narratives produced by the AWARE Narrator. The results, presented in Table 1, and Table 2 demonstrate the LLMs' ability to summarize narratives and predict psychological states. We observe that LLMs perform relatively well on I-PANAS-IF tasks and acceptably on DASS-21 tasks. However, due to the limited data and the fact that this is just a demonstration where our focus is not on prediction optimization, accuracy metric evaluations are not provided in this paper. Additionally, prompt engineering was not optimized, nor were additional examples provided to LLMs to enhance prediction performance. This demonstration illustrates how transforming raw tabular data into narratives using the AWARE Narrator can be beneficial in ubiquitous computing. In

Using the narrative of activities collected from smartphone sensors over the past week, estimate the individual's mental health based on the DASS-21 scale. This scale evaluates three subscales: depression, anxiety, and stress, each with a maximum score of 21. For each subscale, categorize the result into one of the following ranges:
Normal: 0 to 4
Mild: 5 to 6
Moderate: 7 to 10
Severe: 11 to 13
Extremely Severe: 14 and above

Format your output as follows:
Depression: <extent>
Anxiety: <extent>
Stress: <extent>

Narratives data: {narrative data}

Fig. 7. Prompt for DASS-21 prediction

Given a week's narrative of activity collected from smartphone sensors, estimate the scores for the following affective states that the person would report at the end of the week using the I-PANAS-SF scale. The scale ranges from 1 to 5, with 1 representing 'Never' and 5 representing 'Always':

Upset
Hostile
Alert
Ashamed
Inspired
Nervous
Determined
Attentive
Afraid
Active

Narratives: {Narratives for each day of the week}

Fig. 8. Prompt for weekly I-PANAS-SF prediction

comparison to feeding raw numerical data into LLMs, which can exceed limits and reduce the information extracted by the models, narratives offer a clearer and more understandable description. These narratives can be more easily comprehended by humans, facilitating better interpretation and analysis.

Previous work has explored similar approaches to predicting psychological measures. For instance, [10] leveraged LLMs to predict affective states by extracting daily features using RAPIDS, which calculated high-level abstracted features such as the total duration of application usage, the total number of missed calls, and the time spent at home. In contrast, our study directly converts each line of raw data into descriptive English narratives. This approach ensures data completeness, preserving the information of each individual event. The comprehensive narratives produced by the AWARE Narrator can potentially be used in conjunction with LLMs or human inspection for clinical reports, assisting psychiatrists and psychologists in helping individuals with mental health issues. By maintaining detailed event information, our method provides a more thorough basis for psychological assessment and intervention.

5 Future Work and Conclusion

Our exploratory demonstrations with the LLMs were confined to zero-shot prompts. Future work on this idea should explore multi-shot prompting. One approach would be to include some psychological measures of interest as output for each shot input (i.e., set of sensor descriptions). For example, the set of sensor descriptions for each of several

Table 1. Comparison of the Prediction Results of DASS-21 by Different LLMs and Actual Results.

Narrative LLM	Gemini			OpenAI			Claude			Actual Results
Prediction LLM	OpenAI	Gemini	Claude	OpenAI	Gemini	Claude	OpenAI	Gemini	Claude	
Depression	Moderate	Mild	Mild	Moderate	Moderate	Mild	Moderate	Mild	Mild	Normal
Anxiety	Moderate	Moderate	Moderate	Moderate	**Mild**	Moderate	Moderate	Moderate	Normal	Mild
Stress	Moderate	Normal	Normal	Moderate	Moderate	Normal	Moderate	Moderate	Moderate	Mild

Table 2. Comparison of the Prediction results of I-PANAS-IF by Different LLMs and Actual Results.

Narrative LLM	Gemini			OpenAI			Claude			Actual Results
Prediction LLM	OpenAI	Gemini	Claude	OpenAI	Gemini	Claude	OpenAI	Gemini	Claude	
active	**5**	3	4	3	**5**	**5**	**5**	3	**5**	5
determined	**4**	**4**	**4**	3	**4**	**4**	**4**	3	**4**	4
attentive	**4**	3	**4**	3	3	**4**	**4**	3	**4**	4
inspired	3	2	3	2	3	**4**	3	2	**4**	4
alert	4	**3**	4	**3**	4	4	4	**3**	4	3
upset	**2**	3	**2**	**2**	**2**	**2**	**2**	**2**	**2**	2
hostile	1	**2**	**2**	1	1	1	1	1	1	2
ashamed	**1**	2	2	**1**	**1**	**1**	**1**	**1**	**1**	1
nervous	**2**	**2**	3	**2**	3	**2**	**2**	**2**	**2**	2
afraid	**1**	2	2	**1**	**1**	**1**	**1**	**1**	**1**	1

consecutive days could be provided as example inputs, accompanied by psychological assessments taken at the end of those days as example outputs to improve the accuracy of the predictions. Additionally, employing a Retrieval Augmented Generation (RAG) could enhance how contextual information is integrated into the model.

The tasks for the LLM could then include predicting psychological assessment values for subsequent inputs without known outputs. This paradigm may assist in the early detection of mental health issues and mood monitoring. Fine-tuning models can also be developed for personalization or for specific cohorts (e.g., individuals with depressive symptoms) to infer physical and mental health and well-being for self-regulation.

Another line of future work concerns extending the AWARE Narrator tool in terms of the range of descriptions it can generate. This could be done by incorporating other sensors or sources. For example, AWARE-Light includes a novel sensor called the screen reader, which collects text displayed on the screen [7]. This sensor can extract core information and keywords, which can be added to the AWARE Narrator for enhanced narrative content. It could also be done by generating descriptions that combine two or more sensors. For example, “the user was using the Spotify app whilst travelling on a train", which involves both the application sensor and the movement sensors.

Given the highly sensitive nature of the data collected from smartphone sensors, privacy concerns can be mitigated by running the AWARE Narrator and LLM analysis system directly on-device, thus transforming fine-grained sensor data into an abstract summary of activities without leaving the phone.

Our investigation has demonstrated the feasibility and usability of the AWARE Narrator framework. Assisted by large language models (LLMs), the generated narrative descriptions can be analyzed to infer behaviours, contexts, psychological characteristics and mental health. This makes the AWARE Narrator a practical and useful tool not only for visualizing and representing one’s engagement and behaviors with smartphones but also for understanding the

information and environments to which the smartphone user is exposed. It provides a novel way to interpret classical tabular data in digital phenotyping, offering insights into human behaviors and psychological states.

References

[1] Harald Baumeister and Christian Montag. 2019. *Digital phenotyping and mobile sensing*. Springer.

[2] Pasquale Bufano, Marco Laurino, Sara Said, Alessandro Tognetti, and Danilo Menicucci. 2023. Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *Journal of Medical Internet Research* 25 (2023), e46778.

[3] Simon D'Alfonso and Tianyi Zhang. 2024. StudentSense. https://doi.org/10.17605/OSF.IO/DJS2Y.

[4] Xi Fang, Weijie Xu, Fiona Anting Tan, Jiani Zhang, Ziqing Hu, Yanjun Qi, Scott Nickleach, Diego Socolinsky, Srinivasan Sengamedu, and Christos Faloutsos. 2024. Large Language Models(LLMs) on Tabular Data: Prediction, Generation, and Understanding – A Survey. arXiv:2402.17944 [cs.CL] https://arxiv.org/abs/2402.17944

[5] Sydney H Lovibond and Peter F Lovibond. 1995. Depression anxiety stress scales. *Psychological Assessment* (1995).

[6] Joseph Reddington and Nava Tintarev. 2011. Automatically generating stories from sensor data. In *Proceedings of the 16th international conference on Intelligent user interfaces*. 407–410.

[7] Songyan Teng, Tianyi Zhang, Simon D'Alfonso, and Vassilis Kostakos. 2024. Predicting Affective States from Screen Text Sentiment. *arXiv preprint arXiv:2408.12844* (2024).

[8] ER Thompson. 2007. Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS), 38 (2), 227–242.

[9] Niels van Berkel, Simon D'Alfonso, Rio Kurnia Susanto, Denzil Ferreira, and Vassilis Kostakos. 2023. AWARE-Light: A smartphone tool for experience sampling and digital phenotyping. *Personal and Ubiquitous Computing* 27, 2 (2023), 435–445.

[10] Tianyi Zhang, Songyan Teng, Hong Jia, and Simon D'Alfonso. 2024. Leveraging LLMs to Predict Affective States via Smartphone Sensor Features. arXiv:2407.08240 [cs.HC] https://arxiv.org/abs/2407.08240