AWARE-Text: An Android Package for Mobile Phone Based Text Collection and On-Device Processing

Salvatore Giorgi[1,2,*], **Garrick Sherman**[1,*], **Douglas Bellew**[1],
Sharath Chandra Guntuku[2], **Lyle Ungar**[2] and **Brenda Curtis**[1]
[1]National Institute on Drug Abuse, [2]University of Pennsylvania
{sal.giorgi, garrick.sherman, doug.bellew, brenda.curtis}@nih.gov
{sharathg, ungar}@cis.upenn.edu

Abstract

We present the AWARE-Text package, an open-source software package for collecting textual data on Android mobile devices. This package allows for collecting short message service (SMS or text messages) and character-level keystrokes. In addition to collecting this raw data, AWARE-Text is designed for *on device* lexicon processing, which allows one to collect standard textual-based measures (e.g., sentiment, emotions, and topics) without collecting the underlying raw textual data. This is especially important in the case of mobile phones, which can contain sensitive and identifying information. Thus, the AWARE-Text package allows for privacy protection while simultaneously collecting textual information at multiple levels of granularity: person (lifetime history of SMS), conversation (both sides of SMS conversations and group chats), message (single SMS), and character (individual keystrokes entered across applications). Finally, the unique processing environment of mobile devices opens up several methodological and privacy issues, which we discuss.

1 Introduction

Unlike traditional NLP tasks (e.g., machine translation or question answering), NLP in the context of psychological, social, and health sciences is aimed at understanding how textual data can characterize people. This includes stance or sarcasm at document-level (Lynn et al., 2019), state-level tasks, such as emotion prediction (Mohammad, 2016), trait-level tasks, such as personality prediction (Park et al., 2015) or mental health applications (De Choudhury and De, 2014), or even population-level tasks, for example, monitoring the opioid epidemic via social media data (Giorgi et al., 2023). Similarly, keystroke data (or typing dynamics, i.e., a succession of individual depressions on a keyboard) has been used to predict emotions (Epp et al., 2011) and cognition (Brizan et al., 2015). Historically, human-generated textual data for such social science-oriented tasks is collected from social media (e.g., Facebook, Reddit, or Twitter), open-ended survey questions (Kjell et al., 2022), or interviews (Son et al., 2023). However, more recently, short message service (SMS or text messaging) has received attention as a viable data source (Liu et al., 2023; Meyerhoff et al., 2023; Benoit et al., 2020; Nook et al., 2021; Stamatis et al., 2022a,b; Tlachac and Rundensteiner, 2020; Tlachac et al., 2022; Ameer et al., 2022). SMS, and, more generally, mobile phone-based data, is important for Just-in-Time Adaptive Interventions (JITAIs), which can be used to deliver personalized support and interventions in response to a person's changing physical and mental health (Nahum-Shani et al., 2018).

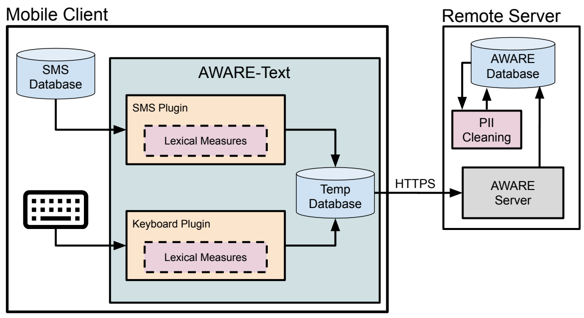

Figure 1: Data flow diagram. Data stored and collected on a mobile client is sent through AWARE-Text which then processes the textual data and transmits raw (i.e., raw SMS or keystrokes) and lexical data to a remote sever via a secure, encrypted connection. Privacy preserving methods are shown in red.

AWARE-Text[1] is an Android mobile phone application (or "app") built to collect passive mobile data (e.g., GPS locations and accelerometer data) with particular emphasis on textual[2] data

[*] Authors contributed equally.

[1]https://github.com/TTRUCurtis/aware-text

[2]To disambiguate *text messages* (or SMS) from *text data* (data in the form of written text), we use the term *textual data*

such as SMS and keystrokes. This app allows researchers to collect raw textual data, both historical data and prospective data *in real time*, as well as lexical-based measures calculated on the device. On-device processing of lexical-based measures, such as sentiment or topics, inherently preserves privacy: summary scores (e.g., sentiment) are transmitted to a remote server and the underlying raw data, which can be highly sensitive and contain personally identifiable information (PII), does not need to leave the mobile device.

In summary, AWARE-TEXT focuses on two types of textual data (SMS data and keyboard input) across four levels of granularity (person, conversation, message, and character). Across all data types and levels AWARE-TEXT offers the ability to collect raw data as well as lexical measures, which inherently preserve privacy, as shown in Figure 2. This package offers researchers the ability to collect fine-grained textual data which can be used to gain insight into tasks across natural language processing, psychology, computational social sciences, and psycho-linguistics.

2 Overall System Design

The AWARE-TEXT package is an extension of the AWARE mobile sensing framework, an open-source package developed to passively collect mobile phone sensor data, such as accelerometer, gyroscope, and GPS data (Ferreira et al., 2015). This extension consists of two on-device plugins (to collect SMS data and process lexical measures) and a series of post-processing scripts for data aggregation and cleaning. Thus, AWARE-TEXT is able to collect everything AWARE is able to collect, plus additional textual data. While the AWARE framework is available for both iOS and Android devices, the AWARE-TEXT package is only available on Android devices due to iOS restrictions on access to raw SMS and keyboard data.

The high-level features of AWARE-TEXT are shown in Figure 1. Here we see that AWARE-TEXT pulls data from both the mobile device's keyboard and local SMS database (described below). Both data types are then optionally processed into lexical measures, after which the raw and processed lexical data are stored in a temporary local database. This data is transferred (via a secure HTTPS connection) to a remote server and stored in a final database. AWARE-TEXT is designed to optimally transfer data whenever wifi connections are available to minimize the amount of cellular data the application uses. The temporary database is then cleared so as not to duplicate data.

when referring to data in text form.

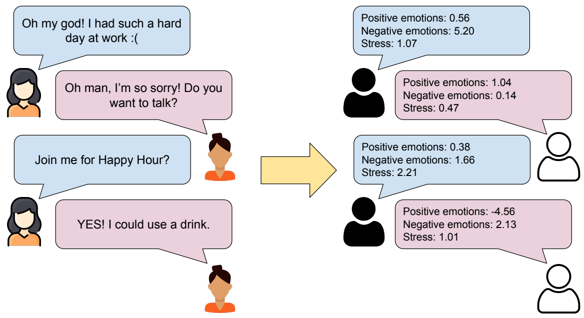

Figure 2: AWARE-TEXT has anonymized both people in the conversation and the exact text written within each utterance, while preserving the conversation structure.

3 Data Types

AWARE-TEXT collects two types of raw text data: SMS and keystrokes. While the keystroke data is available in the original AWARE implementation, the keystrokes lexical processing is novel to AWARE-TEXT, as is the SMS collection.

3.1 SMS

SMS data includes traditional SMS and more recent types of messaging, including MMS (Multimedia Messaging Service) and RCS (Rich Communication Services). This includes group messages (text messages between three or more people) and reactions to messages, such as emojis. Each message is timestamped to indicate the time sent or received. We note that only textual data is collected, and no images or audio files are stored. Finally, information on who is on the opposite end of the received or sent SMS is stored via a hashed identifier. This is done in such as way that hashes are consistent across communications, which enables one to reconstruct conversations or identify SMS messages sent to a particular (non-identifiable) person.

SMS collection can occur retrospectively (the complete history of SMS stored on the mobile device) and prospectively (all SMS messages exchanged while AWARE-TEXT is running). Additionally, SMS is collected from both the person running AWARE-TEXT on their device (i.e., sent messages) as well as from others (i.e., all received messages). Data collection is fully configurable, and all combinations of retrospective/prospective and sent/received are available.

3.2 Keystrokes

Keystrokes are single depressions of a key on a keyboard and include non-standard characters such as deletions and auto-completes, along with information on the time between each key press. Keystroke data is collected per application. This allows one to measure typing dynamics in applications such as Facebook Messenger or the local web browser. For example (as seen in Table 1), if a user searched "Taylor Swift‘", while misspelling the name, then AWARE-TEXT would collect rows for each of "T", "Ta", "Tai", "Ta" (i.e., a deletion occurred), "Tay", etc. No passwords are collected via AWARE-TEXT. Finally, we note that when applying lexical measures to the keystroke data, we only consider the complete keyboard input for a single typing session (for example, input to a search engine) instead of running lexica across each character.

4 Levels of Data Collection

AWARE-TEXT has been designed to enable analysis of textual data at various levels of granularity: person, conversation, messages, and characters. This is true of both raw data and lexical measures. This flexibility is enabled by collecting data at a low level and preserving summary statistics that may be aggregated to higher levels of analysis.

Person-level data is available by aggregating raw text (SMS or keystrokes) or lexical measures across a person's individual inputs (for example, their lifetime SMS history). Conversations can be constructed by combining SMS data between pairs and groups of people. This can include SMS from non-consenting individuals (see Ethical considerations below). Message data (raw or lexical) is obtained by single SMS or complete keyboard input and is the most basic unit available for lexical measures. Finally, raw character-level data is obtained through keystroke inputs.

5 Privacy Preserving Lexical Measures

Running lexical measures *on the device* allows researchers the ability to collect data across each SMS or keystroke without necessitating the collection of the underlying raw data. This raw data could include sensitive information (e.g., revealing search histories or SMS with PII) and data from non-consenting individuals (e.g., SMS from people communicating with the study participant). As shown in Figure 2, the lexical measures can be applied to both ends of the conversation, thus obfuscating both the *exact people* in the conversation and their *exact utterances* while still preserving both the overall conversation and individual turns within the conversation. While mobile phones can collect data in multiple languages, this tokenizer is designed for English text, and thus would need to be replaced for on-device processing of non-English languages.

Device ID	Timestamp	Application Name	Before Text	After Text
1	08/06/2023 12:46:56	Chrome		T
1	08/06/2023 12:46:57	Chrome	T	Ta
1	08/06/2023 12:46:58	Chrome	Ta	Tai
1	08/06/2023 12:46:58	Chrome	Tai	Ta
1	08/06/2023 12:46:59	Chrome	Ta	Tay
2	08/12/2023 07:02:23	Instagram		p
2	08/12/2023 07:02:23	Instagram	p	pi
2	08/12/2023 07:02:23	Instagram	pi	piz
2	08/12/2023 07:02:24	Instagram	piz	pizza

Table 1: Example keystroke data. Each row contains a single depression of a key on the keyboard and includes both the current text and the previous text, allowing one to easily identify deletions and autocompletes.

Preprocessing Applications on the Android OS are Java and, more recently, Kotlin based. Therefore, we use a Java-based tokenizer from the Natural Language Processing for JVM languages (NLP4J) project[3] developed by EmoryNLP. While this tokenizer is not designed for noisy user-generated textual data such as SMS, several key features make it useful for this setting, such as emoji recognition.

Lexical Data Table 2 shows a sample of the SMS lexical data. Here we see both the Device ID (the mobile device running AWARE-TEXT) and the Contact ID as numeric or hashed identifiers, thus removing any identifying information. Total Words and Lexicon Words are, respectively, defined as the total number of words in the message and the number of words in the intersection between the message and the lexicon. This allows one to normalize the lexicon score in various ways. Additionally, this allows one to aggregate scores across levels. For example, one could aggregate the *stress* scores across messages in a given day, normalized using the Total Words, to produce a daily *stress* score. Similarly, this can be done across people, conversations, or applications (via keystroke data).

[3] https://emorynlp.github.io/nlp4j/

Device ID	Lexicon Category	Total Words	Lexicon Words	Score	Contact ID	Type	Timestamp
1	stress	11	5	17.0798	9f27e	sent	08/06/2023 12:46:56
1	happiness	11	7	1.17805	9f27e	sent	08/06/2023 12:46:56
1	stress	4	1	0.382509	c3d17	received	08/07/2023 17:52:01
1	happiness	4	3	0.585531	c3d17	received	08/07/2023 17:52:01
2	loneliness	25	20	-45.5145	73e48;ca96d	sent	08/07/2023 01:43:12
2	life satisfaction	25	6	0.0454235	73e48;ca96d	sent	08/07/2023 01:43:12

Table 2: Example of lexical measures across the SMS data. Lexical measures include stress, happiness, loneliness, and life satisfaction. The person on the other end of the conversation is consistently hashed (e.g., 9f27e) in order to preserve conversations. Group messages include a list of all recipients. Total and Lexicon Words allow for different types of normalization, as well as the aggregation of category scores across time and people.

Post-processing Given the sensitive nature of the raw keystroke features, we include a post-processing script that can be automatically run on the remote AWARE server (Figure 1) to remove potentially identifying information. This script uses spaCy's Named Entity Recognizer and regular expressions[4] to remove mentions of names, numbers, places, etc. The explicit mentions are replaced with their respective category names (e.g., "Taylor Swift" is replaced with {NAME}), and the category names are backpropagated through the data to the first keypress in the explicit mention.

Data Aggregation The lexical data can be further post-processed using the open-source RAPIDS package (Vega et al., 2021). This package is used to process raw mobile sensing data in order to extract behavioral features. In particular, RAPIDS is designed to work with the AWARE and AWARE-TEXT apps. This package can be used to aggregate lexical measures across time, people, and applications and combinations of each. For example, RAPIDS can aggregate lexical measures across hours, days, or even applications and days together.

Prepackaged Lexical Measures AWARE-TEXT comes prepackaged with state-of-the-art lexica for measuring psychological well-being from textual data (see original publications for details): happiness (Giorgi et al., 2021), life satisfaction (Jaidka et al., 2020), loneliness (Guntuku et al., 2019b), politeness (Li et al., 2020), and stress (Guntuku et al., 2019a).

Extending the Lexical Measures The prepackaged lexical measures in AWARE-TEXT can be easily extended to include any measure that can be decomposed into a category with weighted words (note that weights can be trivially set to one for all words). For example, this can include Latent Dirichlet Allocation (LDA) topics, where weights are conditional probabilities estimated through the LDA process. This can also include other popular lexical measures such as LIWC (Pennebaker et al., 2001), the NRC Emotion/Valence-Arousal-Dominance/Sentiment lexica (Mohammad and Kiritchenko, 2015; Mohammad, 2018), and ANEW (Warriner et al., 2013).

[4]https://github.com/madisonmay/CommonRegex

6 Methodological Considerations

Running on-device computation opens up methodological and computational issues. First, while many NLP tools exist in Java, such as the Stanford CoreNLP toolkit (Manning et al., 2014), most modern libraries are written in Python, making them inaccessible in an Android environment. Thus, many new technologies, such as contextual embeddings, are unavailable for on-device processing. Second, one must consider the person using the device. For example, high computation can quickly drain the phone's battery or slow down other applications. Similarly, transferring large amounts of data to a re-

	Facebook	SMS
Age	.68	.45*
Gender†	.91	.80*
Depression	.36	.29
Life Satis.	.21	.14
Stress	.21	.18

Table 3: Product moment correlations (or † accuracy) between language estimates and self-reports across both platforms reported in Liu et al. (2023). * significant difference in bootstrapping test between the SMS and Facebook correlations.

	Open Source	SMS	Keystrokes	On-device Lexical Processing
AWARE-TEXT	✓	✓	✓	✓
AWARE (Ferreira et al., 2015)	✓		✓	
Beiwe (Onnela et al., 2021)	✓	✓		
EARS (Lind et al., 2018)		✓	✓	
Passive Data Kit (Audacious Software, 2018)	✓			✓
m-Path (Mestdagh et al., 2023)		✓		
mindLAMP (Torous et al., 2019)	✓	✓		

Table 4: Comparison of recent mobile sensing platforms in their textual data collecting capabilities. AWARE-TEXT is the only open-source app which collects multiple types of data while offering on-device processing.

mote server can quickly increase data usage and the user's mobile phone bill. These issues can cause participants to uninstall AWARE-TEXT which can lead to low study completion rates. Thus, algorithms should reduce run-time and throughput.

7 Case Study

To date, one study has used the AWARE-TEXT app with a sample of 120 participants who installed AWARE-TEXT, answered a series of self-reports, and shared Facebook language data. This study compared preexisting social media-based lexical models in their ability to predict self-reports when applied to out-of-domain textual data (Liu et al., 2023). It applied five models trained from Facebook language to predict self-reported age, gender, depression, life satisfaction, and stress. Each model was separately applied to SMS and Facebook posts, and the resulting model predictions were compared to self-reports. We report their findings in Table 3. The results from this study show that for three out of five models, the SMS-based predictions did not statistically differ from the Facebook-based predictions, indicating that SMS is a potential data source for investigating social-psychological traits of people. This paper represents a preliminary analysis, and further investigation is needed into the strengths and weaknesses of SMS data.

8 Comparison of Mobile Text Apps

There are several apps which allow for either SMS retrieval or keystroke logging. SMS retrieval apps are typically designed for personal use, such as data backups and phone transfers, such as SMS Backup & Restore[5], or legal discovery, such as Logikcull[6]. Finally, there are also apps used for survey collection via SMS, such as ODK[7]. These apps send questions and receive answers via SMS, and data collection is typically limited to retrieving the survey question responses. Keylogging apps are typically designed for the purpose of monitoring, such as Kidlogger[8] which allows parents to monitor their children's phone activity. Thus, most apps which collect this data are not used for general data collection and research purposes.

Apps which collect SMS or keystroke data that are designed for research purposes are typically in the domain of mobile phone–based sensing software. There are several popular apps in this domain which have been used for social scientific research. In Table 4, we summarize the textual data collecting capabilities of several apps (those updated since 2018[9]). We note that AWARE-TEXT is the only open-source app which collects multiple types of text data while offering on-device processing.

9 Conclusions

AWARE-TEXT is a novel data collection application for Android mobile phones designed to capture textual data through SMS and keystrokes. This application allows researchers to collect data from consenting participants at multiple levels of granularity (person, conversation, message, and character) with the additional ability to collect both raw and aggregate lexical measures which preserve privacy. Over 85% of Americans own smartphones, and a growing number identify as smartphone dependent (i.e., smartphone serves as their primary means of online access). Thus, mobile phones offer a rich data source, which can be used as a lens into the daily lives of large sections of the population.

[5]https://www.synctech.com.au/sms-backup-restore/
[6]https://www.logikcull.com/
[7]https://odk.org/
[8]https://kidlogger.net/
[9]https://w.wiki/7qPn

Ethical Considerations

While AWARE-TEXT allows one to collect anonymized data (in the form of lexical based measures) and offers post-processing cleaning scripts, it does offer the ability to collect raw data. Raw SMS and keystroke data can contain highly sensitive and identifying information, such as names and social security numbers. Similarly, passwords (while not collected in the keystroke data) can be found in SMS data. It also offers the ability to collect data from non-consenting individuals in the form of received SMS. While no identifying information is explicitly collected from these individuals (e.g., all mobile phone numbers are hashed), the SMS data may contain sensitive information. Working with such data therefore requires a high level of care.

Collecting SMS data across conversations involves collecting data from non-consenting individuals. As discussed above, AWARE-TEXT offers the ability to collect de-identified lexical features as opposed to the underlying raw data. If the raw SMS data must be collected then AWARE-TEXT offers post-processing scripts that attempt to automatically remove sensitive and identifying information, which we *highly* recommend using. The resulting text (associated to an individual only with a hashed identifier) may be considered de-identified and, therefore, might not always be considered human subjects research. These distinctions should ultimately be decided by an Institutional Review Board. Note that further privacy preserving actions can be taken in order to help ensure equitable data collection. One study which collected both sides of SMS conversations (Song et al., 2014), offered participants the ability to remove or alter text and asked participants to consider the preferences of their conversation partners when removing sensitive text[10].

While outside of the scope of the AWARE-TEXT app, when storing this data on remote servers we recommended following standard security protocols such limited access, encryption, and two factor authentication. When collecting other measures from participants (such as surveys, demographics, or medical records), we recommend that raw data from AWARE-TEXT be stored separately and, if possible, on separate machines.

Mobile sensing, in general, opens up several ethical and privacy issues, especially in the context of health or when used to collect data from vulnerable populations; see Breslin et al. (2019) and Fuller et al. (2017) for in-depth discussions on such issues. At a minimum, we believe all uses of AWARE-TEXT should obtain approval from an ethical review board (e.g., Institutional Review Board or Ethics Committee). Similarly, researchers should follow informed consent principals when recruiting study participants.

[10]This study did not use AWARE-TEXT.

Acknowledgements

This research was supported by the Intramural Research Program of the NIH, NIDA (ZIA DA000628). We thank Olivia Dodge and Miguel Galván for their programming expertise. We also thank both Denzil Ferreira and the AWARE team, as well as Julio Vega and the RAPIDS team for their guidance.

References

Iqra Ameer, Grigori Sidorov, Helena Gomez-Adorno, and Rao Muhammad Adeel Nawab. 2022. Multi-label emotion classification on code-mixed text: Data and methods. *IEEE Access*, 10:8779–8789.

Audacious Software. 2018. Passive data kit. https://passivedatakit.org/.

James Benoit, Henry K. Onyeaka, Matcheri S. Keshavan, and John B Torous. 2020. Systematic review of digital phenotyping and machine learning in psychosis spectrum illnesses. *Harvard Review of Psychiatry*, 28:296 – 304.

Samantha Breslin, Martine Shareck, and Daniel Fuller. 2019. Research ethics for mobile sensing device use by vulnerable populations. *Social Science & Medicine*, 232:50–57.

David Guy Brizan, Adam Goodkind, Patrick Koch, Kiran Balagani, Vir V Phoha, and Andrew Rosenberg. 2015. Utilizing linguistically enhanced keystroke dynamics to predict typist cognition and demographics. *International Journal of Human-Computer Studies*, 82:57–68.

Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 71–80.

Clayton Epp, Michael Lippold, and Regan L Mandryk. 2011. Identifying emotional states using keystroke dynamics. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 715–724.

Denzil Ferreira, Vassilis Kostakos, and Anind K Dey. 2015. Aware: mobile context instrumentation framework. *Frontiers in ICT*, 2:6.

Daniel Fuller, Martine Shareck, and Kevin Stanley. 2017. Ethical implications of location and accelerometer measurement in health research studies with mobile sensing devices. *Social Science & Medicine*, 191:84–88.

Salvatore Giorgi, Sharath Chandra Guntuku, Johannes C Eichstaedt, Claire Pajot, H Andrew Schwartz, and Lyle H Ungar. 2021. Well-being depends on social comparison: Hierarchical models of twitter language suggest that richer neighbors make you less happy. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 1069–1074.

Salvatore Giorgi, David B Yaden, Johannes C Eichstaedt, Lyle H Ungar, H Andrew Schwartz, Amy Kwarteng, and Brenda Curtis. 2023. Predicting us county opioid poisoning mortality from multi-modal social media and psychological self-report data. *Scientific reports*, 13(1):9027.

Sharath Chandra Guntuku, Anneke Buffone, Kokil Jaidka, Johannes C Eichstaedt, and Lyle H Ungar. 2019a. Understanding and measuring psychological stress using social media. In *Proceedings of the international AAAI conference on web and social media*, volume 13, pages 214–225.

Sharath Chandra Guntuku, Rachelle Schneider, Arthur Pelullo, Jami Young, Vivien Wong, Lyle Ungar, Daniel Polsky, Kevin G Volpp, and Raina Merchant. 2019b. Studying expressions of loneliness in individuals using twitter: an observational study. *BMJ open*, 9(11):e030355.

Kokil Jaidka, Salvatore Giorgi, H Andrew Schwartz, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2020. Estimating geographic subjective well-being from twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences*, 117(19):10165–10171.

Oscar NE Kjell, Sverker Sikström, Katarina Kjell, and H Andrew Schwartz. 2022. Natural language analyzed with ai-based transformers predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy. *Scientific reports*, 12(1):3918.

Mingyang Li, Louis Hickman, Louis Tay, Lyle Ungar, and Sharath Chandra Guntuku. 2020. Studying politeness across cultures using english twitter and mandarin weibo. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–15.

Monika N Lind, Michelle L Byrne, Geordie Wicks, Alec M Smidt, and Nicholas B Allen. 2018. The effortless assessment of risk states (ears) tool: An interpersonal approach to mobile sensing. *JMIR Mental Health*, 5(3):e10334.

Tingting Liu, Salvatore Giorgi, Xiangyu Tao, Sharath Chandra Guntuku, Douglas Bellew, Brenda Curtis, and Lyle Ungar. 2023. Different affordances on facebook and sms text messaging do not impede generalization of language-based predictive models. *Proceedings of the International AAAI Conference on Web and Social Media*, 17(1):1153–1157.

Veronica Lynn, Salvatore Giorgi, Niranjan Balasubramanian, and H Andrew Schwartz. 2019. Tweet classification without the tweet: An empirical examination of user versus document attributes. In *Proceedings of the third workshop on natural language processing and computational social science*, pages 18–28.

Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60.

Merijn Mestdagh, Stijn Verdonck, Maarten Piot, Koen Niemeijer, Ghijs Kilani, Francis Tuerlinckx, Peter Kuppens, and Egon Dejonckheere. 2023. m-path: an easy-to-use and highly tailorable platform for ecological momentary assessment and intervention in behavioral research and clinical practice. *Frontiers in Digital Health*, 5.

Jonah Meyerhoff, Tingting Liu, Caitlin A. Stamatis, Tony Liu, Harry Wang, Yixuan Meng, Brenda L. Curtis, Chris J. Karr, Garrick T. Sherman, Pallavi V. Kulkarni, and David C. Mohr. 2023. Analyzing text message linguistic features: Do people with depression communicate differently with their close and non-close contacts? *Behaviour research and therapy*, 166:104342.

Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)*, pages 174–184.

Saif M Mohammad. 2016. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 201–237. Elsevier.

Saif M Mohammad and Svetlana Kiritchenko. 2015. Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, 31(2):301–326.

Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2018. Just-in-time adaptive interventions (jitais) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, 52(6):446–462.

Erik C. Nook, Thomas Derrick Hull, Matthew K. Nock, and Leah H. Somerville. 2021. Linguistic measures of psychological distance track symptom levels and treatment outcomes in a large set of psychotherapy transcripts. *Proceedings of the National Academy of Sciences of the United States of America*, 119.

Jukka-Pekka Onnela, Caleb Dixon, Keary Griffin, Tucker Jaenicke, Leila Minowada, Sean Esterkin, Alvin Siu, Josh Zagorsky, and Eli Jones. 2021. Beiwe: A data collection platform for high-throughput digital phenotyping. *Journal of Open Source Software*, 6(68):3417.

Gregory Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Michal Kosinski, David J Stillwell, Lyle H Ungar, and Martin EP Seligman. 2015. Automatic personality assessment through social media language. *Journal of personality and social psychology*, 108(6):934.

James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.

Youngseo Son, Sean AP Clouston, Roman Kotov, Johannes C Eichstaedt, Evelyn J Bromet, Benjamin J Luft, and H Andrew Schwartz. 2023. World trade center responders in their own words: predicting ptsd symptom trajectories with ai-based language analyses of interviews. *Psychological medicine*, 53(3):918–926.

Zhiyi Song, Stephanie M Strassel, Haejoong Lee, Kevin Walker, Jonathan Wright, Jennifer Garland, Dana Fore, Brian Gainor, Preston Cabe, Thomas Thomas, et al. 2014. Collecting natural sms and chat conversations in multiple languages: The bolt phase 2 corpus. In *LREC*, pages 1699–1704. Citeseer.

Caitlin A. Stamatis, Jonah Meyerhoff, Tingting Liu, Zhaoyi Hou, Garrick T. Sherman, Brenda L. Curtis, Pallavi V. Kulkarni, and David C. Mohr. 2022a. The association of language style matching in text messages with mood and anxiety symptoms. *Procedia computer science*, 206:151–161.

Caitlin A. Stamatis, Jonah Meyerhoff, Tingting Liu, Garrick T. Sherman, Harry Wang, Tony Liu, Brenda L. Curtis, Pallavi V. Kulkarni, and David C. Mohr. 2022b. Prospective associations of text-message-based sentiment with symptoms of depression, generalized anxiety, and social anxiety. *Depression and Anxiety*, 39:794 – 804.

M. L. Tlachac and Elke A. Rundensteiner. 2020. Screening for depression with retrospectively harvested private versus public text. *IEEE Journal of Biomedical and Health Informatics*, 24:3326–3332.

M. L. Tlachac, Avantika Shrestha, Mahum Shah, Benjamin R. Litterer, and Elke A. Rundensteiner. 2022. Automated construction of lexicons to improve depression screening with text messages. *IEEE Journal of Biomedical and Health Informatics*, 27:2751–2759.

John Torous, Hannah Wisniewski, Bruce Bird, Elizabeth Carpenter, Gary David, Eduardo Elejalde, Dan Fulford, Synthia Guimond, Ryan Hays, Philip Henson, et al. 2019. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. *Journal of Technology in Behavioral Science*, 4:73–85.

Julio Vega, Meng Li, Kwesi Aguillera, Nikunj Goel, Echhit Joshi, Kirtiraj Khandekar, Krina C Durica, Abhineeth R Kunta, and Carissa A Low. 2021. Reproducible analysis pipeline for data streams: open-source software to process data collected with mobile devices. *Frontiers in Digital Health*, 3:769823.

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45:1191–1207.