DYMOND: Identifying the Seams for Human-in-the-Loop Digital Phenotyping for Mental Health

Tunwa Tongtawee*
tongtawee.t@northeastern.edu
Northeastern University
Boston, Massachusetts, USA

Olivia Lin Wang*
owang388@gmail.com
Northeastern University
Boston, Massachusetts, USA

Joyce Hsu
joyce.hsu028@gmail.com
Northeastern University
Boston, Massachusetts, USA

Qiuyue "Joy" Zhong
zhong.qiuy@northeastern.edu
Northeastern University
Boston, Massachusetts, USA

Nicholas C. Jacobson
v.mishra@northeastern.edu
Dartmouth College
Hanover, New Hampshire, USA

Varun Mishra
v.mishra@northeastern.edu
Northeastern University
Boston, Massachusetts, USA

Vedant Das Swain†
v.dasswain@northeastern.edu
Northeastern University
Boston, Massachusetts, USA

Abstract

Passively sensing daily behaviors is a promising approach for digital phenotyping as it helps precisely record mental health challenges continuously and unobtrusively. While the selling point of the passive approach is its seamlessness, we argue that this very quality makes it impersonal, untrustworthy, and disengaging. In response, our ongoing study explores a *seamful* approach, allowing users to interact with their data and influence the underlying models of their mental health. We developed DYMOND, an interface that integrates users into the passive-sensing pipeline, enabling them to configure the model through configurations. Our formative study involves users monitoring their depression and collaborating with experts to contextualize mental health models to their needs. This research aims to prototype human-in-the-loop pipelines that protect user data, provide agency, and increase transparency, fostering a collaborative form of digital phenotyping.

CCS Concepts

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; • **Applied computing →** *Law, social and behavioral sciences*; *Psychology*.

Keywords

Digital phenotyping, Passive sensing, Behavioral mental health, Human-in-the-loop systems, User engagement, Depression

*Authors contributed equally
†Corresponding author

CHI '25 Workshop on Envisioning the Future of Interactive Health, Yokohama, Japan

ACM Reference Format:
Tunwa Tongtawee, Olivia Lin Wang, Joyce Hsu, Qiuyue "Joy" Zhong, Nicholas C. Jacobson, Varun Mishra, and Vedant Das Swain. 2025. DYMOND: Identifying the Seams for Human-in-the-Loop Digital Phenotyping for Mental Health. In *Proceedings of (CHI '25 Workshop on Envisioning the Future of Interactive Health)*. ACM, New York, NY, USA, 5 pages.

1 Introduction

For people facing behavioral mental health challenges, passive sensing presents a new way to leverage their smart devices to generate data that might complement clinical visits and, thus, provide greater insight into their state [18]. This approach provides an opportunity to automatically, continuously and unobtrusively collect large volumes of multimodal data over long periods because it is *seamless*—requires negligible effort from users who continue their daily routine. Yet, long-term adherence to passive sensing tools for behavioral mental health remains a challenge. One of the barriers to the adoption of—and compliance with—these tools is a user's inability to engage with the unperceivable monitoring and inference, which further leads to concerns of privacy, trust, and lack of interest [23]. To mitigate the usability barriers and ethical tensions arising from the impersonal nature of seamless sensing, our study explores a *seamful* approach to passive sensing for digital phenotyping.

We explore a paradigm where users have the opportunity to interact with their data once it is generated so that they can dictate changes upstream to model their health risks. To this end, we developed an interface called DYMOND (DYnamic MONitoring for Depression). This interface situates the user further into the typical passive-sensing machine-learning pipeline by inviting them to evaluate and configure the underlying ML model through configurations, or "seams". This process helps qualify the otherwise data-centric algorithmic estimation of their mental health and highlights several unseen factors that traditional sensing ignores. We are conducting a formative study where people with recent experiences of depression are passively monitoring their mental health in-the-wild, and they periodically collaborate with domain experts to interpret and contextualize the model of their mental health.

The study positions us (i) to characterize users' varying preferences for abstracting their passively generated behavioral data with configurations or "seams", and (ii) to inform the design of tools through which users can articulate these preferences back to the digital phenotyping systems . Our findings will help prototype novel human-in-the-loop pipelines for digital phenotyping that protect users' data, provide more agency, and increase transparency. We envision that our findings to enable new automatic and interactive feedback loops between patients, developers, and clinicians to support a collaborative form of digital phenotyping of behavioral health with passive data.

DYMOND extends the idea that digital phenotyping is patient-centered when patients are the "ultimate gatekeepers" of their data [9]. We interpreted this directive beyond simple user consent. Instead, we want to study how users can dictate the way their data is modeled by AI. This article will exhibit the DYMOND dashboard and present our collaborative passive-sensing research process to elicit further discussion on human-in-the-loop digital health.

2 Background

Passive sensing for digital phenotyping has shown promise in supporting anxiety-related disorders [11], depression [24], and, more recently, for substance use [4]. While patients find value in getting digital insights with negligible effort [17], typical approaches to sensing struggle to retain patients as these are less personal, less intelligible, and more invasive [2]. Once a patient consents to data use in digital phenotyping, they have little clarity on what happens with their data next. The underlying machine-learning algorithms might model unreliable data or even capture sensitive context [14, 21]. These limitations stem from the "top-down" approach of digital phenotyping where clinicians' and developers' assertions supersede the patient's preferences [15]. Users of these systems foster anxieties of uncertain self-presentation in data, which is further muddled with the tension of compromising their privacy [1]. The lack of agency not only raises ethical tensions, but also questions the utility of personalized monitoring.

Patients are known to discontinue treatment when they experience lack of trust, connection, and mutual respect between themselves and clinicians [3]. Existing digital phenotyping paradigms neither give patients any feedback on how their data will be modeled, nor do they give patients an opportunity to give feedback to the model. One way to mitigate this is by moving the patient closer to the locus of control for their data [8]. But, how might systems actually facilitate such an interaction? Recent studies show that users of passive sensing can gain better insight into their health when they see explanations of how a digital phenotyping model works [12]. Similarly, visual progress tracking of behaviors can improve patient compliance [25]. Our method, DYMOND aims to go one step a step further, where patients will not only view their data, but they will be able to participate in how the model learns from their data.

A metaphor for our approach is an emerging clinical practice, where patients and clinicians collaboratively document the visit notes to increase transparency and in turn improve patient trust and adherence to treatment [19]. To implement such an interaction, we borrow the design concept of "seamfulness" [5] to enable users to define boundaries, constraints, or "seams" around their data and the inference model and, thus, gain more agency. As a result, DYMOND takes a "bottom-up" approach to learn how users might want to present themselves for digital phenotyping.

3 Method

We recruited participants for a 6-week study to collect passive sensing data. The study consists three two-week periods, each interleaved with check-in sessions for users to collaborate with researchers and interact with DYMOND. We leveraged the open-sourced data-collection platform AWARE framework [7] to build a smartphone application to passively record users' physical activity, locations, and phone usage. Additionally, participants received a Garmin wearable to help DYMOND retrieve users' sleep patterns. At the time of writing, we have recruited 25 people, 11 of whom have already completed the study. Our recruitment criteria required users to possess a smartphone and have screened for moderate or severe depression based in the past two weeks ($\geq$ 10 on PHQ-9 [13]).

The study involves different aspects of passive, active, and collaborative monitoring of a users' mental health:

- **Full study (passive):** Participants continue their daily routines while our data-logger records their activity.
- **End-of-day (active):** Participants complete an Ecological Momentary Assessment delivered via the phone app to report their depression (PHQ8 [16]) and affect (PANAS [22]).
- **Morning (active):** Participants review DYMOND's depression level estimate (for past days), their behavior patterns, and journal their reflections on how DYMOND is correctly (or incorrectly) representing their depression severity.
- **End-of-period (collaborative):** Researchers interview participants at the end of two-weeks to understand their journals and assist them to configure DYMOND for the next period.

3.1 Technology probe to highlight models' *seams*

DYMOND is essentially technology probe testing users' preferred data configurations for algorithmic estimation of depression. It is intended to understand users, not to accurately forecast depression. Estimates for the first 2 weeks are based on a heuristic model that derives feature weights and values from existing research [10, 24], for example, +ve: phone unlock at night is indicative of high depression levels; -ve: healthy sleep or walking at morning is indicative of low levels. For later weeks, the heuristics were personalized to participant data collected from the first two weeks. Notably, DYMOND is less confident in predicting high levels of depression, but more confident in predicting low levels. Moreover, DYMOND only estimates depression for past days, so that users can evaluate its effectiveness by comparing it to their self-report.

We drew from prior work [6, 20] to design a prototype data-visualization dashboard. This dashboard has 3 key components:

- **Behavior visualization:** The online dashboard shows past behavioral patterns, DYMOND's estimation of depression, and user's self-reported estimates (right pane in Figure 1). We opted for a minimalist visualization that uses a string of dots to represent each activity across three epochs of the users'

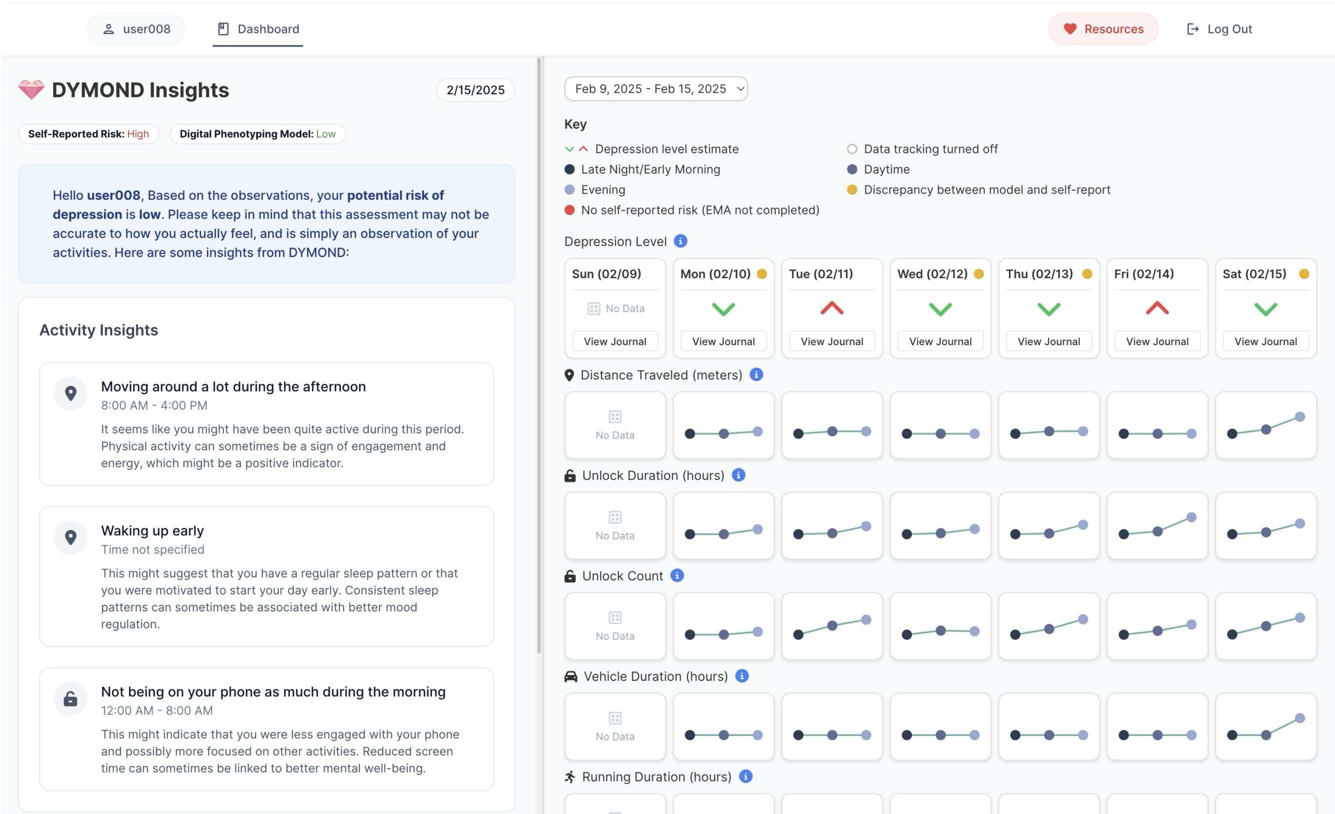

Figure 1: DYMOND allows users to retroactively evaluate how digital phenotyping models might represent their mental state. (Right pane) A minimalist visualization for users to view how DYMOND estimated their depression and how they behaved on those days. (Left pane) For the latest day, DYMOND explains the rationale behind its estimate. In this particular example, DYMOND's estimate did not match what as self-reported by the user—denoted by the yellow dot on the visualization and the self-reported "high" label on top of the insights.

day ("late night/ early morning": 12am–8am; "daytime": 8am–4pm; "evening": 4pm–12am). The top row simply indicates how DYMOND estimated the depression level of the user. A yellow dot is used to signify days when DYMOND's estimate did not match the users' report.

Insights and journaling: To elicit participant reflection, DYMOND provides insight into how user behaviors contribute to their estimated depression. These insights are based on DYMOND's heuristic model. GPT4o articulates the top features in the form of natural language (left pane in Figure 1). Participants can reflect on this information and journal how the model might be (mis)interpreting them (Figure 2).

Configuration panel: The configuration panel can be accessed every two-weeks under the supervision of a researcher. This panel allows users to configure DYMOND's tracking to better align with their actual behavior. In the current iteration, users only have the option to select certain activities that DYMOND should ignore when estimating their depression. Subsequent estimates will reflect their configurations.

3.2 Collaboratively defining seams

Every two weeks, participants check-in with researchers over a virtual interview. This session involves two segments:

Understanding user needs and personalizing configurations: We use the visualization and journal notes to surface which elements of their experience the model is ignoring, misinterpreting, misrepresenting, or even overemphasizing. These 1-on-1 semi-structured sessions help guide participants to define configurations to make DYMOND's model align better to their actual behavior. In additon, it also emphasizes the limitations of the current probe and highlights the needs that we need to design for.

Co-designing new affordances for configurations: Arguably, our probe only allows a limited means to interact with the model or teach it. To elicit new means to improve configurations, we conduct a co-design exercise with participants. In each check-in, we only focus on one of the three components of DYMOND. Prior to the interviews, the authors brainstormed and designed alternative interfaces of different aspects of DYMOND. When this segment begins, we show these variations to participants to invigirate their thoughts. They are tasked to come up with 2-3 designs of their own

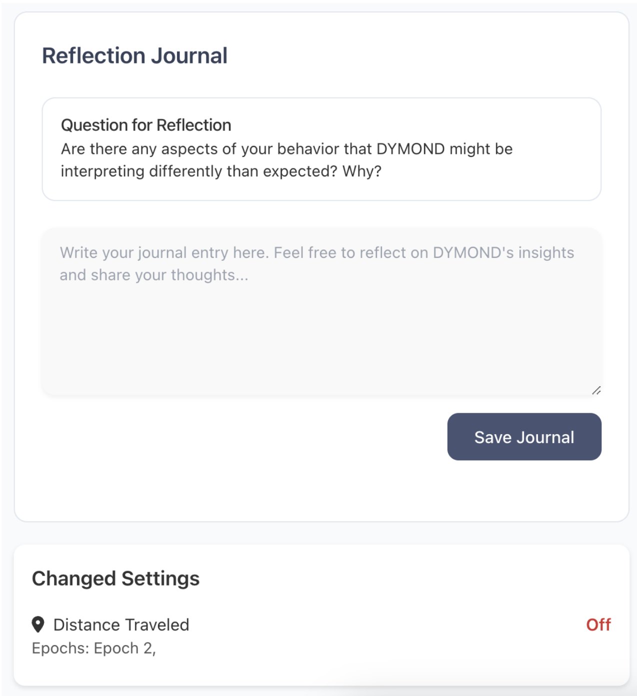

Figure 2: DYMOND purposefully invokes the user to frequently note and express where they disagree with the model or how the model might be incomplete.

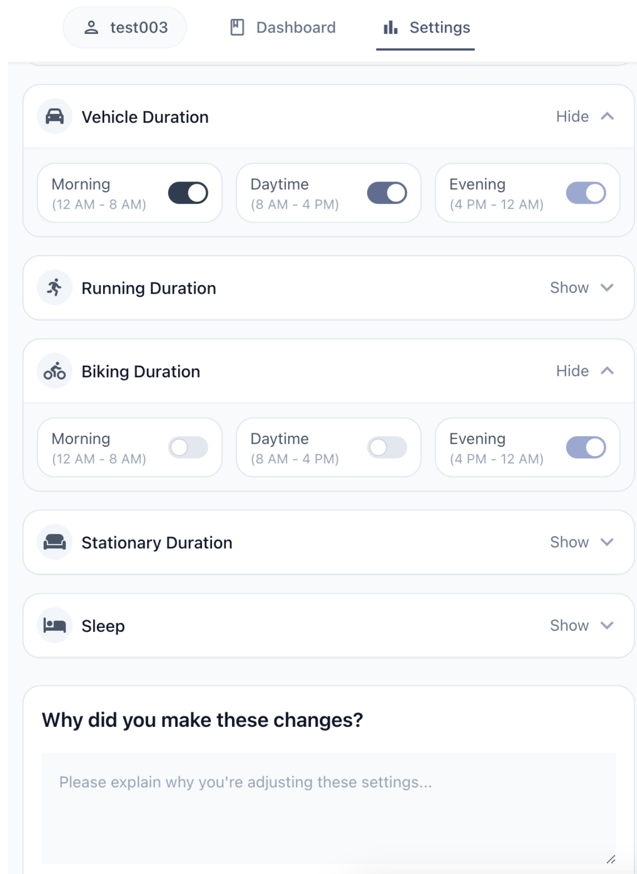

Figure 3: DYMOND provides affordances for users to personalize the model by configuring it. This probe only allows granularly enabling and disabling certain features, but it seeds the idea of potential new interactions.

through rapid prototyping. Then, they describe and expand these designs with the researchers' suggestions.

We repeat this interview process over multiple weeks to provide participants opportunities to revise their configurations as the app gathers more data. After 6 weeks, we conduct an exit interview to learn about participants' overall experiences in co-designing the configurations and their vision of an interactive interface that balances the seamless nature of passive sensing with a seamful human-in-the-loop approach.

Our initial working assumption is to allow 3 kinds of constraints in increasing level of granularity: (i) limiting specific sensors, (ii) limiting certain features within a sensor, and (iii) limiting periods when a feature is computed.

4 Expected Outcomes

Through our study, we aim to learn where digital phenotyping of personal mental health fails, in other words, where do its *seams* lie. We will operationalize common patterns in the choice of configurations between participants, and also highlight discrete variations unique to participants. For instance, we will characterize the motivations behind configurations, are they based on privacy (data is sensitive), reliability (data is not tracked correctly), utility (data is not correctly associated with depression state), or something else? We will also chart how users' configurations evolve over time. Eventually, we hope to arrive at an ontology to abstract passively-sensed behavioral data and align it with the needs of a patient.

Additionally, we aim to compile a set of practical guidelines to design interactions that help users navigate these seams in digital phenotyping. We will study participants' experiences in understanding, conceiving, and articulating constraints through our prototype. The artifacts produced by the co-design activity will aid us in determining opportunities to design interfaces or workflows that can scale over large patient pools without overburdening experts.

5 Conclusion

Our study emphasizes the importance of user engagement and data transparency in advancing human-centered digital phenotyping, particularly for mental health. By integrating users into the data collection and analysis process through the DYMOND interface, we aim to address many implicit concerns of existing passive sensing paradigms. This work is a step towards making digital health more interactive, transparent, and responsive to individual needs.

Acknowledgments

Our research is supported by funds from NIH National Institute of Drug Abuse under award number NIH/NIDA P30DA029926 (through the Center for Technology and Behavioral Health at Dartmouth College), the CORE program at Northeastern University, and the Khoury Apprenticeship program.

References

[1] Daniel A Adler, Emily Tseng, Khatiya C Moon, John Q Young, John M Kane, Emanuel Moss, David C Mohr, and Tanzeem Choudhury. 2022. Burnout and the quantified workplace: Tensions around personal sensing interventions for stress in resident physicians. *Proceedings of the ACM on Human-computer Interaction* 6, CSCW2 (2022), 1–48.

[2] Hege K Andreassen, Marianne Trondsen, Per Egil Kummervold, Deede Gammon, and Per Hjortdahl. 2006. Patients who use e-mediated communication with their doctor: new constructions of trust in the patient-doctor relationship. *Qualitative health research* 16, 2 (2006), 238–248.

[3] Samuel A Ball, Kathleen M Carroll, Monica Canning-Ball, and Bruce J Rounsaville. 2006. Reasons for dropout from drug abuse treatment: Symptoms, personality, and motivation. *Addictive behaviors* 31, 2 (2006), 320–330.

[4] Cynthia I Campbell, Ching-Hua Chen, Sara R Adams, Asma Asyyed, Ninad R Athale, Monique B Does, Saeed Hassanpour, Emily Hichborn, Melanie Jackson-Morris, and Nicholas C Jacobson. 2023. Patient engagement in a multimodal digital phenotyping study of opioid use disorder. *Journal of Medical Internet Research* 25 (2023), e45556.

[5] Matthew Chalmers. 2003. Seamful design and ubicomp infrastructure. In *Proceedings of Ubicomp 2003 workshop at the crossroads: The interaction of HCI and systems issues in Ubicomp*. 577–84.

[6] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. 2017. Understanding self-reflection: how people reflect on personal data through visual data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 173–182.

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

[8] Erik Grönvall and Nervo Verdezoto. 2013. Beyond self-monitoring: understanding non-functional aspects of home-based healthcare technology. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. 587–596.

[9] Kit Huckvale, Svetha Venkatesh, and Helen Christensen. 2019. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. *NPJ digital medicine* 2, 1 (2019), 1–11.

[10] Nicholas C Jacobson and Yeon Joo Chung. 2020. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. *Sensors* 20, 12 (2020), 3572.

[11] Nicholas C Jacobson and Brandon Feng. 2022. Digital phenotyping of generalized anxiety disorder: using artificial intelligence to accurately predict symptom severity using wearable sensors in daily life. *Translational Psychiatry* 12, 1 (2022), 336.

[12] Taewan Kim, Haesoo Kim, Ha Yeon Lee, Hwarang Goh, Shakhboz Abdigapporov, Mingon Jeong, Hyunsung Cho, Kyungsik Han, Youngtae Noh, Sung-Ju Lee, et al. 2022. Prediction for retrospection: Integrating algorithmic stress prediction into personal informatics systems for college students' mental health. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–20.

[13] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. 2001. The phq-9: validity of a brief depression severity measure. *Journal of general internal medicine* 16, 9 (2001), 606–613.

[14] Varun Mishra, Sougata Sen, Grace Chen, Tian Hao, Jeffrey Rogers, Ching-Hua Chen, and David Kotz. 2020. Evaluating the reproducibility of physiological stress detection models. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 4, 4 (2020), 1–29.

[15] David C Mohr, Ken R Weingardt, Madhu R Reddy, and Stephen M Schueller. 2017. Three problems with current digital mental health research... and three things we can do about them. *Psychiatric services* 68, 5 (2017), 427–429.

[16] Matthew D Nemesure, Amanda C Collins, George D Price, Tess Z Griffin, Arvind Pillai, Subigya Nepal, Michael V Heinz, Damien Lekkas, Andrew T Campbell, and Nicholas C Jacobson. 2024. Depressive symptoms as a heterogeneous and constantly evolving dynamical system: Idiographic depressive symptom networks of rapid symptom changes among persons with major depressive disorder. *Journal of psychopathology and clinical science* 133, 2 (2024), 155.

[17] Emily Seltzer, Jesse Goldshear, Sharath Chandra Guntuku, Dave Grande, David A Asch, Elissa V Klinger, and Raina M Merchant. 2019. Patients' willingness to share digital health and non-health data for research: a cross-sectional study. *BMC Medical Informatics and Decision Making* 19, 1 (2019), 1–8.

[18] Claire F Snyder, Albert W Wu, Robert S Miller, Roxanne E Jensen, Elissa T Bantug, and Antonio C Wolff. 2011. The role of informatics in promoting patient-centered care. *Cancer Journal (Sudbury, Mass.)* 17, 4 (2011), 211.

[19] Victoria Stanhope, Chuck Ingoglia, Bill Schmelter, and Steven C Marcus. 2013. Impact of person-centered planning and collaborative documentation on treatment adherence. *Psychiatric Services* 64, 1 (2013), 76–79.

[20] Poorna Talkad Sukumar, Gonzalo J Martinez, Ted Grover, Gloria Mark, Sidney K D'Mello, Nitesh V Chawla, Stephen M Mattingly, and Aaron D Striegel. 2020. Characterizing exploratory behaviors on a personal visualization interface using interaction logs. *EuroVis 2020-Short Papers* (2020).

[21] Vedant Das Swain, Victor Chen, Shrija Mishra, Stephen M Mattingly, Gregory D Abowd, and Munmun De Choudhury. 2022. Semantic gap in predicting mental wellbeing through passive sensing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.

[22] Edmund R Thompson. 2007. Development and validation of an internationally reliable short-form of the positive and negative affect schedule (panas). *Journal of cross-cultural psychology* 38, 2 (2007), 227–242.

[23] John Torous, Jennifer Nicholas, Mark E Larsen, Joseph Firth, and Helen Christensen. 2018. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evidence-based mental health* 21, 3 (2018), 116.

[24] Rui Wang, Weichen Wang, Alex DaSilva, Jeremy F Huckins, William M Kelley, Todd F Heatherton, and Andrew T Campbell. 2018. Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–26.

[25] Katie M White, Ewan Carr, Daniel Leightley, Faith Matcham, Pauline Conde, Yatharth Ranjan, Sara Simblett, Erin Dawe-Lane, Laura Williams, and Claire Henderson. 2024. Engagement with a remote symptom-tracking platform among participants with major depressive disorder: Randomized controlled trial. *JMIR mHealth and uHealth* 12, 1 (2024), e44214.