# Designing Clinician Centered Explanations for Sensing-based Longitudinal Mental Health Monitoring

Esther Kang
KAIST
Daejeon, South Korea
esther.kang@kaist.ac.kr

Youngji Koh
KAIST
Daejeon, South Korea
youngji@kaist.ac.kr

Seungwan Jin
KAIST
Daejeon, South Korea
seungwan@kaist.ac.kr

Uichin Lee
KAIST
Daejeon, South Korea
uclee@kaist.ac.kr

## Abstract

Passive sensing and experience sampling increasingly capture how patients behave between psychiatric visits, and large language models can turn these heterogeneous signals into clinical interpretations. Yet in outpatient care, such interpretations are useful only when clinicians can make sense of them alongside other evidence. Under tight time constraints, clinicians must judge their clinical relevance, verify them against the underlying data, and decide what to explore next. It remains unclear how AI-generated interpretations should be explained to support this reasoning rather than merely justify model outputs. Drawing on semi-structured interviews with 12 psychiatrists, we derive three requirements for explanation design in longitudinal mental health monitoring: rapid understanding within limited consultation time, evidence-grounded clinical interpretation, and support for clinical inquiry. We operationalize these requirements in an interactive dashboard that surfaces the behavioral evidence behind each interpretation, preserves competing interpretations, and carries remaining uncertainty forward as follow-up questions for the consultation. Together, these offer a concrete design point for explaining longitudinal behavioral data to clinicians.

## CCS Concepts

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; *Ubiquitous and mobile computing design and evaluation methods*; • **Applied computing** → Health informatics.

## Keywords

Ubiquitous Computing; Longitudinal Sensing; Patient-Generated Health Data; Mental Health; Clinical Decision Support; Psychiatric Workflows; Human-AI Collaboration

## 1 Introduction

Wearables, smartphone-based experience sampling methods (ESM), and passive behavioral sensing now enable continuous observation of patients' daily lives between psychiatric visits [8, 9, 13]. Such patient-generated health data (PGHD) capture longitudinal changes in sleep, mobility, social interaction, and affective states that are difficult to reconstruct through retrospective recall alone [12, 14]. Recent large language models (LLMs) can further integrate these heterogeneous signals into clinically meaningful interpretations[3, 7].

Such interpretations become useful, however, when clinicians can make sense of them alongside other evidence. Because self-reported symptoms and behavioral observations each provide only a partial picture of a patient's condition [1, 6, 10], clinicians need to examine where these signals reinforce or diverge and what remains to be explored during the consultation.

Designing such explanations is difficult. Too little information makes AI-generated interpretations hard to evaluate, whereas too much detail can overwhelm clinicians during time-constrained outpatient encounters [2, 5]. Prior work addresses this trade-off through progressive disclosure, presenting summaries first and revealing detail on demand [4, 11]. Yet how explanations should support clinicians in navigating clinically relevant findings, underlying evidence, and unresolved uncertainty in longitudinal mental health monitoring remains less clear. To investigate this question, we interviewed 12 psychiatrists and identified three requirements for explanation design: rapid understanding, evidence-grounded interpretation, and support for clinical inquiry. We operationalize these requirements in an interactive dashboard.

## 2 Design Rationale

We conducted semi-structured interviews with 12 psychiatrists recruited from university hospitals and clinics across South Korea (4 female; age $M$=44.5, $SD$=6.43) during the development of the underlying system. The interviews examined clinicians' reasoning and information needs during psychiatric follow-up assessments, including how AI-generated interpretations should be presented during outpatient consultations.

Two researchers independently open-coded the transcripts and iteratively grouped recurring codes into higher-level themes. Through discussion and refinement, we identified three explanation design requirements (DRs); (1) **rapid understanding** within limited consultation time, (2) **evidence-grounded clinical interpretation**, and (3) **support for clinical inquiry**.

### 2.1 DR1. Rapid Understanding within Limited Consultation Time

Most participants recognized the value of longitudinal sensing data for understanding changes between outpatient visits, but emphasized that their volume and heterogeneity made comprehensive

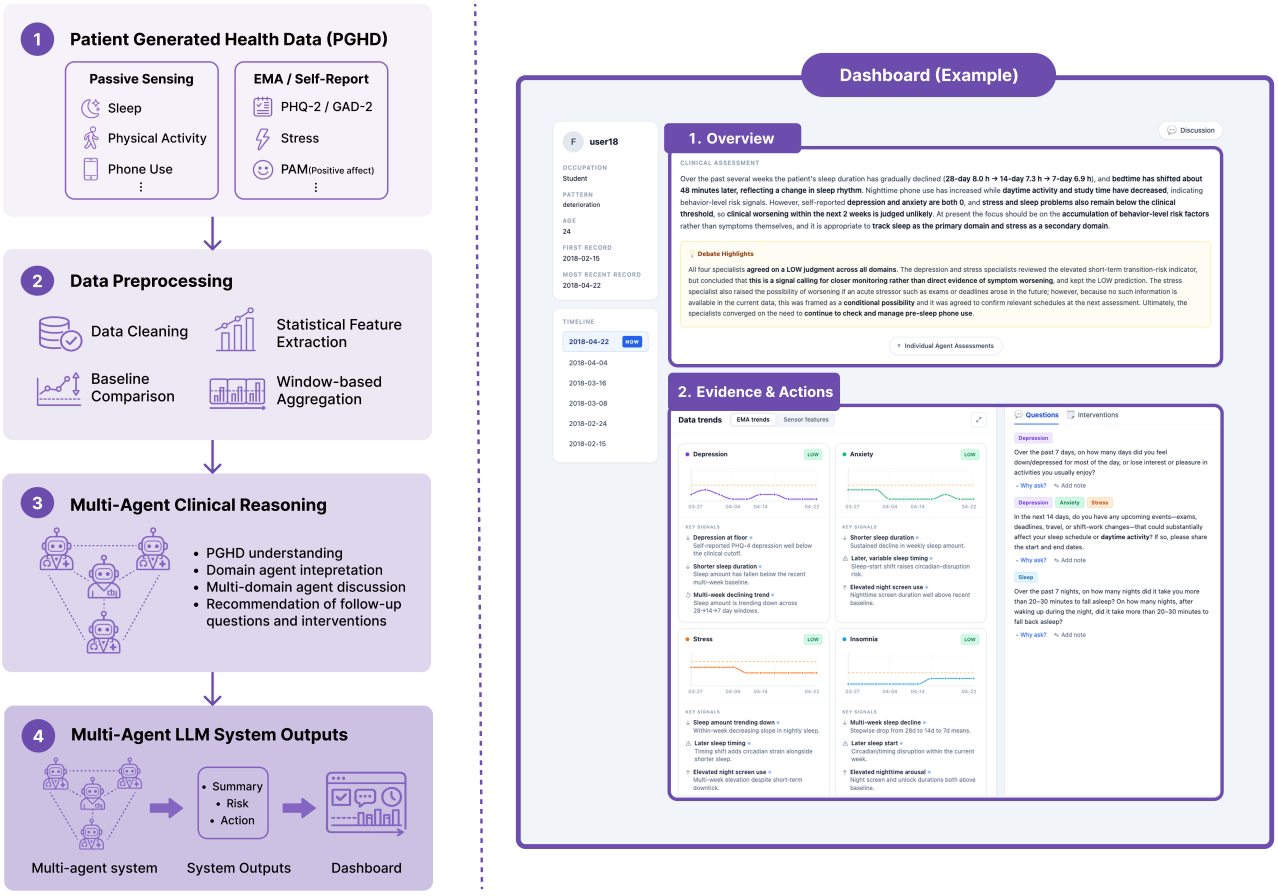



**Figure 1: Overview of the proposed system. (Left) The pipeline processes PGHD from passive sensing and EMA through data preprocessing and multi-agent clinical reasoning to produce structured outputs. (Right) The explanation interface presents these structured outputs through the Overview panel (behavioral summaries and trend cards) and the Evidence & Actions panel (longitudinal trends, follow-up questions, and intervention suggestions).**

review impractical during brief consultations. Rather than manually inspecting numerous behavioral signals, clinicians preferred AI to prioritize clinically relevant findings. Explanations should therefore foreground these findings while preserving access to supporting detail.

## 2.2 DR2. Evidence-grounded Clinical Interpretation

Several participants emphasized examining the reasoning behind an AI-generated assessment rather than prediction accuracy alone. Even when a prediction was incorrect, they considered the system useful if its reasoning remained transparent and open to clinical evaluation. This was particularly important when behavioral signals suggested competing interpretations, as clinicians wanted to examine the evidence behind each perspective before forming their own judgment.

## 2.3 DR3. Support for Clinical Inquiry

Participants valued when the system identified unresolved uncertainties and translated them into follow-up questions for patient conversation. They also viewed intervention suggestions grounded in patients' behavioral patterns as useful for connecting AI-generated interpretations to subsequent clinical care.

# 3 System Design

## 3.1 Underlying Multi-Agent System

The explanation interface is built on a multi-agent clinical reasoning system that synthesizes longitudinal passive sensing and ESM data into clinical interpretations. As shown in Figure 1 (left), multiple domain-specific agents each reason about a particular aspect of the patient's condition, and a moderator agent synthesizes their complementary assessments into an overall clinical assessment together with follow-up questions and intervention recommendations. These outputs are then organized in the explanation interface shown in Figure 1 (right) for use during outpatient consultations.




**Figure 2: Interface features supporting (A) rapid understanding through the synthesized clinical assessment and progressively disclosed supporting rationale, and (B) evidence-grounded clinical interpretation through longitudinal data trends and supporting and counter evidence.**

## 3.2 Explanation Interface

The interface arranges the outputs into two complementary panels. The **Overview** panel summarizes the synthesized clinical assessment, with debate highlights and access to individual agent assessments and the full discussion. The **Evidence & Actions** panel presents the underlying behavioral data through interactive trend visualizations, alongside the system-generated follow-up questions and intervention recommendations. Presenting interpretations next to their supporting evidence lets clinicians inspect both the synthesized assessment and the information from which it was derived.

The interface operationalizes the three design requirements through complementary explanation elements. For **rapid understanding**, the interface leads with the synthesized **Clinical Assessment** (Figure 2), allowing clinicians to grasp the prioritized interpretation first. Supporting rationale, including **Debate Highlights**, **Individual Agent Assessments**, and the **Discussion** view,

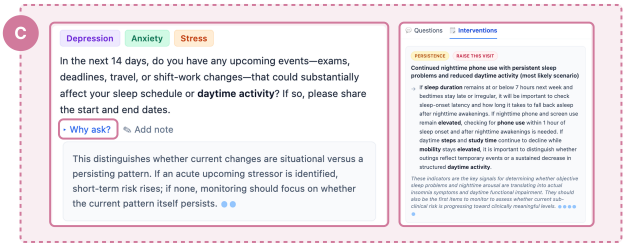



**Figure 3: Support for clinical inquiry through follow-up questions and intervention suggestions grounded in unresolved findings.**

is progressively disclosed, allowing clinicians to access additional context when needed.

For **evidence-grounded clinical interpretation**, the **Data Trends** view keeps longitudinal sensing and self-reported data directly inspectable across configurable time windows. The **Discussion** view presents the individual claims raised during the multi-agent discussion and allows clinicians to inspect the evidence associated with each claim. By hovering over a claim, clinicians can view the specific supporting and counter evidence used in the discussion, making it possible to examine why a particular interpretation was proposed and what evidence may challenge it (Figure 2). Rather than resolving competing signals into a single conclusion, the interface preserves multiple evidence-supported interpretations, so that clinicians can trace and evaluate the reasoning behind each one independently.

For **support for clinical inquiry**, the **Questions & Interventions** view carries unresolved findings forward as follow-up questions, each paired with a **Why ask?** explanation that surfaces the uncertainty motivating it, alongside intervention suggestions grounded in the patient's behavioral patterns (Figure 3). Rather than ending at an AI-generated interpretation, the interface turns what remains uncertain into concrete directions for patient discussion and clinical management.

## 4 Discussion and Conclusion

Our findings suggest that explanations for longitudinal mental health monitoring should do more than justify AI-generated outputs. They should help clinicians quickly identify clinically relevant information, evaluate interpretations against their underlying evidence, and act on remaining uncertainty by carrying it into clinical inquiry. Rather than always resolving heterogeneous signals into a single, definitive account, explanations may benefit from preserving supporting and conflicting evidence. Preserving supporting and conflicting evidence can instead enable clinicians to evaluate alternative interpretations in light of their own clinical judgment. This extends the role of explanation beyond making an AI-generated output understandable toward supporting the clinician's ongoing sensemaking and inquiry during the consultation.

This work has two main limitations, each pointing to future work. First, the interface explains AI-generated interpretations but does not yet represent the data quality, missingness, and contextual limitations of the underlying sensing data. We plan to investigate explanation strategies that communicate such sensing uncertainty explicitly. Second, the interface was demonstrated with longitudinal data from a university student population, leaving its applicability across broader psychiatric settings to be examined. Future work will evaluate the interface with practicing psychiatrists using data from more diverse patient populations. Beyond these, we plan to explore interactive mechanisms that let clinicians further question and refine AI-generated interpretations during the consultation.

## Acknowledgments

This research was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (RS-2025-02305801).

## References

[1] Daniel A Adler, Yuewen Yang, Thalia Viranda, Xuhai Xu, David C Mohr, Anna R Van Meter, Julia C Tartaglia, Nicholas C Jacobson, Fei Wang, Deborah Estrin, et al. 2024. Beyond detection: Towards actionable sensing research in clinical mental healthcare. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 8, 4 (2024), 1–33.

[2] Adrian Bussone, Simone Stumpf, and Dympna O'Sullivan. 2015. The role of explanations on trust and reliance in clinical decision support systems. In *2015 international conference on healthcare informatics*. IEEE, 160–169.

[3] Akshat Choube, Ha Le, Jiachen Li, Kaixin Ji, Vedant Das Swain, and Varun Mishra. 2025. GLOSS: Group of LLMs for open-ended sensemaking of passive sensing data for health and wellbeing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 3 (2025), 1–32.

[4] Michael Chromik and Andreas Butz. 2021. Human-XAI interaction: a review and design principles for explanation user interfaces. In *IFIP Conference on Human-Computer Interaction*. Springer, 619–640.

[5] Jeffrey N Clark, Matthew Wragg, Emily Nielsen, Miquel Perello-Nieto, Nawid Keshtmand, Michael Ambler, Shiv Sharma, Christopher P Bourdeaux, Amberly Brigden, and Raul Santos-Rodriguez. 2024. Exploring the requirements of clinicians for explainable AI decision support systems in intensive care. *arXiv preprint arXiv:2411.11774* (2024).

[6] 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.

[7] Zachary Englhardt, Chengqian Ma, Margaret E Morris, Chun-Cheng Chang, Xuhai" Orson" Xu, Lianhui Qin, Daniel McDuff, Xin Liu, Shwetak Patel, and Vikram Iyer. 2024. From classification to clinical insights: Towards analyzing and reasoning about mobile and behavioral health data with large language models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–25.

[8] David C Mohr, Mi Zhang, and Stephen M Schueller. 2017. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annual review of clinical psychology* 13 (2017), 23–47.

[9] Inez Myin-Germeys, Zuzana Kasanova, Thomas Vaessen, Hugo Vachon, Olivia Kirtley, Wolfgang Viechtbauer, and Ulrich Reininghaus. 2018. Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry* 17, 2 (2018), 123–132.

[10] Ada Ng, Rachel Kornfield, Stephen M Schueller, Alyson K Zalta, Michael Brennan, and Madhu Reddy. 2019. Provider perspectives on integrating sensor-captured patient-generated data in mental health care. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–25.

[11] Cecilia Panigutti, Andrea Beretta, Daniele Fadda, Fosca Giannotti, Dino Pedreschi, Alan Perotti, and Salvatore Rinzivillo. 2023. Co-design of human-centered, explainable AI for clinical decision support. *ACM Transactions on Interactive Intelligent Systems* 13, 4 (2023), 1–35.

[12] Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, and David C Mohr. 2015. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of medical Internet research* 17, 7 (2015), e4273.

[13] ShiYing Shen, Wenhao Qi, Jianwen Zeng, Sixie Li, Xin Liu, Xiaohong Zhu, Chaoqun Dong, Bin Wang, Yankai Shi, Jiani Yao, et al. 2025. Passive sensing for mental health monitoring using machine learning with wearables and smartphones: scoping review. *Journal of medical Internet research* 27 (2025), e77066.

[14] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. 3–14.