SleepHelp Agent: A Clinician-in-the-loop Multi-Agent System for Lifelog-Grounded Insomnia Care in Sleep Psychiatry

Seungeun Chung, Hyuntae Jeong, Se Won Oh, Jeong Mook Lim, Kyung Ju Noh, Sun-Kyung Lee
Digital Convergence Research Lab
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
{schung, htjeong, sewonoh, jmlim21, kjnoh, sklee2014}@email.com

ABSTRACT

Insomnia disorder is a prevalent clinical condition closely linked to depression, anxiety, and broader psychiatric morbidity, yet its behavioral and longitudinal nature makes it difficult to assess from interviews and episodic clinical visits alone. We present SleepHelp Agent, a lifelog-grounded, clinician-in-the-loop multi-agent AI system that supports clinicians in sleep psychiatry through structured report generation and interactive Q&A for insomnia care. The system extracts sleep and behavioral features from multimodal lifelog data, grounds interpretations using a shared retrieval-augmented generation (RAG) knowledge base of clinical evidence, and formats outputs using a two-tier schema that separates *observed* findings directly computed from data from *suggested* LLM-generated interpretations for clinician review. We evaluated SleepHelp Agent in a 12-week pilot insomnia care workflow, where a sleep psychiatrist reviewed AI-generated reports before consultations and provided structured post-session feedback. Preliminary feedback suggests that AI-generated reports can help organize complex longitudinal sleep and behavioral data for pre-consultation review, while highlighting several key requirements for clinical use in insomnia care: standardized terminology, stronger personalization, quantitative grounding, and longitudinal association analysis. Our findings position SleepHelp Agent as a clinician-in-the-loop decision support system for data-informed insomnia care and AI-assisted, CBT-I-informed psychological intervention planning.

CCS CONCEPTS

• Applied computing → Health care information systems

• Computing methodologies → Artificial intelligence

• Human-centered computing → Human computer interaction

KEYWORDS

Insomnia care, Multi-agent AI, Retrieval-augmented generation, Clinician-in-the-loop AI, Multimodal lifelogging

The International Workshop on AI for Cognitive and Mental Health Support (AI4Mental), held in conjunction with ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2026, August 9, 2026, Jeju, Korea.

1 Introduction

Insomnia disorder is among the most prevalent mental health conditions in adults and is recognized as a stand-alone diagnosis in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [1]. It is tightly entangled with depression, anxiety, and other psychiatric conditions through well-established bidirectional pathways: chronic insomnia predicts the subsequent onset of major depressive disorder, and disrupted sleep amplifies emotional dysregulation, daytime cognitive impairment, and stress reactivity [2, 3]. Sleep psychiatry and behavioral sleep medicine therefore sit squarely within mental health practice, where clinicians often need to understand how sleep problems relate to daily activity, stress, naps, caffeine intake, irregular routines, and subjective sleep experience. However, conventional clinical assessment still depends heavily on retrospective interviews, self-reported sleep diaries, and episodic clinical visits. These methods are valuable but may miss longitudinal behavioral patterns that unfold across days or weeks in naturalistic settings.

Recent advances in smartphone-based lifelogging, wearable sensing, and home-based sleep sensing create new opportunities to capture sleep and behavior in daily life. Such multimodal lifelog data can reveal patterns that are difficult to observe during a short consultation, such as irregular sleep timing, fragmented sleep, low daytime activity, excessive naps, evening light exposure, stress-related sleep disruption, and discrepancies between objective sleep metrics and subjective sleep quality. These patterns are particularly relevant to cognitive behavioral therapy for insomnia (CBT-I), the first-line non-pharmacological psychological intervention for insomnia disorder [4], where treatment depends on identifying modifiable behavioral and cognitive factors that maintain sleep difficulties.

Despite its potential, lifelog data is difficult to use directly in clinical practice. A single patient may generate weeks of heterogeneous data across wearable sensors, smartphones, and sleep sensing devices. Clinicians must interpret not only individual sleep metrics, but also temporal relationships among sleep, daytime behavior, subjective symptoms, and contextual factors. Without automated support, reviewing these data streams can increase clinical burden. Therefore, there is a need for AI systems that can transform longitudinal lifelog data into clinically interpretable summaries while making the evidential basis and uncertainty of model-generated interpretations explicit for clinician review.

Large language model (LLM)-based agents offer a promising interface for such decision support, but their use in clinical settings requires careful design. Recent studies of medical LLMs highlight risks such as propagation of erroneous clinical information, underscoring the need for grounding, verification, and clinician review [5]. Clinicians need to know which findings are directly observed from patient data, which interpretations are generated by the model, and what clinical evidence supports each suggestion. This distinction is especially important for mental health and behavioral intervention contexts, where AI systems should provide information that clinicians can incorporate into their own decision-making.

In this paper, we present **SleepHelp Agent**, a lifelog-grounded multi-agent AI designed as a clinician-in-the-loop tool for insomnia decision support in mental health practice. SleepHelp Agent is designed to generate structured sleep analysis reports and support interactive Q&A for clinicians before and during insomnia consultations. The system integrates multimodal data from wearable, smartphone, sleep sensing device, and ecological momentary assessment (EMA) sources; extracts clinically relevant sleep and behavioral features; grounds interpretations using a shared retrieval-augmented generation knowledge store; and produces reports that separate data-derived observations from model-generated clinical interpretations.

Our work makes three main contributions. First, we introduce a multi-agent architecture that supports AI-assisted delivery of insomnia care based on CBT-I principles, converting longitudinal lifelog data into structured reports and interactive Q&A for clinician review. Second, we propose a two-tier report schema that separates *observed findings* directly computed from data from *suggested interpretations* grounded in retrieved clinical evidence and presented for clinician review. Third, we report preliminary clinical feedback from a 12-week pilot in a sleep psychiatry workflow, highlighting both the potential of AI-generated sleep analysis reports for pre-consultation preparation and the design requirements needed for safe deployment in mental health care.

The remainder of this paper is organized as follows. Section 2 reviews related work, Section 3 presents the system overview, Section 4 describes the preliminary evaluation, Section 5 discusses the implications and limitations of the study, and Section 6 concludes the paper.

2 Related Work

2.1 Sleep as an Intervention Target

CBT-I is widely recognized as a standard behavioral intervention for chronic insomnia. The American Academy of Sleep Medicine (AASM) clinical practice guideline [4] strongly recommends multi-component CBT-I for adults with chronic insomnia disorder. Importantly for mental health applications, recent work on digital CBT-I has expanded access to insomnia interventions and demonstrated benefits for sleep outcomes and depressive and anxiety symptoms. A meta-analysis of 22 randomized controlled trials reported significant effects of digital CBT-I on insomnia, depression, and anxiety outcomes [6]. Fully automated digital CBT-I has also shown moderate to large effects on insomnia severity across randomized trials, although therapist-supported or hybrid models may remain more beneficial in some settings [7].

SleepHelp Agent builds on this evidence base by supporting the clinical work required to plan and monitor insomnia interventions. While automated digital CBT-I focuses on delivering intervention content directly to patients, SleepHelp Agent instead supports the clinician's pre-consultation reasoning by summarizing patient-specific sleep and behavioral evidence and generating discussion points that clinicians can adapt for individualized intervention planning.

2.2 Multimodal Lifelog for Sleep Care

Wearable and mobile sensing technologies enable longitudinal monitoring of sleep and behavior in daily life, addressing limitations of episodic clinic visits and retrospective self-reports. Prior work has highlighted the potential of wearable technologies for sleep and circadian biomarker development, while also emphasizing the importance of validation and careful interpretation in free-living environments [8]. However, reviews of wearable AI for sleep disorders report that most studies focus on diagnosis or screening with treatment-oriented applications remaining limited [9].

Multimodal approaches are particularly relevant for insomnia and behavioral sleep medicine because sleep complaints often depend on interactions between nighttime sleep, daytime behavior, subjective experience, and environmental context. A prior study combined consumer sleep sensing device, smartphone-based EMA, and phone usage data over two months, showing the feasibility of integrating multimodal data to characterize habitual sleep behavior [10]. Contactless sleep sensing devices such as the Withings Sleep Analyzer have additionally been validated against polysomnography for sleep-wake estimation and sleep apnea detection in naturalistic settings [11, 12].

SleepHelp Agent extends this line of work by translating multimodal lifelog streams into clinically interpretable summaries for use in an insomnia care workflow. Rather than focusing on passive monitoring or label prediction, our system integrates lifelog data to produce structured features, longitudinal summaries, and report elements that can be reviewed by clinicians, shifting the role of lifelog data from retrospective sensing or classification toward actionable clinical documentation and interactive review.

2.3 Grounded AI for Clinical Decision Support

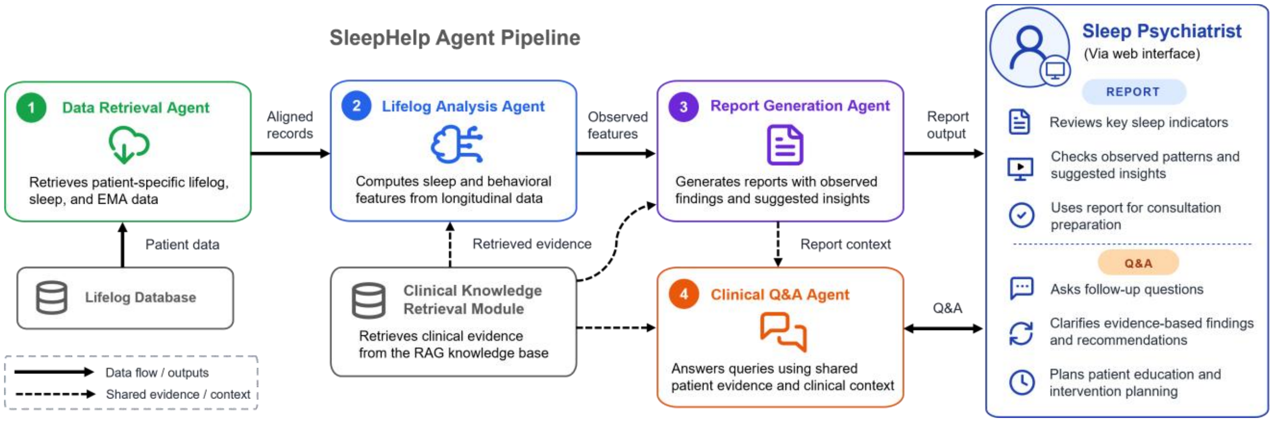

Figure 1. SleepHelp Agent system architecture and evidence-grounded report generation pipeline.

LLMs and agent-based systems have recently been explored for clinical decision support, medical documentation, knowledge retrieval, and workflow optimization. A survey of LLM-based agents in medicine identifies clinical decision support and medical documentation as major application areas, while highlighting multimodal integration, implementation barriers, hallucination management, and ethical considerations as open challenges [13]. Multi-agent clinical decision support systems are also increasingly studied, but systematic reviews note persistent gaps around hierarchical architectures, bidirectional knowledge flow, and explainable AI mechanisms in real-world clinical settings [14]. These gaps are especially consequential for AI tools deployed in mental health workflows [15, 16], where clinician trust and the verifiability of model outputs are central to safe use.

Retrieval-augmented generation (RAG) offers one approach to reducing unsupported generation by grounding LLM outputs in external knowledge sources. Healthcare RAG reviews describe RAG as a method for providing LLMs with external information to improve grounding, while also noting that medical RAG still lacks standardized evaluation practices and must address ethical and safety considerations [17]. Medical RAG benchmarks such as MIRAGE further show that retrieval settings, corpus choice, and retriever LLM combinations can substantially affect performance in medical question answering [18].

In contrast to prior clinical LLM and RAG systems that primarily focus on medical Q&A or documentation, SleepHelp Agent applies these grounding and trust principles to a clinician-in-the-loop workflow in sleep psychiatry, translating longitudinal lifelog data into structured reports and interactive Q&A with explicit separation of observed and inferred content.

3 SleepHelp Agent System Overview

3.1 Multi-agent System Architecture

We adopt a multi-agent architecture to structure the heterogeneous reasoning steps required for lifelog-based sleep decision support. Such reasoning must handle noisy longitudinal data, extract multimodal evidence, ground clinical interpretations in published guidelines, estimate uncertainty in patient-specific findings, and produce safety-aware recommendations. A single end-to-end model may obscure these intermediate steps and make it difficult for clinicians to verify why a recommendation was produced. Therefore, SleepHelp Agent decomposes the workflow into specialized agents, each responsible for a clinically meaningful subtask, so that final recommendations remain grounded in observed evidence and traceable for clinician-in-the-loop review.

The proposed architecture follows a staged pipeline as shown in Figure 1. First, the **Data Retrieval Agent** retrieves lifelog, sleep, and EMA data for a given patient episode. Second, the **Lifelog Analysis Agent** computes observable sleep and behavioral features, including sleep duration and variability, daytime activity, and missing-data indicators. Third, the **Report Generation Agent** synthesizes these lifelog features and clinically grounded interpretations into a standardized clinician report. In parallel, the **Clinical Q&A Agent** supports interactive follow-up questions from clinicians by referencing the same patient-specific evidence, retrieved clinical evidence, and report context. By grounding the report and Q&A interactions in a shared evidence base, the system aims to reduce discrepancies between the written report and subsequent question-answering interactions.

The **Clinical Knowledge Retrieval Module** supplies the Lifelog Analysis Agent, Report Generation Agent, and Clinical Q&A Agent with clinically relevant evidence by dynamically augmenting their prompts. This RAG module indexes consensus sleep medicine guidelines, CBT-I manuals, and peer-reviewed literature as dense vector embeddings. To mitigate clinical mismatches, the retrieved passages are further filtered through a cross-encoder re-ranking layer prior to context integration.

To improve clinical interpretability, SleepHelp Agent formats its outputs using a two-tier schema. *Observed* findings refer to values

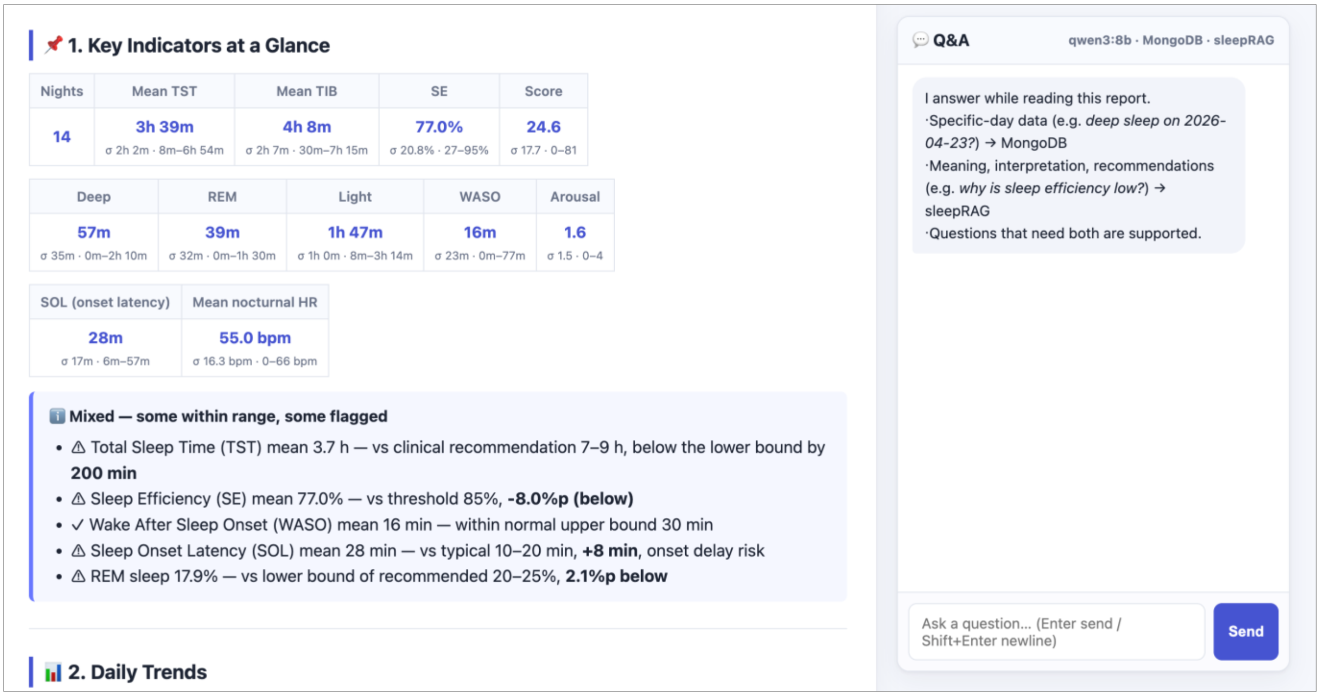

Figure 2. Key-indicator view of the SleepHelp Agent interface.

directly computed from patient data, including sleep and behavioral features, and subjective states from EMA responses. *Suggested* findings refer to LLM-generated clinical interpretations that are supported by retrieved evidence and presented as reviewable suggestions. This separation allows clinicians to verify which findings are directly supported by patient data and which recommendations require clinical judgment.

3.2 Clinician-in-the-loop Workflow

SleepHelp Agent is designed as a clinician-in-the-loop support system for insomnia care, with a focus on pre-consultation report generation and interactive Q&A. The clinician interface presents the generated report and a Q&A panel in a split-screen layout, enabling clinicians to review structured evidence while asking follow-up questions about specific sleep patterns, clinical interpretations, or supporting references.

The report is organized to support rapid clinical review. It includes a narrative summary of the selected monitoring period, tabulated sleep metrics, time-series visualizations of sleep efficiency, wake after sleep onset, and heart rate, two-tier clinical insights with inline citations, recommendation candidates grounded in CBT-I principles, an EMA completion rate indicator, and a reference list. This layout is intended to help clinicians distinguish directly observed lifelog patterns from model-generated interpretations and to trace each recommendation back to patient-specific data or retrieved clinical evidence.

The Q&A Agent supports multi-turn clinician queries anchored to the generated report, patient-specific evidence, and retrieved clinical context. For questions that fall outside the available data or clinical evidence base, the agent is instructed to signal uncertainty rather than generate unsupported recommendations. In the current prototype, clinician feedback is not captured through a dedicated in-system feedback mechanism. Instead, feedback on report usefulness, interpretability, and workflow fit was collected through structured post-session interviews conducted separately.

For each scheduled consultation, SleepHelp Agent generated an analysis report from the participant's accumulated lifelog, sleep, and EMA data. The sleep psychiatrist reviewed the report as a preparation aid, and used it during and after the consultation for patient education, clinical interpretation, and intervention planning. The detailed evaluation procedure including post-session feedback collection is described in Section 4.2.

Figure 2 shows the key-indicator view of the SleepHelp Agent interface for rapid pre-consultation review. The view summarizes core sleep metrics, including total sleep time (TST), time in bed (TIB), sleep efficiency (SE), sleep-stage duration (Deep, REM, Light), wake after sleep onset (WASO), sleep onset latency (SOL), arousal count, nocturnal heart rate, and sleep score, together with their variability and trend during the period. This allows clinicians to quickly identify sleep patterns that may require further review before consultation. Additional interface views, including daily trends, personal pattern analysis, evidence-grounded insights, recommendation candidates, and the interactive Q&A panel, are provided in Appendix A.

3.3 Multimodal Lifelog Evidence Layer

Table 1. Multimodal lifelog features used by SleepHelp Agent.

Data Source	Resolution	Features used	Role in analysis
Galaxy Watch	10-min records; aggregated into five daytime windows*	Heart rate summaries (mean, max, min), step counts (sum per zone), weekday/weekend indicators	Captures daytime physiological state and physical activity patterns
Android Smartphone	10-min records; aggregated into five daytime windows*	Ambient light exposure (mean, sum), screen-on duration (sum)	Captures environmental and behavioral context related to sleep hygiene
EMA Self-Report	Twice daily: post-sleep (AM) and pre-sleep (PM)	Post-sleep: sleep quality, mental and physical fatigue, pain, mood, stress (all in 5-likert scale), dream type (no/good/normal/bad) Pre-sleep: naps (start time, duration), caffeine intake (last intake time), alcohol intake (last intake time, type, amount), exercise near bedtime (start time, intensity)	Captures subjective sleep experience and modifiable behavioral factors
Withings Sleep Analyzer (WSA)	Nightly sleep episode	Time in bed, sleep timing, total sleep time, sleep onset latency, wake after sleep onset, sleep onset latency, wakeup latency, sleep-stage duration, number of awakenings, nocturnal heart rate, sleep efficiency (%), sleep score (0-100)	Provides objective nighttime sleep profile

* Five daytime windows: morning, 09:00–11:59; early afternoon, 12:00–14:59; late afternoon, 15:00–17:59; evening, 18:00–20:59; night, 21:00–23:59.

To support the agent pipeline described previously, SleepHelp Agent converts multimodal lifelog streams into clinically interpretable sleep and behavioral evidence. Participants wore a Galaxy Watch and carried an Android smartphone during daytime activities, while a WSA installed in their living environment recorded nighttime sleep. EMA surveys were completed immediately before and after sleep to capture subjective sleep-related states.

Daytime wearable and smartphone data are first segmented into 10-minute records and then aggregated into five predefined time windows: morning (09:00–11:59), early afternoon (12:00–14:59), late afternoon (15:00–17:59), evening (18:00–20:59), and night (21:00–23:59). This temporal aggregation allows the system to summarize diurnal patterns that may be relevant to sleep health, such as changes in physical activity, daytime heart rate, ambient light exposure, and screen-on behavior across different periods of the day.

From the Galaxy Watch stream, SleepHelp Agent extracts window-level heart rate summaries, including mean, maximum, and minimum heart rate, as well as step-count aggregates within each time window. Smartphone-derived features include ambient light exposure, summarized as mean and cumulative lux values, and screen-on duration, summarized as total minutes per time window.

Nighttime sleep features are derived from the WSA. These include time in bed, sleep start and end time, total sleep time, sleep onset latency, wake-up latency, wake after sleep onset, sleep-stage duration, number of awakenings, nocturnal mean heart rate, sleep efficiency, and sleep score. These features provide the primary objective sleep profile used by the agent to identify sleep continuity, sleep timing, and sleep quality patterns.

EMA responses provide subjective context before and after sleep. Post-sleep surveys capture perceived sleep quality, dream type, fatigue, pain, mood, and stress level. Pre-sleep surveys capture behavioral and contextual factors that may influence sleep, including naps, caffeine intake, alcohol intake, and exercise close to bedtime. By combining objective sensor-derived features with subjective EMA responses, SleepHelp Agent can distinguish directly measured sleep patterns from patient-reported sleep experience and pre-sleep behaviors.

SleepHelp Agent uses timestamps to align daytime lifelog features, nighttime sleep events, and EMA responses at the patient-episode level. Missing or incomplete data are retained as explicit indicators rather than silently imputed, allowing the downstream agents to qualify their interpretations with insufficient evidence. Table 1 illustrates the multimodal data sources used by SleepHelp Agent.

4 Preliminary Evaluation

4.1 Evaluation Setup

This pilot study was conducted in collaboration with a sleep psychiatrist and clinical researchers from the Department of Psychiatry, Chungnam National University Hospital, South Korea. The study protocol was approved by the Institutional Review Board of Chungnam National University Hospital (IRB No. 2023-04-065) and Electronics and Telecommunications Research Institute (N01-202403-01-003), and all participants provided written informed consent. Eligible participants were adults aged 19 to 64 years with clinically significant insomnia symptoms, defined by an Insomnia Severity Index (ISI) score of 15 or higher. Participants were excluded if they had changes in medications expected to affect sleep within the past three months, shift-work schedules, uncontrolled comorbid sleep disorders, moderate or severe suicidal ideation, clinically unstable psychiatric or medical conditions, or pregnancy. To reduce confounding from other sleep disorders, screening also included assessments for conditions that may cause insomnia-like symptoms, including restless legs syndrome, sleep apnea, and circadian rhythm sleep-wake disorders.

A total of 25 individuals were screened, of whom 21 were enrolled. Four individuals were excluded during screening: one did not meet the diagnostic criteria for insomnia, and three were suspected of having other sleep disorders, including circadian rhythm sleep-wake disorder or restless legs syndrome. Among the 21 enrolled participants, one withdrew consent during the study, resulting in 20 participants included in the pilot evaluation.

Participants were followed for 12 weeks and completed five scheduled clinical contacts at weeks 0, 2, 4, 8, and 12. Each visit included consultation with a sleep psychiatrist and self-report clinical assessments. The first and final visits were conducted in person, while intermediate visits could be conducted remotely depending on participant convenience. Throughout the observation period, participants followed a previously published lifelogging protocol [19], which included daytime wearable and smartphone sensing, home-based sleep monitoring, and EMA surveys administered before and after sleep.

SleepHelp Agent was evaluated as a clinician-in-the-loop, pre-consultation decision support tool within this 12-week clinical workflow. A single sleep psychiatrist reviewed the agent-generated report and Q&A outputs prior to each consultation and provided structured post-session feedback through interviews. This deployment-oriented evaluation prioritized observing system behavior in a clinical workflow over benchmarking outputs against fixed ground truth.

4.2 Preliminary Clinical Feedback

We analyzed structured post-session interviews to assess the clinical utility, interpretability, and workflow fit of SleepHelp Agent as a report generation and Q&A-based decision support system for clinician. As the prototype was designed to support pre-consultation review, patient education, and intervention planning based on CBT-I principles, we focused on perceived usefulness and limitations in real-world consultation workflows. The feedback surfaced both the perceived clinical value of lifelog-based AI reports and four specific requirements for improving their reliability and actionability.

The clinician noted that the AI-generated report provided value by analyzing patients' sleep and behavioral patterns outside the consultation room. In particular, the report made it easier to review longitudinal sleep and activity patterns and identify lifestyle factors that may interfere with sleep hygiene, such as irregular sleep timing, excessive naps, low daytime activity, caffeine intake, stress-related sleep disruption, or inconsistent routines. This was perceived as useful for consultation preparation, patient education, and identifying candidate targets for behavioral intervention. The clinician also noted that lifelog-based summaries could complement interview-based assessment with objective daily life evidence and support more data-driven clinical decision making in insomnia care.

Theme 1: Need for terminological and structural consistency

The clinician emphasized the need for greater terminological and structural consistency across generated reports. Several clinically similar concepts were expressed using different terms across sections, such as circadian rhythm versus biological rhythm, and sleep fragmentation versus sleep disruption. This inconsistency reduced readability and made it more difficult to quickly interpret the report during consultation preparation.

Theme 2: Need for individualized recommendation generation

The clinician noted that treatment suggestions should be more individualized. Although the generated reports provided generally appropriate sleep hygiene and behavioral recommendations, these suggestions were often repetitive and did not always sufficiently reflect patient-specific characteristics, longitudinal lifelog patterns, or contextual factors such as daytime activity, naps, caffeine intake, stress level, and sleep timing variability. This feedback suggests that advanced versions should more explicitly connect recommendation generation to patient-specific evidence and clinically meaningful subgroups.

Theme 3: Need for quantitative grounding of interpretations

The clinician requested that the report provide more quantitative evidence for key claims. Some descriptions were perceived as subjective or overly narrative, making it difficult to determine the objective basis for the interpretation. For example, statements about poor sleep quality, irregular rhythm, or increased sleep fragmentation should be accompanied by concrete values such as sleep efficiency, wake after sleep onset, bedtime variability, total sleep time, EMA completion rate, or changes from the patient's own baseline.

Theme 4: Need for longitudinal association analysis

The clinician highlighted the difficulty of manually interpreting relationships among heterogeneous longitudinal variables. The collected lifelog data contained diverse sleep, activity, behavioral, and subjective variables, making it challenging to rapidly identify clinically meaningful associations during a short consultation workflow. In particular, integrating sleep metrics with daytime activity, naps, rest periods, caffeine intake, and stress levels over time required substantial clinical effort when presented only as summaries. The clinician therefore suggested that AI-generated analysis could be especially useful if it automatically identified correlations, temporal patterns, and candidate explanatory factors across variables.

In summary, the preliminary feedback suggests that clinicians perceived value in receiving a synthesized view of patients' sleep and behavioral patterns before consultation. At the same time, the feedback identified practical barriers to clinical use, including inconsistent terminology, insufficient personalization, limited quantitative support for some interpretations, and difficulty interpreting longitudinal relationships across multimodal lifelog variables.

5 Discussion

5.1 Design Implications for SleepHelp Agent

The preliminary clinical feedback suggests several concrete design requirements for the next iteration of SleepHelp Agent. This section translates them into *system-level* design directions for the agent pipeline, report schema, and clinical interaction structure.

First, the system should enforce controlled clinical terminology across the report and Q&A interactions. A shared terminology layer or domain-specific dictionary could help ensure that the Report Generation Agent and Clinical Q&A Agent use consistent representations for key sleep and behavioral concepts. Beyond surface-level normalization, mapping these concepts to medical ontologies such as SNOMED CT [20, 21] or UMLS [22, 23] could provide a semantic layer for agent reasoning, improving consistency across modules and supporting future interoperability.

Second, the system output should be more tightly anchored to patient-specific quantitative evidence, including both individual signals and the longitudinal patterns across signals. When relevant, the system should also indicate temporal patterns across these signals, such as recurring co-occurrence, lagged relationships, or changes over the multi-week observation period. This evidence anchoring can reduce hallucination risk as each interpretive claim becomes auditable against an observed feature or pattern.

Third, to capture multifactorial longitudinal relationships among sleep, daytime behavior, subjective symptoms, and contextual factors, future version could adopt a hierarchical agent structure organized by temporal scale. The treatment trajectory agent would track *long-term* treatment goals and patient trajectories across visits. The consultation planning agent would identify *session-level* behavioral targets and intervention priorities. The lifelog monitoring agent would analyze *daily-to-weekly* lifelog streams, missingness, and signal-level changes.

Finally, due to the lack of dedicated in-system feedback channel, clinician feedback in this study was collected out-of-band interviews, which means we could not measure whether session-level corrections would improve agent outputs over repeated sessions. Future evaluations should test whether structured in-system feedback improves terminology consistency, evidence grounding, report personalization, and clinician trust over time.

5.2 Implications for Clinician-in-the-Loop AI

By summarizing objective and subjective patterns across weeks of daily life, SleepHelp Agent's reports complement the retrospective interviews and episodic visits that anchor conventional insomnia assessment.

Beyond the data itself, the agent-based interaction style changes how clinicians engage with longitudinal lifelog evidence. Whereas dashboard-style lifelog systems present aggregated visualizations and leave interpretation to the clinician, we embed interpretation into the review workflow through a structured report and in-session Q&A. The pre-generated report provides a shared starting point for consultation review, while the follow-up Q&A enables targeted probing of specific time windows, signal combinations, or behavioral hypotheses that arise during review. These modes are complementary rather than redundant: the report provides scope, and Q&A provides depth. And the pilot suggests that clinicians actively use both, drawing on the report for an at-a-glance overview and on Q&A to resolve case-specific questions that a static report could not anticipate.

A further implication arises when SleepHelp Agent is viewed across cases rather than within a single session. In the current prototype, clinical knowledge is retrieved from a fixed RAG knowledge base and used to ground agent outputs; the knowledge base itself does not change. Future systems could also capture clinician-validated insights such as recurring behavioral patterns observed across patients or interpretation rules developed in local practice, and integrate them into institutional knowledge resources after appropriate review. This would create two learning loops: a *short loop* in which clinician corrections improve individual agent outputs, and a *longer loop* in which validated clinical insights update the knowledge base over time. In this way, the clinician becomes both a reviewer of individual reports and a curator of accumulated clinical knowledge.

The implications also extend beyond insomnia care. Sleep disturbance, circadian disruption, daytime inactivity, and stress-related behavioral patterns are closely connected to depression, anxiety, and broader emotional and cognitive functioning. Systems that organize longitudinal sleep and behavior data into clinically interpretable summaries may therefore contribute to broader mental health support by helping clinicians recognize behavioral patterns that are difficult to capture through episodic self-report alone.

5.3 Limitations and Future Work

While the pilot evaluation suggests that SleepHelp Agent can support clinicians by organizing longitudinal sleep and behavioral data for pre-consultation review, several limitations should be considered when interpreting the findings. The present study should be understood as an exploratory pilot evaluation: we observed how a clinician-in-the-loop agent system was used in a 12-week sleep psychiatry workflow rather than benchmarking outputs against fixed ground truth.

First, the pilot evaluation involved a limited number of participants from a single clinical workflow, and therefore the findings should be interpreted cautiously. Although the study provided initial evidence on the feasibility of using SleepHelp Agent as a report generation and Q&A-based decision-support system, the results may not generalize to broader populations, other clinical settings, or patients with different types of sleep complaints. Future studies should evaluate the system with larger and more diverse cohorts, multiple clinicians, and additional clinical sites.

Second, because the study was conducted in daily-life environments, external contextual factors could not be fully controlled. Participants' sleep and behavioral patterns may have been influenced by living conditions, weather, occupational schedules, travel, acute stress events, family routines, or other environmental factors that were not completely captured by the sensing and EMA protocol. These uncontrolled factors may affect both the interpretation of lifelog patterns and the clinical relevance of agent-generated insights. Future versions should incorporate richer contextual information and explicitly indicate when relevant environmental or situational factors are unavailable.

Finally, future deployment should address privacy, security, and auditability for lifelog-based clinical decision support system as SleepHelp Agent processes sensitive longitudinal data that can reveal intimate aspects of daily life. Future systems should therefore implement access control, encryption, audit logging, pseudonymization, and clear data retention policies.

6 Conclusion

We presented SleepHelp Agent, a lifelog-grounded multi-agent AI system designed to support clinicians in insomnia care through structured report generation and interactive Q&A. The system grounds LLM-generated content in retrieved clinical knowledge and uses a two-tier report schema that explicitly separates directly computed observed findings from suggested interpretations. This design helps clinicians trace each interpretation to patient data or retrieved evidence.

We evaluated SleepHelp Agent as a clinician-in-the-loop, pre-consultation decision support tool in a 12-week sleep psychiatry workflow. Preliminary clinical feedback suggests that AI reports generated from lifelog data have practical value for organizing complex longitudinal sleep and behavioral data, identifying sleep-disrupting behaviors, and supporting more data-informed insomnia care. The feedback also points to key directions for future work: standardized clinical terminology, stronger links between interpretations and patient-specific evidence, longitudinal pattern analysis, and structured in-loop feedback channels for clinician correction.

SleepHelp Agent contributes to AI-assisted clinical workflows in mental health by showing how everyday lifelog can be translated into reviewable reports and targeted Q&A for clinicians. Rather than emphasizing automated diagnosis, the system supports clinical interpretation, patient education, and CBT-I-informed behavioral intervention planning. By making longitudinal sleep, behavioral, and subjective patterns more visible during consultation preparation, SleepHelp Agent offers a step toward precision sleep medicine and data-informed mental health support.

ACKNOWLEDGMENTS

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [26ZR1100, Research on Intelligent Industrial Convergence]

REFERENCES

[1] Edition, F., 2013. Diagnostic and statistical manual of mental disorders. *Am Psychiatric Assoc*, *21*(21), pp.591-643.

[2] Baglioni, C., Battagliese, G., Feige, B., Spiegelhalder, K., Nissen, C., Voderholzer, U., Lombardo, C. and Riemann, D., 2011. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. *Journal of affective disorders*, *135*(1-3), pp.10-19.

[3] Hertenstein, E., Feige, B., Gmeiner, T., Kienzler, C., Spiegelhalder, K., Johann, A., Jansson-Fröjmark, M., Palagini, L., Rücker, G., Riemann, D. and Baglioni, C., 2019. Insomnia as a predictor of mental disorders: a systematic review and meta-analysis. *Sleep medicine reviews*, *43*, pp.96-105.

[4] Edinger, J.D., Arnedt, J.T., Bertisch, S.M., Carney, C.E., Harrington, J.J., Lichstein, K.L., Sateia, M.J., Troxel, W.M., Zhou, E.S., Kazmi, U. and Heald, J.L., 2021. Behavioral and psychological treatments for chronic insomnia disorder in adults: an American Academy of Sleep Medicine clinical practice guideline. *Journal of clinical sleep medicine*, *17*(2), pp.255-262.

[5] Omar, M., Sorin, V., Collins, J.D., Reich, D., Freeman, R., Gavin, N., Charney, A., Stump, L., Bragazzi, N.L., Nadkarni, G.N. and Klang, E., 2025. Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support. *Communications Medicine*, *5*(1), p.330.

[6] Lee, S., Oh, J.W., Park, K.M., Lee, S. and Lee, E., 2023. Digital cognitive behavioral therapy for insomnia on depression and anxiety: a systematic review and meta-analysis. *NPJ digital medicine*, *6*(1), p.52.

[7] Hwang, J.W., Lee, G.E., Woo, J.H., Kim, S.M. and Kwon, J.Y., 2025. Systematic review and meta-analysis on fully automated digital cognitive behavioral therapy for insomnia. *npj Digital Medicine*, *8*(1), p.157.

[8] Birrer, V., Elgendi, M., Lambercy, O. and Menon, C., 2024. Evaluating reliability in wearable devices for sleep staging. *NPJ Digital Medicine*, *7*(1), p.74.

[9] Aziz, S., AM Ali, A., Aslam, H., A Abd-alrazaq, A., AlSaad, R., Alajlani, M., Ahmad, R., Khalil, L., Ahmed, A. and Sheikh, J., 2025. Wearable artificial intelligence for sleep disorders: scoping review. *Journal of Medical Internet Research*, *27*, p.e65272.

[10] Massar, S.A., Chua, X.Y., Soon, C.S., Ng, A.S., Ong, J.L., Chee, N.I., Lee, T.S., Ghosh, A. and Chee, M.W., 2021. Trait-like nocturnal sleep behavior identified by combining wearable, phone-use, and self-report data. *NPJ digital medicine*, *4*(1), p.90.

[11] Edouard, P., Campo, D., Bartet, P., Yang, R.Y., Bruyneel, M., Roisman, G. and Escourrou, P., 2021. Validation of the Withings Sleep Analyzer, an under-the-mattress device for the detection of moderate-severe sleep apnea syndrome. *Journal of Clinical Sleep Medicine*, *17*(6), pp.1217-1227.

[12] Stefanos, M.A., de Laboulaye, G., Campo, D., de Gourcuff, M., Escourrou, P., Matrot, B., Islind, A.S. and Geoffroy, P.A., 2026. Evaluation of a Contactless Sleep Monitoring Device for Sleep Stage Detection at Home in a Healthy Population: Prospective Study in Free-Living Conditions. *JMIR Human Factors*, *13*(1), p.e77033.

[13] Wang, W., Ma, Z., Wang, Z., Wu, C., Ji, J., Chen, W., Li, X. and Yuan, Y., 2025. A survey of llm-based agents in medicine: How far are we from baymax?. *Findings of the Association for Computational Linguistics: ACL 2025*, pp.10345-10359.

[14] Silveira, A.L., da Rosa Righi, R. and Da Costa, C.A., 2025. Multi-agent systems for clinical decision support: A systematic review. *Applied Soft Computing*, p.114447.

[15] Stade, E.C., Stirman, S.W., Ungar, L.H., Boland, C.L., Schwartz, H.A., Yaden, D.B., Sedoc, J., DeRubeis, R.J., Willer, R. and Eichstaedt, J.C., 2024. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *NPJ Mental Health Research*, *3*(1), p.12.
[16] Lawrence, H.R., Schneider, R.A., Rubin, S.B., Matarić, M.J., McDuff, D.J. and Bell, M.J., 2024. The opportunities and risks of large language models in mental health. *JMIR Mental Health*, *11*(1), p.e59479.
[17] Amugongo, L.M., Mascheroni, P., Brooks, S., Doering, S. and Seidel, J., 2025. Retrieval augmented generation for large language models in healthcare: A systematic review. *PLOS Digital Health*, *4*(6), p.e0000877.
[18] Xiong, G., Jin, Q., Lu, Z. and Zhang, A., 2024, August. Benchmarking retrieval-augmented generation for medicine. In *Findings of the Association for Computational Linguistics: ACL 2024* (pp. 6233-6251).
[19] Oh, S.W., Jeong, H., Chung, S., Lim, J.M., Noh, K.J., Lee, S. and Jung, G., 2025. Understanding human daily experience through continuous sensing: Etri lifelog dataset 2024. *arXiv preprint arXiv:2508.03698*.
[20] https://docs.snomed.org/
[21] https://www.nlm.nih.gov/healthit/snomedct/international.html
[22] Bodenreider, O., 2004. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research*, *32*(suppl_1), pp.D267-D270.
[23] Jing, X., 2021. The unified medical language system at 30 years and how it is used and published: systematic review and content analysis. *JMIR medical informatics*, *9*(8), p.e20675.

APPENDIX

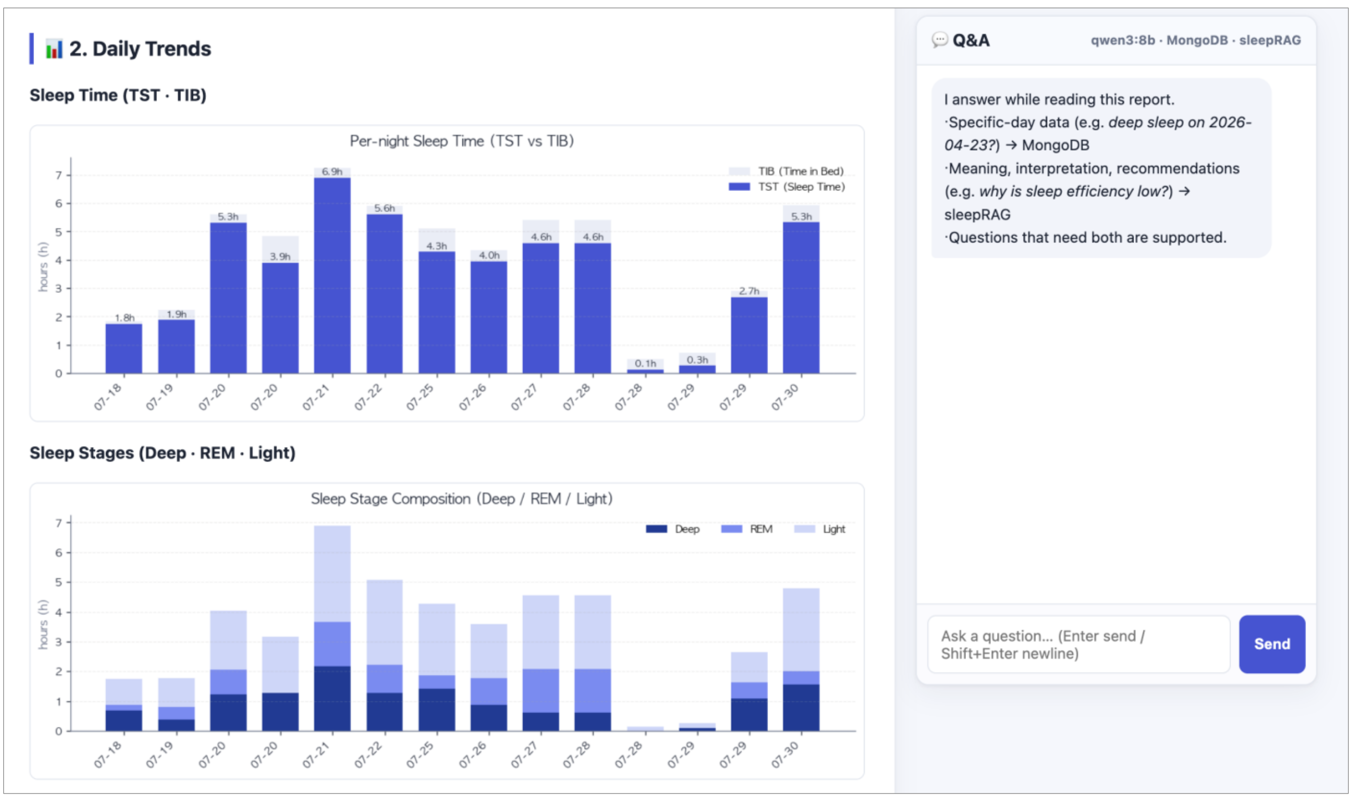

Figure A1. Daily sleep trend view.

Figure A2. Sleep efficiency and nocturnal heart-rate visualization.

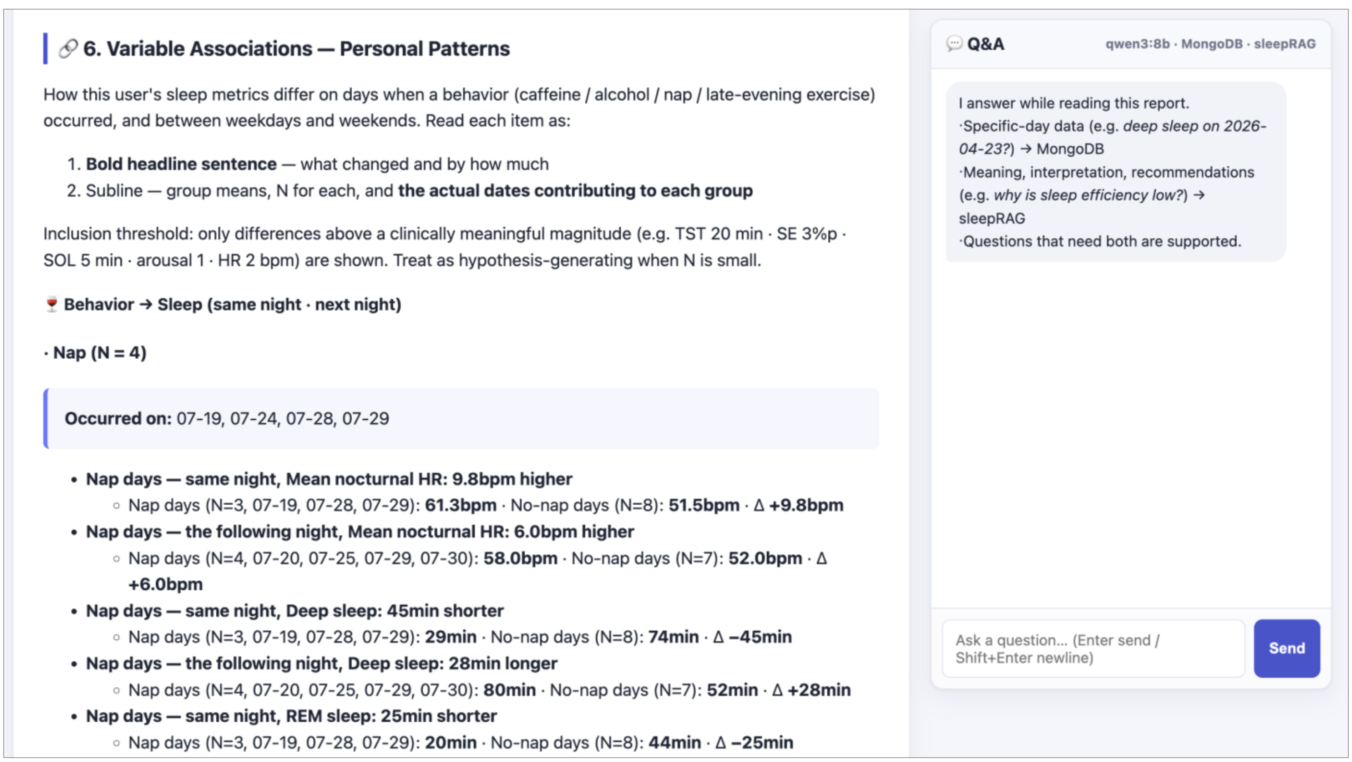

Figure A3. Variable association and personal pattern view.

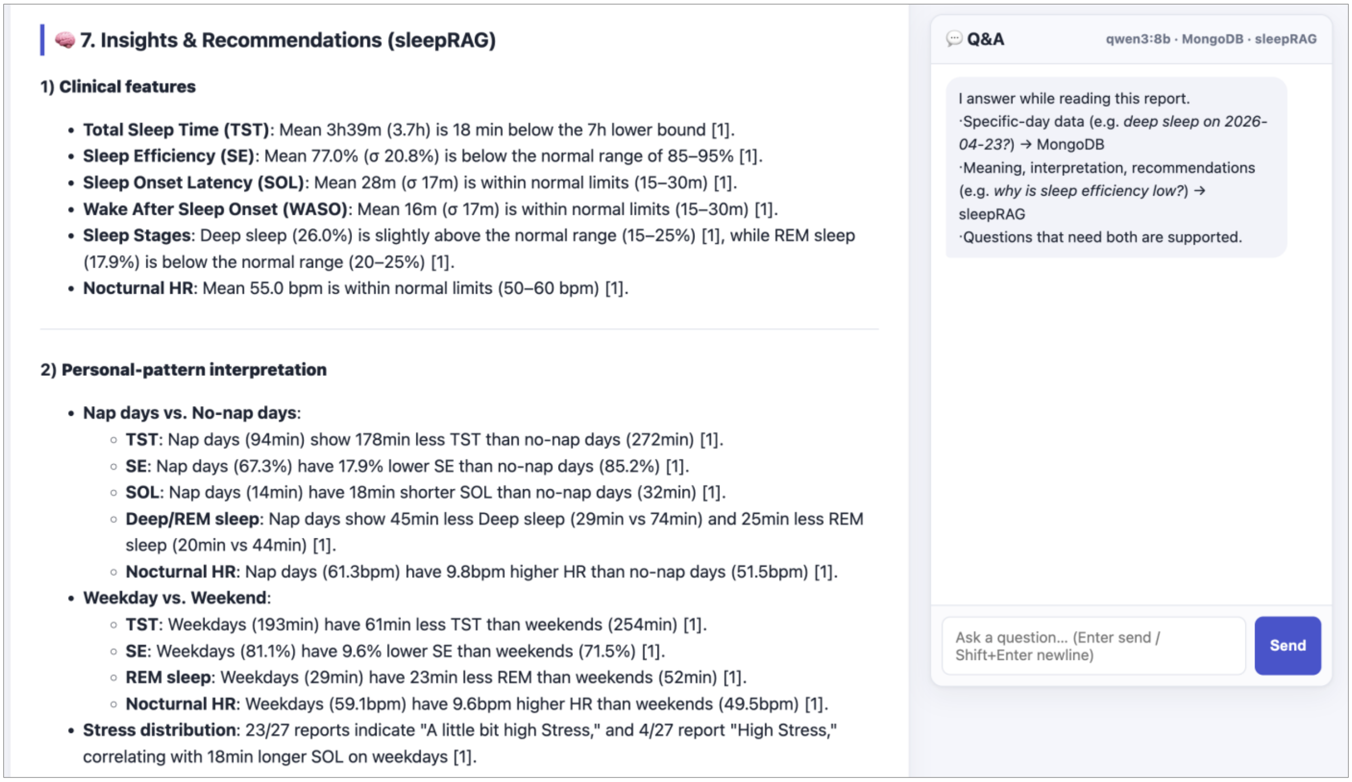

Figure A4. Evidence-grounded insights and recommendation candidates.

A. Additional SleepHelp Agent Interface Views

Figure 2 in the main text presents the key-indicator view used for rapid pre-consultation review. Additional interface views are provided here to illustrate how SleepHelp Agent supports more detailed inspection after clinicians identify patterns that require further review. These views include longitudinal sleep trends, time-series visualizations of sleep efficiency and nocturnal heart rate, personal pattern analyses across behavioral conditions, and evidence-grounded insights with recommendation candidates and interactive Q&A. Together, these supplementary screens show how the system moves from high-level sleep summaries to more detailed review of temporal patterns, behavioral associations, clinical interpretations, and follow-up questions.

Figure A1 shows the daily sleep trend view, which presents longitudinal changes in key sleep metrics across the monitoring period and allows clinicians to inspect day-to-day variation in sleep timing, sleep duration, wake after sleep onset, and related sleep indicators. Figure A2 presents time-series visualizations of sleep efficiency and nocturnal heart rate, supporting review of sleep continuity and physiological patterns over time. Figure A3 shows the variable association and personal pattern view, summarizing patient-specific associations across sleep, daytime behavior, and contextual variables, such as differences between nap and no-nap days or weekday and weekend patterns. Figure A4 illustrates the evidence-grounded insights and recommendation candidates. This presents clinical insights and recommendation candidates generated from patient-specific evidence and retrieved clinical knowledge. The adjacent Q&A panel allows clinicians to ask follow-up questions about specific findings, interpretations, or supporting evidence.