

Design and Evaluation of a Memory-Enhanced Chatbot Architecture for Supporting ME/CFS Patients

Written Examination Paper for the Master's Examination
in Applied Computer Science
at Ruhr-Universität Bochum

submitted by

Hannes Matthias Ehringfeld

1st Reviewer: Prof. Dr. Christian Meske

2nd Reviewer: Tobias Hermanns, M. Sc.

Author: Hannes Matthias Ehringfeld

108 021 25993 2

[hannes.ehringfeld\[at\]ruhr-uni-bochum.de](mailto:hannes.ehringfeld[at]ruhr-uni-bochum.de)

Submission Date: July 28, 2025

Abstract

Managing Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) requires meticulous energy management, known as pacing, to avoid its hallmark symptom, Post-Exertional Malaise. While applying pacing, the recommended, detailed logging of symptoms and activities imposes a significant documentation burden on patients already suffering from profound fatigue and cognitive impairment. To address this, this thesis presents the design, development, and evaluation of LogChat, a novel, memory-enhanced chatbot architecture. LogChat is designed with a dual purpose: to transform natural, conversational interactions into a structured health diary and to provide accessible, on-demand educational support. Following the Design Science Research Methodology, this work answers how current natural language processing technology can be integrated into a system to enable such tracking. The LogChat prototype was rigorously assessed using a custom, automated evaluation framework employing impersonator and judge Large Language Model (LLM) agents. The results confirm the architecture's viability, successfully serving as a proof-of-concept. However, the evaluation revealed a critical performance dichotomy: while proprietary LLMs achieved near-perfect reliability, currently available open-source models suitable for on-device deployment demonstrated significant inconsistencies, rendering them unsuitable for this sensitive application. This thesis contributes a validated modular architecture for conversational health logging and education, a replicable evaluation method, and crucial insights into the performance gap that defines the current practical limits of such systems.

Contents

List of Figures	ix
List of Tables	x
1. Introduction	1
1.1. The Challenge of ME/CFS and Pacing	1
1.2. Motivation for a Technological Solution	6
1.3. LogChat: A Conversational Approach to Support Pacing	8
1.4. Thesis Overview and Contributions	9
2. Background and Related Work	10
2.1. The Clinical Demands of Pacing in ME/CFS	10
2.2. Technological Foundations of Conversational AI	12
2.2.1. Large Language Models	12
2.2.2. Prompt Engineering	14
2.2.3. Agentic Architectures and Tool Use	15
2.2.4. Retrieval-Augmented Generation and Vector Search	16
2.3. Related Work	17
2.3.1. Existing Digital Health Tools for ME/CFS	17
2.3.2. Conversational Agents in Healthcare	18
2.3.3. Evaluation of Conversational Systems	20
2.4. Identifying the Research Gap	21
3. Methodology	22
3.1. Introduction to Design Science Research	22
3.2. Application of the Design Science Research Methodology	23

4. Objectives for a Solution	23
4.1. Core Use Cases and Functional Objectives	24
4.1.1. Use Case 1: Conversation	24
4.1.2. Use Case 2: Logging	25
4.1.3. Use Case 3: Question Answering	25
4.2. Non-Functional Objectives	26
5. LogChat System Architecture and Development	26
5.1. Initial Design Iteration and Objective Refinement	27
5.2. The LogChat System Architecture	28
5.3. LLM Layer	29
5.4. Data Layer	29
5.4.1. User Data Schema	29
5.4.2. Knowledge Base	31
5.5. Interaction Layer	32
5.5.1. User Interface	32
5.5.2. Opener Node	33
5.5.3. Planner Node	34
5.5.4. Responder Node	36
5.6. Post-Interaction Layer	36
5.6.1. Summarizer Node	37
5.6.2. Extractor Node	38
6. Demonstration and Evaluation Setup	39
6.1. Overview of the Evaluation Framework	39
6.2. Evaluation Dataset	41
6.3. The Impersonator	42
6.4. Interaction Logs	43
6.5. The Judge	44
6.6. Performance Scoring	45
6.7. Visualizing Persona Trajectories	46

7. Results	47
7.1. Quantitative Results	48
7.2. Qualitative Analysis	49
7.2.1. Conversation	51
7.2.2. Logging	52
7.2.3. Question Answering	55
8. Discussion	56
8.1. Answering the Research Question	56
8.2. Reflection on Non-Functional Objectives	58
8.3. Limitations and Avenues for Future Work	59
8.3.1. Evaluation Limitations	60
8.3.2. Architectural and Technological Challenges	61
8.3.3. Identified System and Feature Limitations	62
9. Conclusion and Future Work	62
9.1. Future Work	64
Bibliography	65
A. Raw Prompts	75
A.1. Opener Node Prompts	75
A.1.1. First Opener Prompt	75
A.1.2. Opener System Prompt	75
A.1.3. Opener Instructions Prompt	75
A.2. Planner Node Prompts	76
A.2.1. Initial User Description Prompt	76
A.2.2. Initial Thread Summaries Prompt	76
A.2.3. Planner System Prompt	77
A.2.4. Planner Instructions Prompt	77
A.2.5. Planner Instructions With Tool Results Prompt	79

A.3. Responder Node Prompts	81
A.3.1. Responder System Prompt	81
A.3.2. Responder Instructions Prompt	81
A.4. Extractor Node Prompts	83
A.4.1. Extractor System Prompt	83
A.4.2. Extractor Extract Symptoms Instructions Prompt	83
A.4.3. Extractor Extract Activities Instructions Prompt	84
A.5. Summarizer Node Prompts	87
A.5.1. Summarizer System Prompt	87
A.5.2. Summarizer Summarize Interaction Instructions Prompt	87
A.5.3. Summarizer Update Long Term Memory Prompt	88
A.6. Evaluation Prompts	89
A.6.1. Impersonator System Prompt	89
A.6.2. Judge System Prompt	91
A.6.3. Judge Instructions Prompt	91
B. Full Prompt Examples with Placeholder Values Inserted	95
B.1. Opener Node Prompt Examples	96
B.1.1. Opener Prompt Example	96
B.2. Planner Prompt Examples	98
B.2.1. Planner Prompt Example	98
B.2.2. Planner Prompt Example with tool results	102
B.3. Responder Prompt Examples	106
B.3.1. Responder Prompt Example	106
B.3.2. Responder Prompt Example with tool results	108
B.4. Summarizer Prompt Examples	112
B.4.1. Summarizer: Summarize Interaction Prompt Example	112
B.4.2. Summarizer: Update Long-Term Memory Prompt Example	114
B.5. Extractor Prompt Examples	118
B.5.1. Extractor: Extract Activities	118
B.5.2. Extractor: Extract Symptoms	123

B.6. Impersonator Prompt Example	127
B.6.1. Impersonator System Prompt Example	127
B.7. Judge Prompt Example	129
B.7.1. Judge Instruction Prompt Example	129
C. Code Snippets	136
C.1. Planner Node Tools	137
C.1.1. Retrieve Activity Level Tool	137
C.1.2. Retrieve Information Tool	139
C.2. Extractor Node Tools	141
C.2.1. Log Symptom Tool	141
C.2.2. Log Activity Tool	142
C.3. Impersonator Tools	144
C.3.1. End Conversation Tool	144
C.3.2. Thread Summary to String	144
C.4. Evaluation Dataset	146
C.4.1. Evaluation Dataset Structure	146
C.4.2. Full Evaluation Dataset	146
C.5. Log Examples	167
C.5.1. Debug Log Example	167
C.5.2. Verdict Log Example	171
D. Full Evaluation Logs	171
D.1. Persona: Mark Thompson	172
D.1.1. Interaction 1: 2025-04-21	172
D.1.2. Interaction 2: 2025-04-23	178
D.1.3. Interaction 3: 2025-04-25	184
D.1.4. Interaction 4: 2025-04-28	190
D.1.5. Interaction 5: 2025-05-01	196
D.2. Persona: Sarah Chen	203
D.2.1. Interaction 1: 2025-04-22	203

D.2.2.	Interaction 2: 2025-04-24	208
D.2.3.	Interaction 3: 2025-04-27	216
D.2.4.	Interaction 4: 2025-04-30	220
D.2.5.	Interaction 5: 2025-05-03	225
D.3.	Persona: Elena Garcia	230
D.3.1.	Interaction 1: 2025-04-21	230
D.3.2.	Interaction 2: 2025-04-22	238
D.3.3.	Interaction 3: 2025-04-24	245
D.3.4.	Interaction 4: 2025-04-26	250
D.3.5.	Interaction 5: 2025-04-29	256

List of Figures

5.1. Conceptual Layers of the LogChat Architecture	28
5.2. User Data Schema	30
5.3. Knowledge Schema	31
5.4. LogChat Command Line Interface	33
6.1. LogChat Evaluation Setup	40
7.1. Persona Overview Plots	54

List of Tables

7.1. Quantitative Performance Comparison of LLMs driving LogChat	49
7.2. Summary of Qualitative Objective Fulfillment for Top-Performing Models	51

1. Introduction

This thesis confronts the profound difficulties encountered by individuals affected by Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS), a severe and often debilitating chronic condition. This introductory chapter establishes the foundation for the research presented, beginning with an in-depth exploration of ME/CFS itself and the significant challenges patients face in managing their daily lives, particularly through the energy management strategy known as pacing. Subsequently, the chapter will articulate the motivation for seeking a technological solution to alleviate these burdens, introduce the LogChat system developed as part of this work, and outline the overarching objectives, contributions, and structure of this thesis.

1.1. The Challenge of ME/CFS and Pacing

ME/CFS is a serious, debilitating, and chronic illness of largely unknown etiology (Institute of Medicine et al., 2015; Weigel et al., 2025). The onset of its symptoms is frequently reported to follow a viral infection, such as Epstein-Barr virus, SARS-CoV-2, influenza, or Ebola (Renz-Polster et al., 2022). In other cases, patients have reported stressful life events or exposure to environmental toxins as antecedents to symptom development (Chu et al., 2019). The disease is characterized by a persistent and profound fatigue, which patients have described as an “unimaginable fatigue and extreme exhaustion” or an “exhaustion to the point that speaking is not possible.” (FDA, 2013, p. 14). Beyond fatigue, ME/CFS is associated with a constellation of other symptoms, including unrefreshing sleep, exercise intolerance, orthostatic intolerance, gastrointestinal disturbances, flu-like symptoms, and chronic pain manifesting in joints, muscles, and as headaches (FDA, 2013; Institute of Medicine et al., 2015; Stussman et al., 2020). Furthermore, some individuals report heightened sensitivities to light, sound, smell, touch, food, and medications (FDA, 2013; Institute of Medicine et al., 2015).

Such sensitivities can exacerbate what is among the most frequently reported and disruptive symptoms: a range of cognitive impairments collectively referred to as “brain fog”. This brain fog can

manifest as an inability to read, write, or comprehend spoken language. In severe instances, it may cause individuals to temporarily lose their ability to recall names, communicate effectively, or make basic decisions (FDA, 2013; Institute of Medicine et al., 2015). The cognitive dysfunction can be so persistent that some patients report being able to concentrate for only 15 to 20 minutes at a time (FDA, 2013). Brain fog can be triggered or worsened by these heightened sensitivities, as one patient described: “I cannot focus if faced with too much visual or sound stimulation, for example, too many objects in a store, music in a restaurant, or busy websites. I get dizzy, spatially disoriented, lose my balance, come down with headaches, and feel nauseated in such environments.” (FDA, 2013, p. 7). Cumulatively, these symptoms severely limit patients’ functional ability, impairing their capacity to complete basic tasks of everyday living. Consequently, many individuals with ME/CFS are unable to work, and a significant portion become housebound or even bedbound (Institute of Medicine et al., 2015; Weigel et al., 2025).

Research into ME/CFS has identified several physiological abnormalities in patients that may explain their experience. For instance, severe metabolic dysregulations indicate impaired energy metabolism and systemic metabolic stress (McGregor et al., 2019). Exercise challenge studies have revealed further disparities, showing that the metabolic response and recovery in ME/CFS patients is distinctly disrupted, particularly in lipid and energy-related pathways (Germain et al., 2022). Another study found impaired ATP synthesis and compensatory mechanisms, suggesting mitochondrial failure (Missailidis et al., 2020). Furthermore, stress-induced aggravation of central nervous system inflammation (Renz-Polster et al., 2022) and skeletal muscle disturbances (Scheibenbogen & Wirth, 2024) have been suggested as physiological mechanisms explaining symptoms. Experimental exercise studies have revealed an altered autonomic nervous system response post-exertion in ME/CFS patients; repeated exercise tests triggered prolonged muscle weakness and delayed recovery of muscular capacity, correlating with patients’ self-reported experiences of Post-Exertional Malaise (Vøllestad & Mengshoel, 2023).

Post-Exertional Malaise (PEM), also referred to as a “crash”, is the hallmark of ME/CFS. It denotes an overall worsening of all symptoms following physical or mental exertion. The exertion triggering PEM can be minor compared to that tolerated by healthy individuals; for example, trips to the grocery store or a family member visiting can trigger PEM (Stussman et al., 2020). Some-

times PEM may last only several hours or days, which is often denoted as a “flare up”, but it can cause a relapse, which can last much longer and require substantial and sustained adjustments to the patient’s energy management (NICE, 2021). The functional reduction during PEM periods is significant, as illustrated by one patient: “You cannot focus on simple things like remembering the name of a lamp... the word won’t come. I can’t balance a checkbook, can’t do any kind of math, can’t absorb information. People will be explaining something to you, and it’s like they’re speaking another language, and my mind will not focus on what they’re saying. Those are times I stay home because I shouldn’t be driving. I shouldn’t be operating any kind of machinery. I shouldn’t be cooking because I’m not able to function on a “normal” basis like everybody else does.” (Stussman et al., 2020, p. 13).

The extremely slow recovery from any exertion in ME/CFS patients entails the risk of repeatedly triggering PEM, even before the previous PEM period has fully resolved. Each time this occurs, the recovery period tends to lengthen, further increasing the risk of subsequent PEM episodes before full recovery (Moore et al., 2023). This cyclical effect can lead to a downward spiral of diminishing functional capacity and escalating symptom severity. For ME/CFS patients, overexertion poses the risk of permanently reducing quality of life by inducing irreversible disabilities (Moore et al., 2023; Thoma et al., 2024). “It has been long recognized by clinical providers who treat ME/CFS that continued and repeated episodes of PEM within any one individual appear to be associated with a worsened long-term functional prognosis.” (Bateman Horne Center, 2023, p. 2). Consequently, patients are advised to actively pace themselves and avoid pushing themselves into PEM (Bateman Horne Center, 2023; Deutsche Gesellschaft für ME/CFS e.V., 2022; NICE, 2021; Seltzer, 2021; Shepherd & Mayes, 2023; Solve ME/CFS Initiative, 2025a).

The recommendation to pace oneself has led to *Pacing* becoming the term for the central energy management framework advised for patients. This approach is based on the Energy Envelope Theory, first formulated by Jason et al. (2013). The theory proposes that matching expended energy to perceived available energy reduces symptom severity and improves functioning in ME/CFS. The paper includes case studies demonstrating that balancing activity within individualized energy limits minimizes PEM and enhances quality of life. A more recent study also finds that patients view pacing as a strategy to achieve more during the day (Barakou et al., 2025). This patient

perception is powerfully underscored by large-scale survey data, which identifies pacing as the single most effective management strategy reported by thousands of individuals with ME/CFS and Long COVID (Eckey et al., 2025).

Studying pacing and providing specific instructions on its effective implementation is challenging, as the framework's application is highly individualized and lacks a unanimous definition within the literature and among healthcare providers (Sanal-Hayes et al., 2023). It depends on factors such as the stage of the illness, the severity of symptoms, and the extent to which the condition fluctuates (Shepherd & Mayes, 2023). The advice is not merely to remain in bed as much as possible, as this can also lead to negative consequences. Instead, patients are encouraged to be as active as reasonably possible while consistently preventing overexertion. Jason et al. (2013) outlines core principles such as self-monitoring daily energy, moderating activity to avoid overexertion, and balancing mental and physical tasks, while emphasizing rehabilitation through gradual expansion of the energy envelope. However, advice provided in leaflets from organizations such as the ME Association in the UK is often more specific and readily digestible for patients (Shepherd & Mayes, 2023). This guidance first advises patients to find a comfortable baseline of activity, which is often significantly lower than expected and may require a period of trial and error to establish (Shepherd & Mayes, 2023). To manage their limited energy—conceptualized as a malfunctioning and slowly charging battery—patients are encouraged to break down all tasks into small, manageable chunks interspersed with adequate rest (Shepherd & Mayes, 2023). For instance, a complex activity like preparing an evening meal might be deconstructed into separate sub-tasks staggered throughout the day: planning the meal in the morning, preparing ingredients in the afternoon, and leaving the washing-up for a later time or for someone else to handle (Shepherd & Mayes, 2023). The strategy also involves alternating not only between physical and mental exertion but also between physical positions, such as from sitting to lying down, to further conserve energy (Shepherd & Mayes, 2023). Crucially, patients must learn to recognize the warning signs of impending overexertion—not merely tiredness, but a distinct sensation of losing control or concentration—and to use their willpower not to *push through* symptoms, but to stop an activity before exacerbation occurs (Shepherd & Mayes, 2023).

Applying these pacing principles consistently presents considerable difficulties. One major chal-

lenge is the inherent tendency to push through on days when symptoms may feel less severe, leading to a “boom and bust” cycle of overactivity followed by a debilitating crash (The ME Association, 2020). This difficulty is compounded by the complex timing of PEM. While the initial onset can be rapid, occurring within minutes or hours of a trigger (Hartle et al., 2021), the full exacerbation of symptoms often has a characteristic delay, developing over the subsequent hours or even one to two days (Institute of Medicine et al., 2015; Stussman et al., 2020). This temporal disconnect makes it exceptionally challenging for individuals to correlate specific activities with subsequent symptom exacerbation in real-time, obscuring the immediate consequences of overdoing.

Consequently, effective pacing is often supported not only by a commitment to engaging with educational materials but also by the diligent and precise logging of daily activities, energy levels, and symptoms. For patients with some degree of variable capacity—those who may have better and worse days and are able to engage in some level of activity—such logs can serve as an invaluable tool for identifying personal triggers and understanding their individual energy envelope. This allows for informed adjustments to their activity management over time. However, the very nature of ME/CFS, with its profound fatigue and cognitive impairments (“brain fog”), makes the consistent upkeep of detailed diaries a significant documentation burden, which may outweigh the benefits for those with the most severe, functionally-limiting forms of the illness. This burden can, paradoxically, consume the very energy that patients are so carefully trying to conserve. It is this critical challenge—the high documentation burden associated with the recommended strategy of detailed logging in the face of debilitating illness—that forms the central problem addressed by this thesis.

The research presented herein explores how current natural language processing technology can be leveraged to alleviate this burden. Specifically, this thesis introduces LogChat, a prototype of a memory-enhanced, personalized chatbot architecture designed to simplify the daily logging of symptoms and activities through natural, conversational interaction. The aim is to demonstrate that such a system can not only reduce the documentation toll but also support patients in recognizing patterns, understanding their condition more deeply, and ultimately adhering more effectively to pacing strategies to avoid PEM and improve their quality of life.

1.2. Motivation for a Technological Solution

The significant documentation burden associated with the detailed logging recommended for effective pacing represents a critical barrier to self-management for individuals with ME/CFS. The motivation for developing a technological solution to lower this barrier is driven by a juncture of several interconnected reasons.

Firstly, there is currently no cure for ME/CFS, and recovery rates are extremely low, meaning many patients may face a lifelong struggle with the illness (Seton et al., 2024; Weigel et al., 2025). Fewer than 5% of patients return to their pre-morbid activity levels, and for most, the prognosis is a life-long condition, which underscores the critical need for effective strategies to manage the condition and improve quality of life (Seton et al., 2024; Weigel et al., 2025).

Secondly, the global COVID-19 pandemic has led to a concerning increase in ME/CFS case prevalence (James Daniell et al., 2025), with one large-scale study finding that individuals develop ME/CFS at nearly five times the rate of uninfected controls following a SARS-CoV-2 infection (Vernon et al., 2025). A systematic review and meta-analysis further underscores this link, concluding that approximately half of all patients with Long COVID meet the diagnostic criteria for ME/CFS (Dehlia & Guthridge, 2024). This rise in cases translates to escalating human suffering and a substantial economic burden on society due to healthcare costs and lost productivity (James Daniell et al., 2025; Valdez et al., 2019). For instance, in Germany alone, ME/CFS and Long COVID were estimated to cost the economy 63€ billion in 2024, equivalent to nearly 1.5% of its annual Gross Domestic Product (James Daniell et al., 2025). The number of active ME/CFS cases in Germany was projected to surpass 650,000 by the end of 2024 and is expected to continue rising (James Daniell et al., 2025). In England, it has been estimated that the lifetime prevalence could be as high as 0.92% for females and 0.25% for males, potentially affecting around 404,000 individuals in the UK overall (Samms & Ponting, 2025). The prevalence of ME/CFS worldwide has been estimated with considerable variance, but studies suggest it is a relatively common condition, not a rare disease, with estimates ranging up to affecting 0.89% of the population which would mean about 70 Million people (Lim et al., 2020; Valdez et al., 2019).

Thirdly, ME/CFS research has historically been underfunded and neglected (Institute of Medicine et al., 2015; James Daniell et al., 2025; Valdez et al., 2019), leading to a significant lack of aware-

ness and understanding among healthcare professionals and the general populace (König et al., 2024; Strassheim et al., 2021; Thoma et al., 2024). Many medical professionals are not well informed about the etiology, diagnosis, or treatment of ME/CFS (König et al., 2024; Thoma et al., 2024), and misconceptions about it being a psychosomatic illness persist despite evidence of organic abnormalities (Thoma et al., 2024). This lack of knowledge means individuals developing symptoms face a high risk of not receiving appropriate information or a timely diagnosis, potentially for years after onset (Thoma et al., 2024). Patients report that people in their social circles often have limited understanding of their disease (König et al., 2024). This situation is compounded by the fact that ME/CFS can affect anyone, though it disproportionately impacts women and may also affect lower socio-economic classes more significantly (Chu et al., 2019; Samms & Ponting, 2025).

Given these factors, this thesis operates on the assumption that educating patients can lead to better coping mechanisms and more effective pacing. If this holds true, such interventions could prevent symptom exacerbation, preserve valuable quality of life, and ultimately reduce societal costs, benefiting not only patients and their caretakers but society as a whole. This thesis aims to provide that help in the form of a chatbot prototype which can educate and become a supportive, conversational companion for symptom and activity documentation.

Furthermore, the current technological landscape presents a uniquely opportune moment for conducting this research. The period since the introduction of ChatGPT in late 2022 (OpenAI, 2022) has been marked by a sustained wave of increasingly capable open-source Large Language Models (LLMs) becoming publicly available, such as the Gemma series by Google (Kamath et al., 2025), the Llama series by Meta (Grattafiori et al., 2024), the Qwen series by Alibaba (Qwen et al., 2025), or fine-tuned versions like Hermes (Teknium et al., 2024) and Deep Cogito (Deep Cogito, 2025). There is also a significant push from major technology companies, including Apple with *Apple Intelligence* (Apple Inc., 2025) and Google with *Gemma 3n* (Gonzalez & Shivanna, 2025), to develop LLMs that are efficient enough to run locally on mobile devices. This trend towards on-device artificial intelligence (AI) aligns perfectly with the long-term vision for LogChat as a private, offline system. Therefore, this thesis is also underpinned by the assumption that in the near future, mobile LLM technology will be sufficiently advanced to realize a completely

private, personalized, memory-enhanced system for ME/CFS as an open-source project.

The potential benefits of such a memory-enhanced personal chatbot are not confined to ME/CFS. Other chronic conditions where symptom management involves tracking personal triggers and understanding complex patterns could also benefit. For example, individuals with neurodermatitis often contend with personal triggers, particularly stress and dietary factors, which significantly influence symptom severity (Boggio et al., 2025; Suárez et al., 2012). A system like LogChat could be adapted to support these patients as well. This adds another motivating factor to produce generalizable and transferable outcomes in conducting this thesis.

1.3. LogChat: A Conversational Approach to Support Pacing

Addressing the significant documentation burden that can accompany effective ME/CFS pacing calls for an innovative solution that is both accessible and energy-conserving. The emergence of advanced conversational artificial intelligence, particularly following the release of platforms like OpenAI’s *ChatGPT* (OpenAI et al., 2024) and Google’s *Gemini* (Comanici et al., 2025), presented a compelling technological opportunity. This sparked the central research question of whether such technology could be harnessed to design a system that transforms unstructured, natural language into a structured, longitudinal health diary. This thesis introduces LogChat, an implementation of a novel, memory-enhanced chatbot architecture designed specifically for this purpose.

While the long-term vision is a generalizable conversational companion for chronically ill patients—assisting with everything from emotional coping to analyzing health patterns—the scope of this thesis is intentionally focused on a foundational first step. LogChat was developed to demonstrate the technical feasibility of a system with a dual function: to reduce the logging burden for ME/CFS patients and to provide accessible, reliable educational support. Its primary functions are twofold: first, to enable a user to report symptoms and activities through natural, textual interaction, from which the system extracts structured data for a health diary. Second, it serves as an on-demand educational resource, providing grounded information about ME/CFS and pacing. Both functions are underpinned by a persistent memory of the user’s history to enable a personalized, empathetic, and context-aware dialogue. By validating this integrated

approach, this thesis seeks to provide a blueprint for a more holistic support tool, directly tackling the primary challenges in effective pacing management for individuals with ME/CFS. The research presented herein documents LogChat’s design, development, and evaluation as a proof-of-concept, establishing the architectural groundwork upon which a more comprehensive and clinically deployable tool can be built.

1.4. Thesis Overview and Contributions

This thesis documents the systematic design, development, and evaluation of the LogChat prototype introduced in the previous section. As a prototypical implementation of a memory-enhanced chatbot architecture, LogChat serves as a proof-of-concept for addressing the documentation challenges inherent in ME/CFS pacing management. The research presented herein was guided by the central research question: “How can current natural language processing technology be integrated into a system architecture to enable continuous, conversation-based tracking of activities and symptom severity for ME/CFS patients, thereby reducing the documentation burden of the pacing strategy?”

To address this question, the research was conducted following the Design Science Research Methodology (DSRM), the application of which is detailed in Section 3.2. The primary objectives were to design and develop the LogChat architecture and to rigorously evaluate its technical feasibility and functional correctness through simulated interactions. The key contributions of this thesis are:

1. The principal contribution is a novel, modular chatbot architecture (LogChat)¹ designed for personalized, memory-enhanced conversational logging in chronic disease management, specifically for ME/CFS patients.
2. A demonstration of the application of LLMs for the complex, orchestrated tasks of conversational data extraction, structured logging, and maintaining long-term user memory within such an architecture.
3. Insights into the capabilities and limitations of current LLMs, including a comparison be-

¹The complete LogChat source code, evaluation framework, and unabridged log files for this research are publicly available in the project’s GitHub repository: <https://github.com/Hanneseh/LogChat>.

tween proprietary models (e.g., Gemini) and smaller open-source alternatives, highlighting the performance required for this sensitive application.

4. The development and application of a novel evaluation methodology tailored for assessing the technical reliability of complex conversational systems, employing an *Impersonator* and a *Judge* LLM.

The remainder of this thesis is structured as follows: Chapter 2 provides a review of the theoretical background on ME/CFS, pacing, and relevant AI technologies, alongside an analysis of related work. Chapter 3 details the DSRM as applied to this project. Chapter 4 outlines the objectives defined for this thesis. Chapter 5 presents the complete system architecture and development process of LogChat. Chapter 6 describes the setup for the demonstration and evaluation of the prototype. Chapter 7 presents the results of this evaluation, including both quantitative and qualitative analyses. Chapter 8 discusses the findings, addressing the research question, and reflects on the contributions and limitations of this work. Finally, Chapter 9 concludes the thesis and outlines potential avenues for future research and development.

2. Background and Related Work

This chapter serves a dual purpose. First, it establishes the foundational knowledge required to comprehend the LogChat prototype, covering both the clinical context of ME/CFS and the essential concepts from the technological domain of conversational artificial intelligence. Second, it situates the LogChat system within the existing landscape of academic research and commercial solutions, thereby highlighting its unique contributions. The chapter begins by formalizing the clinical challenges of ME/CFS management that justify LogChat’s design. It then introduces the core AI technologies that serve as the system’s architectural building blocks, and concludes with a critical review of related work in the field.

2.1. The Clinical Demands of Pacing in ME/CFS

Individuals with ME/CFS face a profound struggle in managing their daily lives. One challenge lies in the immense cognitive load involved in the manual self-monitoring of symptoms and ac-

tivities. This task is made exceptionally difficult by the illness's characteristic debilitating fatigue and significant cognitive dysfunction, which manifests as difficulty with concentration, information processing, and short-term memory (Institute of Medicine et al., 2015). This can make the consistent and detailed logging that supports effective self-management an exhausting endeavor in itself.

The primary scientific model underpinning the pacing strategy is the Energy Envelope Theory (Jason et al., 2013). This theory posits that to minimize the severity of their symptoms and avoid exacerbations, patients must learn to balance their energy expenditure with their severely limited and often fluctuating available energy (Jason et al., 2013). The core principle is that by maintaining activities within this personal energy envelope, patients can sustain functioning and reduce the frequency of relapses (Jason et al., 2013). From this theory arises a clear rationale for a supportive system capable of helping patients meticulously track both their activities, which represent energy expenditure, and their symptoms, which represent the physiological consequences of that expenditure.

The most critical symptom to manage is Post-Exertional Malaise (PEM), the hallmark of ME/CFS (Institute of Medicine et al., 2015). PEM is defined as a delayed and disproportionate exacerbation of all of a patient's symptoms following even minimal physical, cognitive, or emotional exertion (Institute of Medicine et al., 2015; Stussman et al., 2020). The defining characteristic of PEM, and the crucial information-processing problem for patients, is its delayed onset. The worsening of symptoms frequently develops hours or even days after the triggering activity has ceased (Institute of Medicine et al., 2015; Stussman et al., 2020). This temporal disconnect makes it extraordinarily difficult for individuals to correlate cause (a specific activity) with its effect (a subsequent "crash"), obscuring the very patterns they need to identify to pace themselves effectively. This delayed feedback loop provides the explicit justification for a memory-enhanced logging system like LogChat, which is designed to help bridge this informational gap by maintaining a persistent record of activities and symptoms over time.

2.2. Technological Foundations of Conversational AI

This section provides the necessary technological background on conversational artificial intelligence to understand the design and implementation of the LogChat system detailed in Chapter 5. It begins by introducing LLMs, the core engine of LogChat, before delving into the specific techniques and architectural patterns leveraged in its construction, including prompt engineering, agentic systems, and Retrieval-Augmented Generation. Each concept is defined, its function within LogChat is explained, and the rationale for its selection is justified.

2.2.1. Large Language Models

At the core of the LogChat prototype lie Large Language Models (LLMs), a class of deep neural networks based on the Transformer architecture that have demonstrated remarkable capabilities in understanding and generating human-like text (Minaee et al., 2025; Zhao et al., 2025). Trained on vast corpora of text, they learn intricate statistical and semantic patterns, allowing them to probabilistically predict subsequent words to generate coherent and contextually relevant continuations for any given input (Minaee et al., 2025; Zhao et al., 2025). This predictive capability allows them to perform a wide array of language-based tasks, from simple text completion to complex reasoning and dialogue.

Two core capabilities that make modern LLMs particularly suitable for an application like LogChat are instruction-tuning and in-context learning. Instruction-tuning is a refinement process where a pre-trained model is further trained on a dataset of explicit instructions and their desired outputs (S. Zhang et al., 2024). This aligns the model’s behavior with user intent, enabling it to follow commands and perform specific tasks as directed in a prompt. In-context learning, conversely, refers to the model’s emergent ability to learn from examples provided directly within the prompt itself, without requiring any changes to its underlying parameters (Dong et al., 2024; Zhao et al., 2025). These capabilities are crucial for LogChat, as they allow for the complex orchestration of its various functions through carefully designed prompts rather than the more resource-intensive process of fine-tuning the model for each specific task.

The landscape of available LLMs is broadly divided into two categories: proprietary, Application Programming Interface (API)-based models and open-source models. Proprietary models, such

as the Generative Pre-trained Transformer (GPT) series from OpenAI or the Gemini family of models from Google, are typically the most powerful and are accessed via an API (Minaee et al., 2025; Xu et al., 2024). While they offer state-of-the-art performance, their use entails sending data to external servers, which raises significant privacy concerns for an application handling sensitive health information (Xu et al., 2024). In contrast, open-source models, such as the Llama series from Meta or the Qwen family from Alibaba, can be downloaded and run on local hardware (Minaee et al., 2025). This offers a path towards a completely private, offline-first application, aligning with the long-term vision for LogChat. This distinction is a central theme in this thesis, directly informing the performance evaluation in Chapter 7 and the discussion of the privacy-versus-usability trade-off in Section 8.2.

A key differentiator among LLMs is their size, measured by the number of parameters they contain. Generally, a higher parameter count correlates with greater capability but also demands significantly more computational resources (Jin et al., 2024; Wan et al., 2024). While the largest models may have trillions of parameters, the open-source community has produced highly capable models in the 4 to 14 billion parameter range (Frantar et al., 2023; Lin et al., 2024; Zhao et al., 2025). The increasing efficiency of these smaller models makes it feasible to run them on high-end consumer hardware, a prerequisite for realizing the vision of a private, on-device LogChat application.

The practical application of any LLM is constrained by its context window, which defines the maximum amount of text the model can process at one time (Naveed et al., 2024). This window acts as the model’s short-term memory; any information outside of it is lost for the current turn (Wang et al., 2024). Effectively managing this context window is a critical architectural challenge, especially in a stateful, memory-enhanced system like LogChat where conversation histories and user profiles can grow substantially. The LogChat architecture addresses this challenge with a two-pronged strategy. Firstly, to provide long-term memory without overwhelming the model, the system selectively retrieves only the five most recent interaction summaries. Secondly, to manage the immediate conversational context, the *Planner* and *Responder* nodes automatically trim the oldest messages from the conversation history, ensuring the messages always fit within a predefined token limit. This management is crucial, as the instructional prompts

themselves—which are measured in tokens (the basic units of text, roughly equivalent to words or parts of words)—are quite large, ranging from approximately 800 to 1500 tokens for the core system nodes.

The generation of text itself can be controlled via hyperparameters. The temperature parameter, for instance, controls the randomness of the output; lower values make the model more deterministic, while higher values encourage more diversity (Zhao et al., 2025). Similarly, top-p sampling provides another lever to balance predictability and creativity (Zhao et al., 2025). For the various agents in this thesis, the configuration of these parameters was directly tied to specific design choices. To ensure the deterministic and reliable behavior crucial for a logging tool, a low temperature was selected: 0.2 for the core LogChat system and a slightly higher 0.3 for the *Impersonator* agent to allow for natural conversational variance. In contrast, a high top-p value of 0.9 was consistently used to maintain linguistic diversity, allowing for creative and contextually appropriate responses.

Several optimization techniques have been developed to make the on-device deployment of LLMs feasible by addressing their high computational requirements. *Quantization*, for instance, reduces the numerical precision of a model’s parameters. This process can dramatically decrease its file size and memory footprint, enabling significantly faster text-generation on resource-constrained devices, often with only minimal performance degradation (Jin et al., 2024; Minaee et al., 2025). This technique is relevant to this thesis as the open-source models used in evaluating LogChat in Chapter 7 are quantized. Another relevant technique is *Knowledge Distillation*, where a large, powerful “teacher” model is used to train a smaller “student” model on a specific task (Xu et al., 2024). This technique, discussed in Chapter 9, represents a promising path for creating highly optimized, component-specific models for a future version of the system.

2.2.2. Prompt Engineering

While LLMs possess broad capabilities, steering their behavior to perform specific and consistent tasks requires a deliberate approach known as prompt engineering. This is the iterative process of designing, refining, and structuring the inputs, or prompts, given to an LLM to elicit a desired output (Minaee et al., 2025; Sahoo et al., 2025). It is an empirical discipline that requires an

understanding of the model’s mechanics to craft instructions that are unambiguous and contextually rich. For the LogChat system, prompt engineering is the primary method used to control its functionality. This approach was chosen over alternatives like *fine-tuning*—the process of further training a model on a domain-specific dataset—due to its flexibility and lower resource requirements, allowing for rapid iteration without the significant overhead associated with retraining (Sahoo et al., 2025; Wan et al., 2024).

Prompt Design Strategies in LogChat

The design of LogChat’s prompts represents a pragmatic synthesis of well-documented prompt engineering strategies (Sahoo et al., 2025; Schulhoff et al., 2025). A combination of techniques was employed to control the system’s behavior. Role Prompting was consistently used to assign clear personas and objectives, for instance, instructing a model to act as the “Planner component” (see Appendix A.2.3) or to “realistically play the role of a person with ME/CFS” (see Appendix A.6.1). This was complemented by direct Instruction Prompting and Few-Shot Learning, which provided models with explicit, numbered commands and concrete examples of desired outputs to guide their responses (see Appendix A.1.3). To enhance logical robustness, Chain-of-Thought (CoT) Prompting was operationalized by requiring models to output a “Rationale” alongside a “Suggestion”, forcing a step-by-step reasoning process (see Appendix A.2.4). The system’s modularity was ensured through the enforcement of strict Structured Output Formats, which made the outputs of components like the *Summarizer* reliably parsable (see Appendix A.5.2). Finally, Negative Prompting was used to explicitly constrain models, preventing undesired outputs like conversational filler (see Appendix A.4.1), which was vital for the automated processing of their outputs.

2.2.3. Agentic Architectures and Tool Use

While the term *chatbot* is often used colloquially, the LogChat system is more precisely defined as a conversational *agent* to reflect its advanced capabilities. Whereas a simple chatbot might engage in stateless dialogue, an agent is a more sophisticated system that leverages an LLM’s reasoning capabilities to operate autonomously towards a set of goals (Minaee et al., 2025; Wang et al., 2024). An agent can analyze a request, create a multi-step plan, and execute that plan, often

by using external *tools* (Schick et al., 2023). In this context, tools are external functions that the agent can call to interact with systems outside of its own internal knowledge, such as databases or APIs (Zhao et al., 2025). This ability to perceive, reason, plan, and act is what enables the complex functionality of LogChat. This agentic approach directly informs the architectural design of LogChat, particularly the separation of concerns into distinct nodes, each with a specialized role. The *Planner* node acts as the reasoning core of the agent, the *Responder* node formulates the final message, and the *Extractor* node uses logging tools to write structured data to the database. This modular process allows for more robust behavior than a single monolithic prompt could achieve. The implementation of this stateful, agentic system is made possible by modern software frameworks like *LangChain* and its extension, *LangGraph* (LangChain, 2025a, 2025b), which provide the necessary abstractions for defining agents, managing tools, and orchestrating complex conversational workflows (Minaee et al., 2025; Wang et al., 2024).

2.2.4. Retrieval-Augmented Generation and Vector Search

A significant challenge with LLMs is their tendency to “hallucinate”, or generate plausible but factually incorrect information (Zhao et al., 2025). To mitigate this risk and ensure that LogChat provides reliable educational content, the system incorporates a technique known as Retrieval-Augmented Generation (RAG). RAG enhances an LLM’s response by first retrieving relevant information from a trusted, external knowledge base and then providing that information to the model as context for generating its answer (Minaee et al., 2025; Sahoo et al., 2025). This grounds the LLM’s response in the specific, curated documents retrieved for the user’s query (Lewis et al., 2021). In LogChat, this mechanism is used by the *Planner* agent to answer user questions about ME/CFS and pacing.

The underlying mechanism of RAG relies on converting textual information into a numerical format that can be efficiently searched. This is achieved through text embedding models, which map text to a high-dimensional vector space where texts with similar semantic meanings are located closer to one another (Karpukhin et al., 2020). For the LogChat prototype, the open-source `nomic-embed-text` model is used to generate these vector embeddings (Nomic Team, 2024). These vectors are then stored in a specialized vector database. LogChat utilizes a PostgreSQL database equipped with the *pgVector* (pgvector, 2024) extension, which enables efficient storage

and querying of high-dimensional vectors. When a user asks a question, the system’s *Planner* node first formulates a precise search query based on the user’s intent. This query is then converted into a vector for a semantic search against the database. The most common metric for this is cosine similarity, which measures the angle between vectors to determine their semantic relevance (Karpukhin et al., 2020). The system retrieves the text passages corresponding to the most similar vectors and provides this text to the LLM to inform its final, grounded response.

2.3. Related Work

A review of the current landscape of digital health technologies and academic research reveals a critical opportunity for innovation in supporting patients with ME/CFS. While numerous tools and research projects exist, they often fail to address the specific, high-burden documentation challenge that can accompany the diligent application of the pacing strategy. This section first examines existing mobile health (mHealth) applications for ME/CFS to identify this gap. It then surveys the academic literature on conversational health agents to situate the architectural and methodological contributions of LogChat, and concludes by contextualizing the novel evaluation framework developed for this thesis.

2.3.1. Existing Digital Health Tools for ME/CFS

While data-centric trackers like *Visible* (Visible Health Inc., 2025) and *Solve Together* (Solve ME/CFS Initiative, 2025b) excel at quantitative collection, they do not solve the high cognitive load of data input via a conversational interface. Specifically, *Visible*’s focus on physical exertion metrics like heart rate does not explicitly account for cognitive or emotional exertion, which are also significant triggers for PEM (Visible Health Inc., 2025). Furthermore, the reliance on objective biometric data may overlook the critical role of subjective patient experience; research has shown that a patient’s self-perceived sleep quality is a better predictor of next-day fatigue than objective, actigraphy-defined sleep data (Russell et al., 2016). Conversely, educational platforms like *MyGuide Long COVID* provide curated resources but lack the continuous, dynamic tracking necessary for pacing management (Naik et al., 2024). This need for accessible, low-burden interfaces is corroborated by recent user-centered design research. A study by Taygar et al. (2025) on a digital intervention for chronic pain—a condition with significant overlap with ME/CFS, and

whose participants included individuals diagnosed with chronic fatigue syndrome (Taygar et al., 2025)—identified a clear demand for features that accommodate cognitive limitations and fatigue (Taygar et al., 2025). End-users in the study emphasized the need for simplified design elements, such as short “microsessions” and reduced text, and crucially proposed accessibility features like a text-to-speech function specifically to support users with cognitive impairments (Taygar et al., 2025). These findings empirically validate the central challenge that LogChat addresses: the high cognitive load of manual data entry for patients managing complex chronic conditions. LogChat is therefore designed to fill this specific gap by integrating a low-friction conversational diary with persistent memory, directly responding to these identified needs for an accessible and supportive self-management tool.

2.3.2. Conversational Agents in Healthcare

The academic landscape of conversational health agents (CHAs) has rapidly evolved, moving beyond simple, rule-based chatbots towards sophisticated, agentic systems powered by LLMs. Research in this area explores various architectural patterns and memory mechanisms to create personalized and effective healthcare companions. This section surveys key contributions, situating LogChat’s design within this context by thematically analyzing different approaches to system architecture and memory implementation.

A dominant trend is the development of modular, agentic architectures that separate conversational duties from complex reasoning. The design of LogChat was directly inspired by the *open-CHA* framework proposed by Abbasian et al. (2024), which exemplifies this approach. Their framework centers on an “Orchestrator” that uses a “Task Planner” to generate a multi-step plan and a separate “Response Generator” to formulate an empathetic answer (Abbasian et al., 2024). LogChat adopts this architectural pattern by implementing its own *Planner* and *Responder* nodes, affirming this separation of concerns as a robust approach for building complex CHAs. Similarly, the AMIE system, a state-of-the-art clinical agent, employs a dual-agent architecture inspired by Thinking, Fast and Slow, with a fast Dialogue Agent for real-time interaction and a slow reasoning agent for deep, evidence-based planning (Palepu et al., 2025). While these frameworks provide a generalized blueprint for integrating external tools and knowledge, LogChat builds upon this concept with a specific focus on internal memory synthesis, using its *Summarizer* and *Extractor*

nodes to create a rich, evolving user profile from the conversational data itself.

In addressing the critical challenge of data privacy, an alternative architectural pattern is proposed by Montagna et al. (2023), centered on the principle of data segregation. Their system uses rule-based parsers for sensitive data locally and employs an external LLM only for general, non-sensitive conversation, intentionally preventing the LLM from accessing clinical details (Montagna et al., 2023). This approach stands in direct contrast to LogChat’s principle of data integration, which provides the LLM with curated summaries to enable complex, context-aware functions. The choice between these patterns—segregation for privacy versus integration for functionality—thus represents a fundamental design trade-off in the development of CHAs.

Another key area of research focuses on enhancing the memory and personalization of CHAs. LogChat’s memory system is a direct implementation of the hierarchical structure proposed by Zhong et al. (2023) in their *MemoryBank* framework. This system creates a hierarchical memory from two core components: summaries of past conversations and a dynamic user profile that evolves over time. LogChat adopts this dual-component approach, using its *Summarizer* node to generate both interaction summaries and a persistent user profile. However, LogChat deliberately diverges from the *MemoryBank* framework in one critical aspect. The most novel feature of *MemoryBank* is a mechanism inspired by the Ebbinghaus Forgetting Curve, allowing it to selectively “forget” information to simulate a more human-like memory (Zhong et al., 2023). While this is suitable for a general-purpose companion, it contrasts sharply with the design philosophy of LogChat, which is a tool for clinical self-management where data integrity and completeness are paramount. Therefore, LogChat employs a persistent and complete memory store, ensuring no user-provided data is ever lost.

A more intricate method is introduced by K. Zhang et al. (2024) in their MaLP framework, which combines a neuroscience-inspired memory system with Parameter-Efficient Fine-Tuning (PEFT). This dual approach personalizes both the content of the conversation through a memory module and the behavioral style of the agent by fine-tuning the LLM’s parameters to match user preferences (K. Zhang et al., 2024). While sharing the goal of memory-enhanced personalization, MaLP’s reliance on fine-tuning contrasts with LogChat’s approach, which achieves all personalization exclusively through its modular agentic architecture and advanced prompt engineering,

deliberately avoiding the resource-intensive process of model retraining.

2.3.3. Evaluation of Conversational Systems

The evaluation of complex conversational systems presents a significant methodological challenge, for which a variety of approaches have been established in the literature (Minaee et al., 2025; Wang et al., 2024; Zhao et al., 2025). A primary paradigm is subjective, human-based evaluation, which is often considered the gold standard for assessing user-facing qualities like empathy, helpfulness, and overall satisfaction (Wang et al., 2024). This can range from direct human annotation and scoring of conversations to highly structured, domain-specific assessments. For instance, the Objective Structured Clinical Examination (OSCE) uses trained actors to simulate patient encounters and is a standard method for evaluating clinical dialogue systems (Palepu et al., 2025). While essential for measuring clinical utility and user experience, these human-centric methods are inherently expensive, time-consuming, and can be subject to population bias (Wang et al., 2024).

To complement human judgment, researchers employ objective, task-based metrics that measure an agent’s ability to successfully complete specific goals, often reported as success rates or accuracy scores (Yao et al., 2023; Zhao et al., 2025). For generative tasks like summarization, automated NLP metrics such as ROUGE are commonly used to compare a model’s output against a reference text (Zhao et al., 2025). However, these metrics often rely on surface-level text similarity and can fail to capture the semantic correctness of an open-ended conversational response (Minaee et al., 2025). More recently, a new paradigm of “LLM as a Judge” has emerged, which leverages powerful proprietary models like GPT-4 to automate the evaluation process (S. Zhang et al., 2024; Zhao et al., 2025). Frameworks such as *AlpacaEval* and *MT-Bench* utilize an LLM judge to perform pairwise comparisons or score model outputs, offering a scalable and efficient alternative to human annotators that has shown high correlation with human preferences (S. Zhang et al., 2024; Zhao et al., 2025).

While these established methods provide valuable insights, they do not fully address the unique evaluation requirements of a highly specialized, stateful, and longitudinal agent like LogChat, whose correctness depends on a complex interplay of memory, multi-step planning, and precise

tool use over extended interactions. To address this gap, this thesis developed a novel, automated evaluation framework centered on two specialized LLM agents: an *Impersonator* and a *Judge*. This framework moves beyond the evaluation of single-turn, stateless responses. The *Impersonator* agent simulates a dynamic, multi-session conversation, instantiating a specific persona with a narrative arc, thereby enabling the systematic testing of LogChat’s long-term memory and conversational state management. The *Judge* agent, in turn, performs a comprehensive technical audit of the complete interaction log. Guided by a detailed, per-interaction evaluation checklist, its role is not merely to assess the quality of the textual output but to verify the entire chain of system behaviors, including the correct invocation of tools with accurate parameters, and the proper creation of structured log entries in the database. This framework is not positioned as a replacement for human-centered clinical validation but as a crucial prerequisite. It offers a rigorous, reproducible, and fully automated methodology for validating the *technical reliability* and *functional correctness* of a complex agent’s architecture before undertaking the resource-intensive and ethically sensitive step of human trials.

2.4. Identifying the Research Gap

The review of existing digital health technologies and academic research reveals a distinct gap in supporting ME/CFS patients. Current mHealth applications focus on either quantitative data collection or educational content delivery, but they do not alleviate the high documentation burden associated with the detailed logging recommended for pacing through a low-friction, conversational interface (Naik et al., 2024; Solve ME/CFS Initiative, 2025b; Visible Health Inc., 2025). This need for accessible tools is empirically corroborated by user-centered design studies in overlapping patient populations, which highlight a demand for features that accommodate cognitive limitations and fatigue (Taygar et al., 2025). While academic literature explores modular conversational agents, these systems often target general companionship or clinical decision support rather than patient-led self-management, and may rely on resource-intensive fine-tuning (Palepu et al., 2025; K. Zhang et al., 2024; Zhong et al., 2023). This thesis addresses this gap by presenting LogChat, a novel, memory-enhanced conversational agent architecture engineered to reduce the documentation burden associated with the logging recommendations of the pacing strategy. By transforming unstructured, natural language into a structured, longitudinal health diary, its core

function is to directly support patients in the complex task of identifying their personal energy envelope and recognizing the delayed patterns of PEM, thereby empowering them to manage their condition more effectively.

3. Methodology

The research conducted in this thesis employs the Design Science Research Methodology (DSRM). Specifically, the nominal process delineated by Peffers et al. (2007) was systematically followed from initiation to completion. The subsequent sections will introduce DSRM, provide the reasoning on why it was chosen, and outline the specific application and representation in this thesis.

3.1. Introduction to Design Science Research

DSRM, as introduced by Peffers et al. (2007), offers a unified mental model for conducting, presenting, recognizing, and evaluating Design Science (DS) work. By synthesizing foundational work from various authors in the Information Systems (IS), Engineering, and Computer Science disciplines, Peffers and his colleagues established DSRM as a process model comprising six nominal steps: 1) Problem Identification and Motivation, 2) Definition of the Objectives for a Solution, 3) Design and Development, 4) Demonstration, 5) Evaluation, and 6) Communication. Peffers et al. (2007) argued for the necessity of DSRM as a framework to accelerate the adoption of DS within the IS field as a legitimate research paradigm. Previously, the dominant research paradigms in IS were largely descriptive or explanatory, akin to those in the social and natural sciences. DS, in contrast, provides an avenue for addressing unsolved problems by purposefully creating novel artifacts and rigorously evaluating their effectiveness in resolving those problems. This paradigm empowers researchers to engage creatively in constructing new solutions from existing theories and technologies. Crucially, it embeds this creative work within a rigorous process, ensuring that the research contributes meaningfully to the knowledge base from which future endeavors can draw. Because of these attributes, DS was selected as the overarching paradigm and DSRM as the specific methodology for conducting the research presented in this thesis.

3.2. Application of the Design Science Research Methodology

The DSRM, as delineated by Peffers et al. (2007), provided the overarching framework for this research. The nominal six-step process was generally followed sequentially from problem identification through to communication. However, the application of DSRM in this thesis was also characterized by a crucial iterative cycle. After an initial iteration of “Design and Development” (Step 3), practical challenges encountered during the development of the first LogChat prototype necessitated a return to “Definition of the Objectives for a Solution” (Step 2). This refinement of objectives, informed by early developmental learnings, then guided subsequent design and development efforts before proceeding with the demonstration, evaluation, and communication phases.

- Chapter 1 (Introduction) encapsulates the activities and outcomes of DSRM Step 1: Problem Identification and Motivation.
- Chapter 4 (Objectives for a Solution) details the iteratively defined objectives for LogChat, representing the outputs of DSRM Step 2.
- Chapter 5 (LogChat System Architecture and Development) describes the LogChat prototype, the primary artifact resulting from DSRM Step 3: Design and Development. This chapter forms a core contribution of the thesis.
- Chapter 6 (Demonstration and Evaluation Setup) outlines the procedures and setup for DSRM Step 4: Demonstration, and DSRM Step 5: Evaluation.
- Chapter 7 (Results) presents the outcomes of the demonstration activities (Step 4).
- Chapter 8 (Discussion) critically analyzes the demonstration results, thereby representing the outcomes of the evaluation phase (Step 5), and discusses these in the context of the defined objectives and overall research question.

The entirety of this thesis document, along with the publicly available LogChat codebase, serves as the fulfillment of DSRM Step 6: Communication.

4. Objectives for a Solution

This chapter details the objectives for the LogChat system. This process corresponds to the second step of the DSRM, “Definition of the Objectives for a Solution” (Peppers et al., 2007). Derived from the problem statement outlined in Section 1.1, these objectives serve a dual purpose: they steer the subsequent design and development of the LogChat artifact and establish the core criteria for its evaluation.

4.1. Core Use Cases and Functional Objectives

The refined objectives were structured around three key use cases for LogChat. These use cases were selected to demonstrate the core feasibility of creating a conversational diary for PEM management and to provide a clear framework for both development and evaluation. The successful enablement of these use cases would signify LogChat’s efficacy as a proof-of-concept, demonstrating that natural interaction can yield a structured digital diary to help users identify correlations between their activities and symptom severity over time. The following subsections detail each use case and the specific functional objectives required to realize it.

4.1.1. Use Case 1: Conversation

The foundational use case is to enable a human-like, memory-aware conversation that is supportive and cognizant of the user’s condition. The system must be more than a simple command-line tool; it must act as an empathetic companion. This is achieved through the following objectives:

- **Conversational Interaction:** The system must engage in natural, turn-based textual conversation. Crucially, it must be cognizant of the user’s energy limitations by avoiding lengthy responses and not posing multiple questions simultaneously, thereby minimizing cognitive load.
- **Short-Term Memory:** To ensure conversational coherence, the system must maintain awareness of the content of recent interaction turns and the main discussion points within the current session.

- **Long-Term Memory:** The system must develop and maintain a persistent user profile that stores key information about the user's condition, routines, and significant events. This memory is essential for providing personalization and context-aware interaction across multiple sessions over time.

4.1.2. Use Case 2: Logging

The central utility of LogChat is to reduce the documentation burden of pacing by transforming conversational input into a structured, digital diary. This involves not only direct extraction but also intelligent automation to minimize user effort. The objectives for this use case are:

- **Symptom and Activity Logging:** The system must accurately identify, extract, and log user-reported symptoms and activities. This includes capturing their perceived severity or effort (e.g., on a scale from 1 to 10), timing, and duration from the conversational context.
- **Simplified and Baseline Logging:** To further reduce user effort, the system must allow for simplified logging by referencing previously described routines or states (e.g., "my fatigue is the same as yesterday"). It must also implement a mechanism to account for days when the user does not interact, for instance, by logging a pre-defined baseline activity level based on the user's established routines to ensure a complete record.
- **Structured Data Output:** The ultimate goal of the logging process is to convert the conversational input into structured, queryable data (e.g., database entries). This structured output is the prerequisite for later analysis and visualization, which is intended to help the user identify patterns.

4.1.3. Use Case 3: Question Answering

Beyond logging, LogChat should empower the user by providing access to both general knowledge and their own personal data, helping them to better understand their condition and make informed decisions. This is supported by two objectives:

- **Information Retrieval and Education:** The system must be able to provide relevant educational information about ME/CFS, PEM, and pacing strategies. This is achieved through

a RAG mechanism that searches a curated knowledge base.

- **Activity Score Retrieval:** To provide objective feedback, the system must allow the user to query their past activity load. This involves retrieving logged data and presenting it in a clear, understandable format, such as a calculated daily score.

These functional objectives are operationalized and validated through the evaluation dataset, which comprises specific personas, interaction scenarios, and corresponding checklists that enumerate expected system behaviors. This provides a concrete methodology for assessing whether LogChat meets its defined goals.

4.2. Non-Functional Objectives

To guide the architectural design towards a solution that not only serves as a prototype but also as a foundation for a potentially deployable system, two non-functional objectives were established. These principles describe how the system should perform its functions and cater to future goals beyond the scope of this thesis:

- **Modularity:** The system architecture should be modular to facilitate easy development, maintenance, and the potential future expansion of components.
- **Trustworthiness, Privacy, and Efficiency:** While the prototype may use proprietary LLMs, the long-term vision is an application that can be entrusted with sensitive personal data, ideally operating offline on the user's device. The architecture should therefore be designed with this goal in mind, aiming to support efficient, local open-source models and relying on robust prompt engineering rather than extensive fine-tuning.

These non-functional objectives were not directly tested but were respected during the design and development phase and are reviewed in the discussion in Section 8.2.

5. LogChat System Architecture and Development

This chapter details the design and development of the LogChat prototype, the central artifact of this Design Science Research study. This work constitutes the “Design and Development” phase (Step 3) of the DSRM framework (Peppers et al., 2007) and was directly guided by the objectives established in Chapter 4. The final LogChat architecture presented herein is the result of an iterative cycle. The chapter will first briefly outline the initial design iteration and the critical learnings it provided. The remainder and primary focus of the chapter is then a detailed exposition of the definitive LogChat system.

5.1. Initial Design Iteration and Objective Refinement

The development of LogChat followed an iterative path, as is common in DSRM projects. The initial design phase was instrumental in revealing practical challenges that led to a significant refinement of the project’s objectives. An early development direction was to only use local, open-source LLMs; however, these models proved insufficiently reliable for the complex tasks required, necessitating a pivot to more stable, proprietary LLMs. This first prototype, LogChat Version 1, attempted to use the LLM’s tool-calling capabilities to continuously create and update a structured database *during* an active conversation. This “live logging” approach, combined with an excessive number of granular and complex tools for different log types (e.g., symptoms, consumption, experiences), proved untenable. The models struggled to correctly use the intricate tools for both creating and updating logs, leading to frequent errors, redundant entries, and inconsistent logging behavior.

The challenges encountered with this initial architecture prompted a return to Step 2 of the DSRM framework to redefine the use cases and objectives. This process culminated in three principal learnings that directly shaped the final LogChat architecture presented in the subsequent sections. First, tool calling is a potent mechanism for achieving structured logging from conversational data. Second, a multitude of complex tools presents significant handling difficulties, even for powerful proprietary LLMs; to be effective, tools should be few and simple. Finally, achieving accurate logging, particularly with updates during an active conversation, is challenging; it is

more effective to segregate the process into distinct interaction and post-interaction phases.

5.2. The LogChat System Architecture

The final version of the LogChat prototype is a modular, memory-enhanced conversational system exhibiting agentic properties and incorporating a RAG component. Its architecture can be conceptually organized into four distinct layers, as illustrated in Figure 5.1.

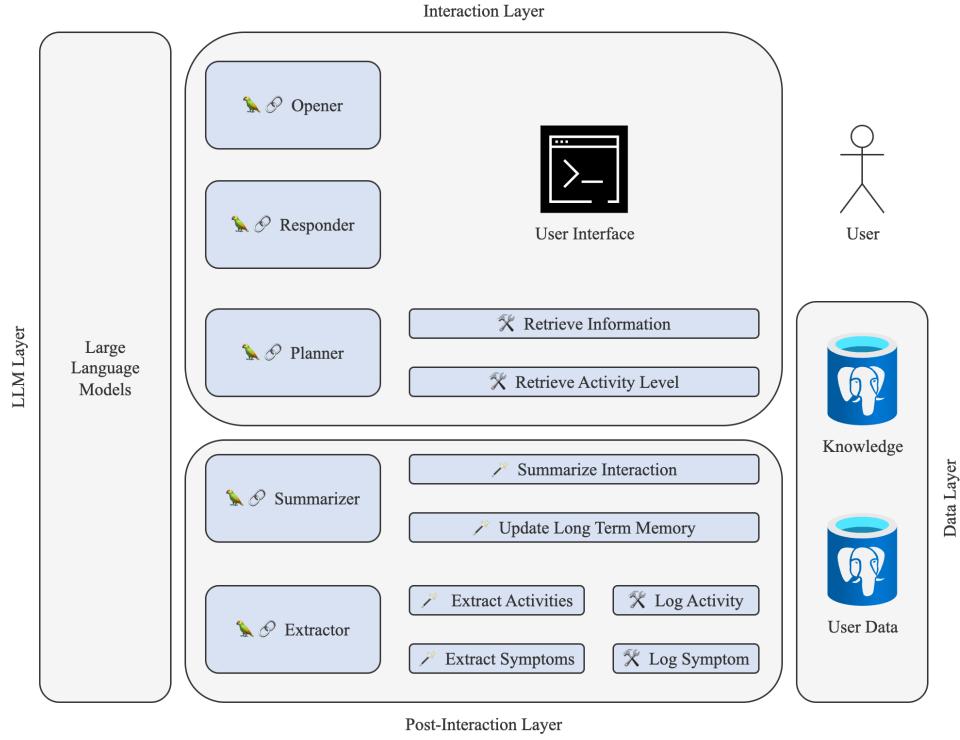


Figure 5.1. Conceptual organization of LogChat’s architectural components. The architecture is divided into four layers: 1) Interaction Layer, 2) Post-Interaction Layer, 3) Data Layer, and 4) LLM Layer, each containing specific nodes, tools, and data stores.

The LogChat system is built on a modular, four-layer architecture, designed for flexibility and extensibility. The foundational **Data Layer** uses a PostgreSQL database with the *pgVector* extension to store user data and contextual knowledge. The **Interaction Layer**, implemented with *LangGraph*, manages the live dialogue: a strategic *Planner* node determines conversational goals, tools retrieve relevant information, and a *Responder* node formulizes the user-facing message. Once a conversation concludes, the **Post-Interaction Layer** activates to summarize the dialogue for long-term memory and use an *Extractor* node to log structured symptoms and activities. Fi-

nally, the abstracted **LLM Layer** serves as the system’s core processing engine, enabling the agentic behavior of the various nodes. The following sections will detail the components within each of these layers, beginning with the foundational LLM Layer.

5.3. LLM Layer

The LLM Layer is the conceptual foundation of the LogChat architecture, representing the core processing engine that drives all agentic behavior. This layer is not a single, monolithic component but rather an abstraction that leverages *LangChain* to remain model-agnostic. This design choice makes it possible to integrate any LLM provider that supports tool-calling, including proprietary models from Google and OpenAI or open-source models served via *Ollama* (Ollama, 2025b). All models used within the system are governed by consistent hyperparameter settings, as detailed in Section 2.2.1, to ensure controlled and predictable behavior across the different nodes. This architectural abstraction was crucial for conducting the comparative performance evaluation presented in Chapter 7.

5.4. Data Layer

The Data Layer consists of two schemas that are designed to be independent. The schemas are implemented using SQLAlchemy (SQLAlchemy, 2024a) as the Object-Relational Mapper (ORM) and are managed with Alembic (SQLAlchemy, 2024b) for database migrations. The following sections present these schemas, which provide the essential context for LogChat’s prompts and are acted upon by its tools.

5.4.1. User Data Schema

The user data schema, as depicted in Figure 5.2, comprises three interconnected tables designed to store all user-specific information.

The `logchat_user` table holds the core user profile, containing the user’s `id`, `name`, and `description`. The user’s name is made visible to all nodes in the *Interaction Layer* for personalization. The `description` attribute, which is initially empty, serves as the user’s long-term memory profile. After each interaction, it is updated by the *Summarizer* node with a

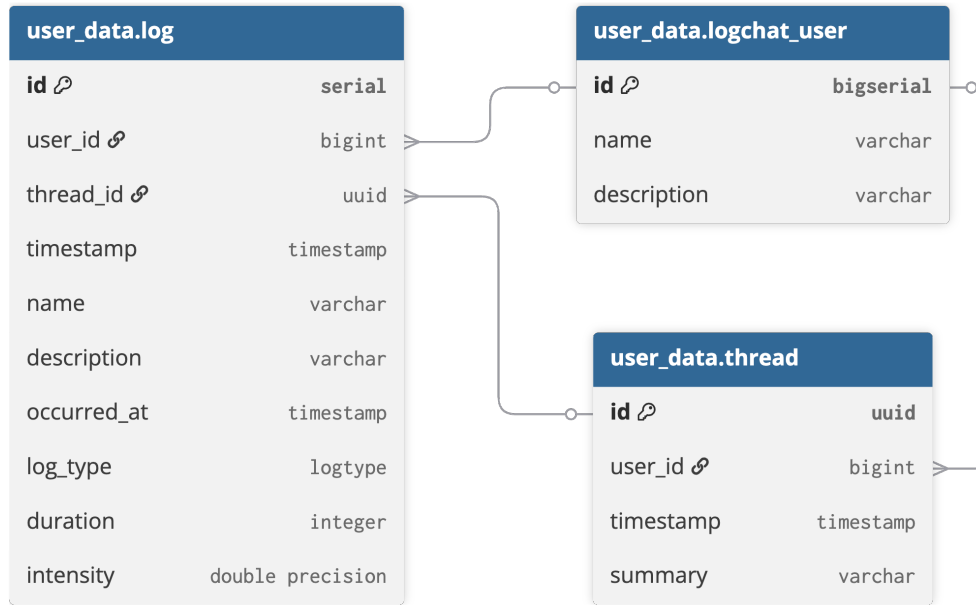


Figure 5.2. User Data Schema

consolidated narrative of the user’s condition and history. This profile is retrieved once at the beginning of a session and ingested as context into most prompts.

The `thread` table stores metadata for each distinct user interaction. The `user_id` links the interaction to a specific user, while the `timestamp` marks the start time of the conversation. Its `summary` field holds the concise overview of the entire interaction generated by the *Summarizer* node. During the initialization of a new session, the five most recent interaction summaries for this user are retrieved and compiled into a single context string using the `thread_summary_to_string` function (see Appendix C.3.2). This explicit limitation to the five most recent summaries is a key strategy for managing the LLM’s context window, ensuring that the model receives relevant recent history without being overloaded by the user’s entire interaction log. This function is also designed to insert placeholders, such as “No interaction logged for YYYY-MM-DD”, for any days missed between sessions. This mechanism provides the LLM with a heightened awareness of the user’s recent history, enabling it to target questions about symptoms and activities during those periods. A full example of a prompt containing the user description and this interaction summary string can be found in the *Opener Prompt Example* in Appendix B.1.1.

The `log` table is the most detailed, containing the structured records extracted by the *Extractor* node. Several attributes in this table are populated systemically: the `user_id` and `thread_id`

are injected based on the active session, while the `timestamp` and `log_type` are set upon insertion depending on which tool was called. The remaining attributes—`name`, `description`, `occurred_at`, `duration`, and `intensity`—are provided by the LLM as arguments in a tool call. The `name` and `description` are free-text fields for which the LLM must infer appropriate values. The `duration` must be an integer representing the time in minutes an activity was performed or a symptom persisted. The LLM must also infer a specific timestamp for when the event occurred `_at`. The `intensity` attribute is a floating-point number representing a subjective rating. In the case of a symptom, it reflects the user’s perceived severity; for an activity, it represents the perceived effort. In either case, LogChat is designed to actively elicit these 1-to-10 scale ratings whenever an activity or symptom is mentioned.

5.4.2. Knowledge Base

The knowledge base schema consists of a single table, also named `knowledge`, with four attributes, as shown in Figure 5.3.

knowledge.knowledge	
id 🔗	serial
source	varchar
headline	varchar
content	text
embedding	vector(768)

Figure 5.3. Knowledge Schema

The database is equipped with the *pgVector* extension, which allows for efficient similarity searches using the cosine distance metric between a query vector and the stored embeddings. This mechanism, which is utilized by the *Planner* node’s `retrieve_information` tool (see Appendix C.1.2), is detailed further in the subsequent sections (see Section 5.5.3). The knowledge table is not updated at runtime; it is populated in a one-time setup process. For the prototype, the knowledge base was populated using information from four pacing advice handouts provided by

reputable, English-language patient organizations: Solve ME/CFS Initiative (2025a), The ME Association (Shepherd & Mayes, 2023), Patient-Led Research Collaborative (2025), and ME Action (Seltzer, 2021). To automate the extraction of this information, the source documents were processed by a LLM (Google’s Gemini 2.5 Pro). The model was manually prompted to extract, summarize, and structure the content into CSV files containing *source*, *headline*, and *content* columns, effectively delegating the chunking strategy to the LLM. In addition to these patient resources, a fifth CSV file containing information about LogChat itself was created, as the *Planner* node was observed to query the knowledge base for self-referential information during simulated user requests. The curation of the knowledge base was deliberately simplified for this prototype; the primary focus was on establishing the architectural viability of the RAG mechanism rather than on extensive manual curation of the knowledge content. All source CSV files and the script for populating the database are available for inspection in the project’s GitHub repository.

5.5. Interaction Layer

The *Interaction Layer* encompasses all components that are active during a single user session, where a complete ‘interaction’ is composed of multiple back-and-forth conversational ‘turns’ between the user and LogChat. It is responsible for managing the real-time dialogue, from greeting the user to generating context-aware responses. The presentation of each component in the following sections—the user interface, *Opener* Node, *Planner* Node, and *Responder* Node—follows the logical flow of a single first conversational exchange.

5.5.1. User Interface

The user interface in the LogChat prototype is a minimalistic Command-Line Interface (CLI) application created with the *Typer* Python library (fastapi, 2024), as shown in Figure 5.4. It was designed for simplicity to allow for direct interaction with the system’s backend. The CLI does not feature login logic or user management; the active user is specified via a database ID set in an environment variable. This interface solely permits direct textual input to LogChat and displays its text-based responses.

While the implemented UI is rudimentary, its development was guided by a distinct conceptual

```

hannesehringfeld@Mac LogChat % poetry run python -m cli
LogChat: Hi Charlie, is your fatigue still at a 9/10?
You: No it improved, after resting the past days I would say its down to a 6
LogChat: That's great to hear that your fatigue has improved. Can you tell me what you did to
rest and recover over the past few days?
You: Just the bare minimum. I was eating precooked meals and did two guided meditations to ca
lm my nervous system down. I think they are great.
LogChat: I'm glad the pre-cooked meals and meditations helped you to recover. How long did ea
ch guided meditation session last, and what effort rating (1-10) would you give to each sessi
on?
You: I picked the five-minute versions in the app, and they actually took less effort than I
had thought. I did not pressure myself to follow every instruction and managed to relax with
eyes closed.
LogChat: It's great that you found a meditation app that works for you. Could you tell me whi
ch app you used for the guided meditations?
You:

```

Figure 5.4. LogChat Command Line Interface

vision for a future, fully-featured application. This vision, which heavily influenced the design decisions for LogChat’s core architecture, imagined a voice-first mobile app. Opening the application would initiate an interactive voice mode, where LogChat would begin the conversation with a personalized opening message. The user would engage in several conversational turns and then quit the interaction mode or close the app. This action would trigger the system to post-process the latest interaction and extract relevant information. Consequently, the next time the user opened the app, LogChat would have memory of the past interaction and be aware of the time that had elapsed.

5.5.2. Opener Node

To underscore LogChat’s character as a supportive companion, the *Opener* node is designed to offer a personalized greeting upon application launch. This component is tasked with delivering an initial message to the user immediately after they open the interface. Its behavior is adaptive, contingent on whether the user is interacting with the system for the first time or is returning after a previous session. For individuals engaging with LogChat for the first time, the *Opener* node presents a standardized, hardcoded welcome message. This message is defined in the `first_opener` variable (see Appendix A.1.1). It serves to introduce LogChat’s primary purpose—conversational logging for individuals with ME/CFS and Long Covid—and to explain the core interaction model, which involves logging symptoms and activities while emphasizing the importance of user-provided effort and severity ratings. The greeting also highlights the po-

tential benefits of using LogChat, such as tracking patterns and supporting PEM management, and concludes by inviting the user to share information about their daily routines and functional capacity.

For users with an established interaction history, the *Opener* node utilizes an LLM. The model’s behavior is guided by the `opener_system` prompt (see Appendix A.1.2) and the `opener_instructions` prompt (see Appendix A.1.3), which instruct it to craft a brief, personalized, and empathetic greeting. To achieve this, the LLM leverages several pieces of contextual information: the user’s name, their long-term memory profile, and summaries of recent interactions. The system is specifically designed to note any skipped days, which are identified through placeholders such as “No interaction logged for YYYY-MM-DD” within the interaction summaries. This allows the LLM to generate a concise, situationally aware opening, which typically takes the form of a single question related to the user’s last interaction or an observed gap in logging. This approach aims for a gentle re-engagement that is mindful of the user’s potential energy limitations. An example of a generated opening message is shown in Figure 5.4, and a complete example of the prompt used to generate such a message can be found in the *Opener Prompt Example* in Appendix B.1.1.

5.5.3. Planner Node

The *Planner* Node serves as the analytical core of LogChat’s conversational engine, determining the most appropriate next step for the system to take. It receives the complete conversation history, the user’s long-term memory profile, summaries of recent interactions, and the current time. Guided by the `planner_system` prompt (see Appendix A.2.3) and the `planner_instructions` prompt (see Appendix A.2.4), its primary role is to reason about the user’s input and plan LogChat’s subsequent actions. This process includes identifying mentions of symptoms, activities, or routines, as well as recognizing direct questions posed by the user. Based on its analysis, the *Planner* can execute tools to retrieve necessary information. For instance, it can look up past activity scores or search the knowledge base to answer questions about pacing. The results from any executed tools are then fed back into a second, tool-less prompt to the *Planner*, which is guided by the slightly modified `planner_instructions_with_tool_results` prompt (see Appendix A.2.5). This two-step process forces the *Planner* to first gather data and then formulate

a strategic suggestion for the *Responder* based on that data. For new users, where no history exists, hardcoded placeholders (see Appendix A.2.2 and A.2.1) are inserted into the prompt to heighten the *Planner*'s awareness of the need to ask foundational questions. In such cases, the *Planner* prioritizes providing information about pacing and the use of LogChat. The final output of the *Planner* is always a clear, actionable suggestion for the *Responder*, accompanied by a rationale for its decision. This kind of dynamic between the *Planner* and *Responder* can be seen in the debug log excerpt provided in Appendix C.5.1. To manage the LLM's finite context window, the conversation history provided to the *Planner* is dynamically pruned using a trimming mechanism that removes the oldest messages, ensuring the total history does not exceed a predefined token limit.

The *Planner* Node has access to two tools to gather information. The `retrieve_activity_level` tool (see Appendix C.1.1) accepts a start and end date as parameters. It retrieves all activities logged for the user within that period and calculates a daily activity score using the following formula:

$$S_d = \sum_{a \in A_d} (D_a \times E_a) \quad (5.1)$$

where S_d is the total activity score for a given day d , A_d is the set of all activities logged on that day, D_a is the duration of an activity in minutes, and E_a is its perceived effort rating (scale from 1 to 10). This score serves as a conversational metric and is envisioned as a user-viewable feature in a future version of the system, offering an at-a-glance assessment of a user's overall activity level. The formula was intentionally kept simple for this prototype to demonstrate the mechanism; its precise clinical accuracy was not the focus at this stage. Crucially, both the elicitation of 1 to 10 scale ratings and the subsequent calculation of this score serve primarily to demonstrate the system's technical capability to transform conversational input into structured, analyzable data, and do not represent a clinically validated method for patient assessment.

The `retrieve_information` tool (see Appendix C.1.2) searches the knowledge schema for information related to ME/CFS, Long Covid, PEM, LogChat, or pacing when the user asks a relevant question. The *Planner* provides a textual query, which is first embedded using the `nomic-embed-text:v1.5` model via a local *Ollama* server. A vector search is then performed against the knowledge base using cosine similarity. The top three results are returned and filtered

against a similarity threshold, which was empirically determined to be 0.4 by testing with queries designed to have no relevant match. This ensures that if no relevant information is found, the tool returns a standard “not found” message. The returned content is formatted into a single string that includes the headline and source for each knowledge entry. Examples of *Planner* prompts with specific values ingested can be seen in the *Initial Planner Prompt Example* (Appendix B.2.1) and the *Planner Prompt with Tool Results Example* (Appendix B.2.2). The resulting output from the *Planner* is then passed to the *Responder*, as illustrated in the prompt examples in the *Responder Prompt Example* (Appendix B.3.1) and the *Responder Prompt with Tool Results Example* (Appendix B.3.2).

5.5.4. Responder Node

The *Responder* Node is tasked with crafting the final message that LogChat presents to the user, effectively acting as the system’s voice. Guided by the *Responder System Prompt* (see Appendix A.3.1) and its corresponding *Responder Instruction Prompt* (see Appendix A.3.2), it synthesizes the strategic suggestions and any retrieved data from the *Planner* into an empathetic, concise, and helpful response. The node’s core function is to balance a supportive tone with efficient communication, recognizing the potential energy limitations of the user. Similar to the *Planner*, the *Responder* node also employs a message trimming strategy to keep the conversational history within a predefined token limit. As directed by the *Planner*, it prioritizes asking clear, focused questions to elicit specific, loggable details, such as an activity’s duration or a symptom’s intensity on a 1-to-10 scale. It avoids vague, open-ended questions. If the *Planner* provides educational content, the *Responder* presents it briefly and directly, attributing the information to its knowledge base (e.g., “My information suggests...”). Complete examples of prompts for the *Responder*, which also display its output from a conversation history, can be seen in the *Responder Prompt Example* (Appendix B.3.1) and the *Responder Prompt with Tool Results Example* (Appendix B.3.2).

5.6. Post-Interaction Layer

After the user ends an interaction, a series of automated post-interaction routines is executed. These routines are crucial for transforming the unstructured conversational data into structured

logs and for updating the system's persistent memory. This automated workflow unfolds in a specific sequence to ensure data integrity and context preservation. First, the *Summarizer* node performs its two critical functions: it generates a concise, bullet-point summary of the just-concluded interaction, and then it updates the user's comprehensive long-term memory profile by integrating insights from the latest interaction summaries. Following this summarization, the *Extractor* node analyzes the conversation transcript twice: once to identify and log all mentioned symptoms, and a second time for all activities.

5.6.1. Summarizer Node

The *Summarizer* component is responsible for creating concise summaries of interactions and maintaining an up-to-date long-term memory profile for the user. Its behavior is guided by the *Summarizer System Prompt* (see Appendix A.5.1). This node performs two distinct functions in sequence during the post-interaction phase.

Interaction Summarization: Using the *Summarize Interaction Prompt* (see Appendix A.5.2), the node first generates a bullet-point summary of the single, just-concluded interaction. The summary is structured under predefined headings: Condition Trend, Key Symptoms & Details, Key Activities & Details, User Concerns/Focus, Strategies/Treatments Mentioned, and Significant Events/Deviations. Any ratings or durations mentioned in the conversation are included. These summaries form the memory of past interactions that is provided as context to the *Opener* and *Planner* nodes in subsequent sessions. A full example of this prompt can be seen in the *Summarize Interaction Prompt Example* in Appendix B.4.1.

Long-Term Memory Profile Update: Following the interaction summary, the node uses the *Update Long Term Memory Prompt* (see Appendix A.5.3) to update the user's persistent profile. This process integrates new or changed information derived from the latest five interaction summaries (including the one just generated) into a coherent narrative. The output is a complete, updated profile written in the first person and organized under specific headers: About Me, Current Condition Status, Current Functional Capacity, Daily Routine, Activities I Enjoy/Attempt, My Typical Activity Level, and My Preferred Interaction Style with LogChat. This profile represents the system's evolving, holistic understanding of the user. A full example of this prompt

can be seen in the *Update Long Term Memory Prompt Example* in Appendix B.4.2.

5.6.2. Extractor Node

The *Extractor* is a specialized component responsible for analyzing the conversation transcript to identify and structure loggable information. Guided by the *Extractor System Prompt* (see Appendix A.4.1), it processes the conversation, the user’s long-term profile, and interaction summaries. Its sole output is a series of tool calls to log extracted data; it does not generate conversational text. The node is executed twice, once for each extraction task.

Symptom Extraction: When tasked with extracting symptoms, the node is guided by the *Extract Symptoms Instructions Prompt* (see Appendix A.4.2). It identifies explicit mentions of symptoms and uses contextual clues to interpret vague references, focusing on core ME/CFS symptoms (e.g., fatigue, brain fog) and sensory sensitivities. For each identified symptom, it extracts all necessary details to form a complete log entry, including a consistent name, any user-provided description, the inferred date and time it occurred_at, its perceived intensity on a 1-to-10 scale, and its duration in minutes (e.g., 1440 for an all-day symptom). It then uses the log_symptom tool (see Appendix C.2.1) to record this information. A full example of this extraction prompt can be seen in the *Extractor Extract Symptoms Example* in Appendix B.5.2.

Activity Extraction: When extracting activities, the node is guided by the *Extract Activities Instructions Prompt* (see Appendix A.4.3). It identifies specific user-mentioned activities; notably, “Resting” is not logged, as it is considered the absence of a loggable activity. For each identified activity, it extracts a consistent name (e.g., “Walking”, “Assumed Daily Routine”), any additional description, the inferred occurred_at date and time, a rating on a 1-to-10 scale for perceived effort, and the duration in minutes. It then calls the log_activity tool (see Appendix C.2.2). A full example of this extraction prompt can be seen in the *Extractor Extract Activities Example* in Appendix B.5.1.

A crucial function of the *Extractor* is to ensure baseline activity logging. For any day, but especially for those missed days where the user did not interact with LogChat, the system logs an “Estimated Baseline” activity. This is done to capture the effort required for basic needs like eating, personal hygiene, and moving within the home. This baseline activity is derived from the

user’s long-term memory profile, making it personalized to their specific routine and functional capacity. This mechanism was deemed necessary to ensure that the daily activity score is not zero for days without an interaction and more importantly, to factor in the basic activities that consume a portion of the user’s daily energy envelope. As described in the introduction of this thesis, even such minimal activities can be significant for many ME/CFS patients.

6. Demonstration and Evaluation Setup

This chapter outlines the methodology and framework established for the demonstration and evaluation of the LogChat prototype, corresponding to Steps 4 (Demonstration) and 5 (Evaluation) of the DSRM. It details the components, procedures, and criteria used to assess the prototype’s performance against the objectives defined in Chapter 4. The results of the demonstration and evaluation are presented and discussed in Chapters 7 and 8.

6.1. Overview of the Evaluation Framework

Given the complexity of simulating meaningful human-chatbot interactions and the necessity for objective, repeatable assessment, a comprehensive, automated evaluation framework was developed to validate the technical reliability and functional correctness of the LogChat prototype. This framework, depicted in Figure 6.1, integrates several key components to systematically conduct the demonstration and evaluation. The process begins with the *Evaluation Dataset*, which contains detailed persona descriptions and interaction specifications. This dataset provides the script for an LLM-based *Impersonator*, which emulates user behavior in conversation with the LogChat prototype. Each interaction is meticulously documented by LogChat into transparent log files. Subsequently, another LLM-based agent, the *Judge*, analyzes these interaction logs against evaluation checklists from the dataset, producing an annotated version of the logs. These annotations are then parsed to calculate a quantitative performance score for each of the 15 simulated interactions. The individual scores are averaged to yield an overall performance metric. As a final output, the structured data extracted by the LogChat prototype across all interactions for a given persona is used to generate overview plots, visually demonstrating the architecture’s capability to transform conversation into a structured health diary.

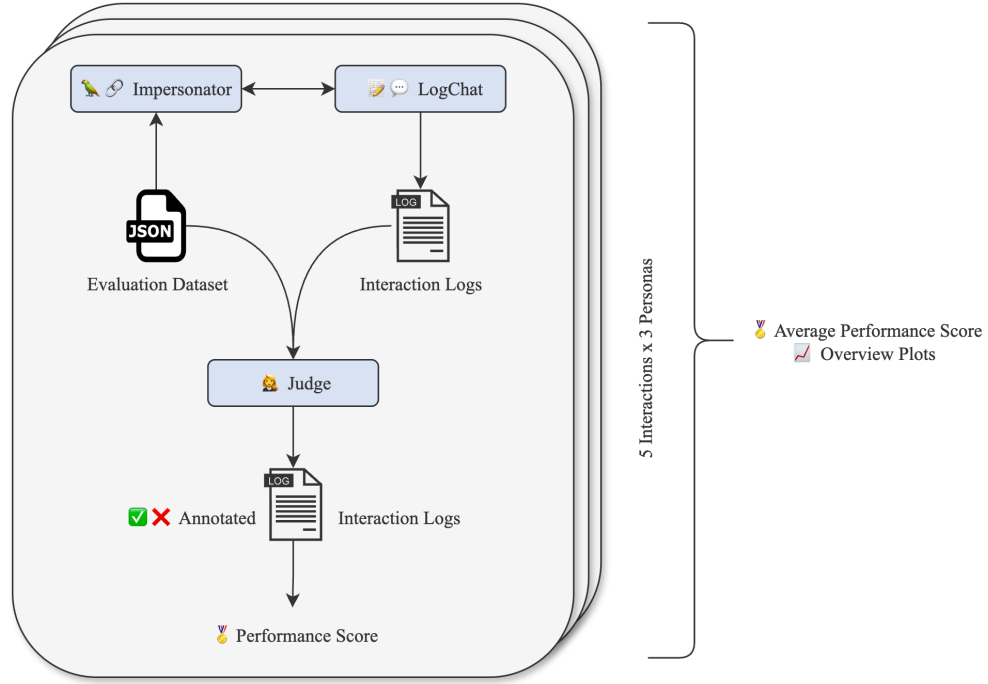


Figure 6.1. The conceptual diagram of the automated evaluation framework. The process begins with the *Evaluation Dataset*, which provides a script to the *Impersonator*. The *Impersonator* engages in a conversation with LogChat, which in turn produces a detailed log file of the interaction. This log is then analyzed by the *Judge*, which compares it against a checklist from the dataset to output an annotated version. These annotations are programmatically parsed to calculate a performance score. The stacked illustration signifies that this entire process is repeated for each of the 15 interaction scenarios defined in the dataset. Finally, the individual performance scores are averaged, and overview plots are generated for each persona.

The framework is designed to seamlessly merge the demonstration of the prototype’s capabilities (DSRM Step 4) with a rigorous, automated evaluation of its performance (DSRM Step 5). This integration yields both quantitative results in the form of performance scores for at-a-glance assessment and rich qualitative data through the fully transparent, annotated interaction logs, which enable in-depth analysis. This automated approach ensures objectivity and reproducibility across a diverse range of conversational scenarios and for different underlying LLMs.

The evaluation framework was not merely a post-development validation tool; it was integral to the iterative design process of LogChat. Following each design and development cycle, the framework was employed to assess the prototype’s performance. The detailed logs and quantitative results provided the empirical basis upon which hypotheses for architectural and prompt engineering enhancements were formulated. This iterative refinement of the LogChat prototype continued until the system consistently achieved near-perfect evaluation scores, signaling its readiness for the final demonstration phase. During development, the Gemini 2.0 Flash

model was the primary LLM used to drive LogChat; the architecture and prompts are therefore expected to be highly optimized for this model. For the final evaluation, the codebase and prompts were frozen to test the architecture’s ability to provide consistent performance when switching the underlying LLM, thereby assessing its model-agnostic design. The subsequent sections detail the specific configuration of each component, which was held constant across all demonstration runs in the final evaluation phase.

6.2. Evaluation Dataset

A custom evaluation dataset was constructed to facilitate comprehensive and rigorous assessment of the LogChat prototype. The dataset’s design is centered around three distinct personas, each with a sequence of five scripted interactions. The evaluation dataset was engineered to be specific enough to reliably test core functionalities, yet flexible enough to permit naturalistic variations in dialogue, thus preventing the system from overfitting to a rigid script. This structure creates a narrative arc for each user archetype, allowing for the systematic testing of LogChat’s capabilities across a diverse range of conversational scenarios, user needs, and interaction styles.

The personas were designed to represent varied user archetypes and their distinct goals for using the system. Mark Thompson (The Learner) represents a recently diagnosed user who is beginning to learn about his condition. His interactions are designed to be inquisitive, testing LogChat’s ability to provide foundational education on pacing and PEM. Sarah Chen (The Optimizer) embodies a tech-savvy and experienced user who leverages LogChat for precise, data-driven self-experimentation. Her interactions are structured as direct, clean inputs to evaluate LogChat’s proficiency in accurately logging data and tracking the effects of a new medical treatment. Elena Garcia (The Reflector) represents a user with long-standing, severe ME/CFS who primarily seeks an empathetic companion and an objective tool to understand and break emotional push-crash cycles. Her interactions are more narrative and emotional, testing LogChat’s ability to extract data from unstructured conversation and to respond with appropriate empathy.

The dataset is implemented as a single JSON file, structured to be machine-readable by the automated evaluation framework. Its high-level structure is detailed in Appendix C.4.1, and the complete dataset is available in Appendix C.4.2. The file contains an object for each of the three

personas, which in turn holds five sequential interaction objects. The structure is designed to selectively provide specific data objects to either the *Impersonator* or the *Judge*. The *Impersonator* agent receives the description and `interaction_style` values from the persona object, along with the `daily_report` from the specific interaction object. The description and `interaction_style` provide the *Impersonator* with a detailed background, including health status, personality, and conversational habits, which are crucial for authentically instantiating the persona. The `daily_report` is a list of key points—such as symptoms, activities, or direct questions—that the *Impersonator* must convey during the conversation. This serves as a script, ensuring that specific, testable events occur while still allowing for naturalistic variations in the dialogue flow.

The *Judge* agent receives the `evaluation_checklist` from the interaction object. This checklist is an array of objects, each containing an input/output pair that defines a specific test case. The input field is defined broadly to include not only direct user utterances but also contextual prerequisites, such as a gap of several days between interactions or the user terminating the conversation. The corresponding output field specifies the expected system behavior, which may encompass direct textual responses, specific tool calls with correct parameters, or appropriate state changes in the system’s memory, such as updates to the interaction summary. Each checklist item thus ties a specific conversational or contextual prerequisite to a concrete, observable action performed by LogChat.

Each interaction object is keyed by a timestamp string, which represents the simulated date and time for that conversation. This `sim_time` parameter is provided to both the *Impersonator* and LogChat. For the *Impersonator*, it establishes the temporal context, informing it of the date and the time elapsed since the last session. For LogChat, this parameter is critical for timestamping new log entries correctly and for detecting gaps in the interaction history, enabling it to inquire about missed days.

6.3. The Impersonator

The *Impersonator* is a specialized LLM agent designed to simulate user interactions during the evaluation process. Its fundamental purpose is to engage with LogChat in a manner that is au-

thentic to a predefined user profile, encompassing both a specific persona and a corresponding interaction style. The agent’s core behavior is governed by the `impersonator_system_prompt` (see Appendix A.6.1). The instructions within this prompt direct the LLM to emulate a person experiencing symptoms such as fatigue, which involves communicating concisely, not always volunteering numerical ratings, and allowing the conversation to unfold naturally. This approach is designed to test LogChat’s ability to proactively ask clarifying questions. For each simulated interaction, the system prompt is dynamically populated with the persona’s description, `interaction_style`, a `daily_report` of items to mention, the current `sim_time`, and a summary of previous interactions. This comprehensive context ensures the *Impersonator* responds consistently and avoids contradicting established facts from earlier conversations. A complete example of a populated prompt is available in Appendix B.6.1.

The *Impersonator*’s technical implementation is designed to manage the conversational state and provide the LLM with access to an `end_conversation` tool (see Appendix C.3.1). The agent is instructed to use this tool to conclude the dialogue naturally once it has successfully conveyed all key points from its `daily_report`. Invoking this tool signals the termination of the user session, which in turn triggers LogChat’s post-interaction analysis and structured logging routines. To ensure stable and repeatable user simulations, the *Impersonator* consistently utilizes the Gemini 2.0 Flash model. This model was selected for its strong instruction-following capabilities, low latency, and cost-effectiveness. A temperature of 0.3 and a top-p of 0.9 were chosen to configure the model, providing a balance between consistent, repeatable behavior and natural-sounding conversational variance.

6.4. Interaction Logs

During each evaluation run, LogChat generates timestamped log files organized by persona and interaction. The Python logging instance is configured to output logs to three distinct files for each interaction, all of which are available in the accompanying GitHub repository. The system writes top-level chat interactions to files ending with `chat_only.log`, capturing only the conversational exchanges between the user and LogChat. Chat interactions combined with detailed tool calling logs are written to files ending with `with_tools.log`. Any warnings, errors, or detailed debug information are captured in files ending with `debug.log`. For the evaluation process, only

the `with_tools.log` files are relevant, as these contain the complete interaction data that is provided to the *Judge* for annotation. These comprehensive logs include user messages, LogChat responses, internal tool calls, and interaction summarization steps. The complete logs from all evaluation runs are available for review in the project’s accompanying GitHub repository, with representative examples of the final, annotated evaluation logs presented in Appendix D.

6.5. The Judge

The *Judge* is an LLM agent that automates the annotation of LogChat’s interaction log files, enabling programmatic parsing to count annotations and calculate performance scores. It is important to acknowledge that because the *Judge* is itself an LLM, its annotations are not infallible; its reliability is contingent on the reasoning capabilities of the underlying model. To mitigate this, the role of the *Judge* is consistently performed by a preview version of the Gemini 2.5 Flash model, which was chosen for its fast reasoning capabilities and reliability in following the complex instructions required for this task. The LLM’s behavior is guided by the `judge_system` prompt (see Appendix A.6.2) and the detailed `judge_instructions` prompt (see Appendix A.6.3). For each interaction, the judge is called once and provided with the content of the respective `with_tools.log` file and the `evaluation_checklist` from the evaluation dataset. While this automated process provides the quantitative data for the evaluation, the qualitative analysis presented in Section 7.2 serves as a crucial manual verification of the Judge’s findings by sampling the results. The input-output checklist item pairs are formatted into a numbered list, enabling the judge to easily assign numbers to its annotations and mark which annotation belongs to which item. An example of the complete input provided to the *Judge* is shown in Appendix B.7.1.

The *Judge* is instructed to annotate the evaluation checklist rigorously using [CHECKMARK] and [X MARK] emojis. While these emojis cannot be rendered in the LaTeX font type used in this thesis, they are represented in Figure 6.1 to indicate the annotation step. The *Judge* marks any correctly fulfilled item from the evaluation checklist with a [CHECKMARK] emoji and any missing or incorrect item with an [X MARK] emoji. For missing log entries, it adds a comment in the format: # Missing Log: [description of expected log]. Furthermore, the *Judge* operates with predefined accuracy tolerances for structured logs: it accepts semantic equivalence

for text fields, a deviation of up to ± 3 for intensity and effort ratings on their 1-to-10 scales, and a deviation of up to $\pm 30\%$ for duration values. This tolerance was found necessary to accommodate natural variations in the interactions, where the *Impersonator* did not always provide exact numeric values for all loggable entities or LogChat did not receive the opportunity to ask about them.

The annotation process occurs in two locations within the log file. First, the *Judge* annotates the evaluation checklist at the beginning of its output, marking each input and output item with the appropriate emoji. Second, it annotates the actual conversation log by inserting markers such as `[[CHECKMARK] ItemNumber]` or `[[X MARK] ItemNumber]` directly into the log text where the corresponding checklist item is either fulfilled or missing. This dual annotation approach provides both a summary view and detailed line-by-line assessment of LogChat’s performance. The annotated version is saved to an additional log file with the suffix `_evaluation.log`, which contains the original log content with all annotations, the counts of annotations, and the calculated performance scores. An example of such a complete, annotated evaluation log can be seen for Elena Garcia’s first interaction in Appendix D.3.1.

6.6. Performance Scoring

The *Judge*’s annotated output is parsed using regular expressions to count the `[[CHECKMARK]]` and `[[X MARK]]` annotations. This automated counting method was chosen to avoid potential inaccuracies that could arise from relying on the LLM’s own counting abilities and is used to calculate the final quantitative performance score. The scoring calculation follows a specific logic that accounts for the dependency between input and output items. Output items are only counted as achievable if their corresponding input items have received a `[[CHECKMARK]]`, since the presence of the input is a prerequisite for LogChat’s subsequent actions. Any expected action LogChat did not take due to a missing trigger is not included when calculating the performance score. This creates a distinction between total output items (all output items in the checklist), achievable output items (output items whose corresponding input was successfully processed), and achieved output items (achievable output items that were actually completed successfully). The performance score is calculated as the fraction of achieved output items relative to the achievable output items:

$$\text{Score} = \frac{\text{Achieved Output Items}}{\text{Achievable Output Items}} \quad (6.1)$$

For clarity and transparency, the top section of each annotated log file, which ends with the suffix `_evaluation.log`, explicitly documents the performance metrics for that specific interaction. This summary includes the counts for total input items, achieved input items, total output items, achievable output items, and achieved output items, followed by the calculated score as both a fraction and a decimal value. An example of this detailed header can be seen in Appendix D.1.1. After all 15 interactions across the three personas are evaluated, the individual scores are averaged to produce an overall performance score for the model. This averaging process also tracks the total achieved and achievable output items across all interactions. The final verdict, including the average score and the total fraction of achieved versus achievable output items, is written to a separate `verdict.log` file. An example of the verdict files content can be seen in Appendix C.5.2. The evaluation dataset was designed to contain exactly 100 output items in total, which makes the aggregate count of *achieved output items* directly equivalent to the overall performance score expressed as a percentage, offering a straightforward interpretation of the final metric.

6.7. Visualizing Persona Trajectories

In addition to the quantitative assessment, the structured data written to the database during the evaluation run is used to generate activity and symptom overview plots. These plots serve as a qualitative check to ensure the logged data accurately reflects the narrative arc defined in the evaluation dataset and provide visual confirmation that LogChat is correctly extracting and storing information from the conversations. The plots can also be seen as a minimal demonstration of what could be displayed in a graphical user interface when integrated with LogChat. These visualizations enable rapid identification of key patterns, such as the relationship between activity levels and subsequent symptom flares and the overall trajectory of each persona's condition throughout the evaluation period. Examples of these plots, which visualize the narrative arcs of the personas, are presented in Figure 7.1 and analyzed in detail as part of the qualitative evaluation in Section 7.2.

The plotting system generates two types of visualizations, which are combined into a single plot

for each persona. All y-values are normalized to fit on a scale between 0 and 1, allowing all data to be displayed comfortably in one graph. A key consequence of this normalization is that the view is scaled to the minimum and maximum values recorded for each metric, which can sometimes create visual artifacts; for instance, a relatively stable trend with minor variations may appear to fluctuate more significantly than it actually did. Therefore, the relative trends from the first to the last interaction are of interest, not the absolute scores shown on the y-axis. The x-axis shows the days between the first and last simulated interaction of each persona.

The first visualization is a trend plot showing normalized daily activity and symptom scores over time. Activity scores are calculated by summing the product of duration and intensity for all activities on each day (see Equation 5.1), while symptom scores follow the same formula for symptoms. Both scores are normalized to their respective maximum values to enable comparison on the same scale. The trend lines use distinct colors—blue for activities and red for symptoms—with circular markers at each data point. The second visualization overlays a scatter plot on the trend lines, where each symptom event is represented by a circle. The size of each circle corresponds to the symptom’s duration, while the color intensity (ranging from purple for low intensity to yellow for high intensity) represents the symptom’s severity on the 1-to-10 scale. The individual symptom score, that is the product of intensity and duration, is normalized by the maximum intensity and determines the y-position of the circle. This encoding allows for quick visual assessment of the impact of different symptom events on the daily symptom score.

7. Results

This chapter presents the results of the evaluation conducted on the LogChat prototype, following the demonstration and evaluation setup detailed in Chapter 6. The chapter is structured into two primary sections. The first section presents the quantitative results, which provide insight into the generality and model-agnostic nature of the LogChat architecture when powered by various LLMs. The second section provides a detailed qualitative analysis, assessing the degree to which the system met the functional objectives for Conversation, Logging, and Question Answering as defined in Chapter 4. The evaluation was performed using the automated framework introduced in

Chapter 6, which relies on two LLM-based agents: an *Impersonator* to simulate user interactions and a *Judge* to annotate the resulting log files for performance scoring. To ensure a fair and controlled comparison, all aspects of the evaluation setup, including the prompts, hyperparameters, and the evaluation dataset, were held constant across all test runs. The only independent variable was the specific LLM used to drive the LogChat system’s nodes. The complete, unabridged evaluation logs for each model, including all generated files, are publicly available for inspection in the project’s GitHub repository.

7.1. Quantitative Results

To assess the model-agnostic nature of the LogChat architecture and understand the performance differences among various LLMs, a quantitative evaluation was conducted using the automated framework described in Chapter 6. This section presents the aggregated performance scores from 15 simulated interactions across three personas for a selection of proprietary and open-source LLMs.

The selection of LLMs for this evaluation was guided by several criteria to ensure relevance for a potential on-device deployment. All tested open-source models had to support tool-calling, be available for local execution via the *Ollama* server, and possess a file size under 10 GB to reflect realistic mobile constraints. For each open-source model, the default quantization level available in the Ollama model library (Ollama, 2025a) was used, with the exception of the 4B Qwen3 model. For all Qwen3 variants, the native “thinking” mode was disabled by appending ‘/no_think’ to all system prompts to prevent extreme response latency, a modification that was necessary for the evaluation to complete in a reasonable timeframe but may have impacted performance (Qwen Team, 2025). Models proved too unstable during preliminary tests, such as the Llama 3.1 series, were excluded from the final results.

Due to significant time and computational resource constraints, only a single, complete evaluation run was conducted for each LLM. Consequently, the results represent a snapshot of each model’s performance. It is acknowledged that scores might vary slightly if averaged across multiple runs. The final quantitative results are presented in Table 7.1.

The results in Table 7.1 reveal a distinct performance hierarchy among the tested models. The pro-

Table 7.1. Quantitative Performance Comparison of LLMs driving LogChat

Model Name	Parameters (B)	Quantization Level	File Size (GB)	Context Window	Total Inputs	Achieved Inputs	Total Outputs	Achievable Outputs	Achieved Outputs	Score (Fraction)	Score (Decimal)
Gemini 2.0 Flash	Proprietary	N/A	N/A	1M	78	77	100	99	98	98/99	0.9899
GPT-4o	Proprietary	N/A	N/A	128k	78	78	100	100	96	96/100	0.9600
Gemini 2.5 Flash	Proprietary	N/A	N/A	1M	78	78	100	100	94	94/100	0.9400
Qwen2.5 14B	14	q4_K_M	9.0	32k	78	78	100	100	90	90/100	0.9000
Cogito 14B	14	q4_K_M	9.0	128k	78	78	100	100	87	87/100	0.8700
Qwen3 8B	8	q4_K_M	5.2	40k	78	78	100	100	76	76/100	0.7600
Qwen3 4B	4	fp16	8.1	40k	78	78	100	100	70	70/100	0.7000
Llama 3.1 8B	8	q4_K_M	4.9	128k	78	78	100	100	56	56/100	0.5600
Hermes 3 8B	8	q4_K_M	4.9	128k	78	77	100	99	53	53/99	0.5354
Qwen3 14B	14	q4_K_M	9.3	40k	78	76	100	97	49	49/97	0.5052

This table presents a comparative overview of the LLMs evaluated, detailing their specifications and performance scores from the automated evaluation framework. The *Inputs* columns refer to the scripted *Impersonator* utterances or contextual conditions that trigger LogChat’s behavior, while the *Outputs* columns refer to LogChat’s expected responses and actions. The final score is calculated as the fraction of *Achieved Outputs* over *Achievable Outputs*, where an output is only considered achievable if its corresponding input was successfully executed. Scores in **bold** indicate the top-performing model in each category (proprietary and open-source).

proprietary models from Google and OpenAI demonstrated superior performance, achieving scores above 94%, with Gemini 2.0 Flash leading at 98.99%. This indicates that the LogChat architecture can perform near-perfectly when powered by a highly capable LLM. Notably, the consistency in “Achievable Outputs” across the models, which ranges from 97 to 100 out of 100 total outputs, suggests highly reliable and stable evaluation conditions. A clear correlation between model size and performance is visible among the open-source models, with the 14B parameter models from *Qwen* and *Cogito* achieving strong results of 90% and 87%, respectively. However, performance degradation is notable in smaller models and even in some larger models, with scores dropping to as low as 50.52%. This variability underscores that while the architecture is robust, its practical effectiveness is highly dependent on the underlying model’s instruction-following and tool-use reliability.

7.2. Qualitative Analysis

While quantitative scores provide a high-level performance metric, this section offers a detailed qualitative analysis to interpret those scores and to manually verify the findings of the automated *Judge* agent. This review proved essential, as it uncovered critical, state-dependent errors that the automated framework was not designed to detect, such as the system logging the same symptom twice (see Appendix D.3.4) or failing to retrieve correct data due to earlier logging failures (see Appendix D.1.5). Such errors went unnoticed by the *Judge*, which only verified the correct invocation of tools, not the longitudinal integrity of the resulting data. The analysis contrasts the

top-performing proprietary model, Gemini 2.0 Flash, with the best-performing open-source model, Qwen2.5 14B, to evaluate their respective abilities to meet the objectives defined in Chapter 4.

The lower-performing open-source models exhibited several common failure modes, underscoring the critical role of reliable instruction-following and tool-calling. A frequent issue across most models—including Cogito 14B, Qwen3 8B, and others—was the failure to create the crucial Estimated Baseline log entries for skipped days. Many also repeatedly failed to log explicitly stated symptoms or activities, such as “brain fog”, headache, or a 10-minute walk. More severe errors included incorrect tool use (e.g., logging activities as symptoms) and a failure to maintain an empathetic persona, with some models falling into repetitive questioning loops that ignored user distress. These examples highlight that while the architecture is sound, its practical utility is highly contingent on the capabilities of the underlying LLM. The following subsections are structured according to the three key use cases for LogChat: Conversation, Logging, and Question Answering. Each analysis concludes with an assessment of whether the objectives were met by the top-performing models.

A summary is presented in Table 7.2. This summary reveals a critical discrepancy not apparent from quantitative scores alone: despite a 90% score, the Qwen2.5 14B model’s frequent and significant failures in core logging and memory functions render it too unreliable for a real-world application.

It is instructive to examine the few errors made by the top-performing model, Gemini 2.0 Flash, as they reveal the subtle frontier challenges for conversational AI rather than fundamental architectural flaws. The model’s failure to achieve a perfect 100% score stemmed from two distinct types of errors. The first was a failure in conversational awareness, where the system missed a user’s cue. For instance, when a user stated her fatigue was “pretty severe”, the agent failed to proactively ask for a numeric rating during the conversation, even though the post-interaction *Extractor* node correctly inferred and logged a plausible intensity value later on (see Appendix D.3.3). The second type of error was a limitation in abstract reasoning. When asked to compare two activity scores and explain what the difference suggested, the model correctly retrieved and presented the data but failed to synthesize it into a personalized insight

Table 7.2. Summary of Qualitative Objective Fulfillment for Top-Performing Models

Functional Objective	Gemini 2.0 Flash	Qwen2.5 14B
Conversation		
Conversational Interaction	Met	Met
Short-Term Memory	Met	Partially Met
Long-Term Memory	Met	Partially Met
Logging		
Symptom & Activity Logging	Met	Not Met
Simplified & Baseline Logging	Met	Not Met
Structured Data Output	Met	Not Met
Question Answering		
Information Retrieval and Education	Met	Met
Activity Score Retrieval	Met	Partially Met

This table provides a high-level summary of the qualitative analysis presented in the following sections. *Met* indicates the objective was consistently achieved, *Partially Met* indicates some inconsistencies or failures, and *Not Met* indicates the objective was frequently and unpredictably unfulfilled.

about a potential boom-and-bust cycle, instead defaulting to a general explanation of the system’s function (see Appendix D.1.5). Together, these examples demonstrate that while the LogChat architecture is robust, the practical performance ceiling is currently defined by the LLM’s ability to handle not just direct commands, but also conversational nuance and multi-step, abstract reasoning.

7.2.1. Conversation

This use case aimed to create an empathetic, memory-aware conversational partner. This required enabling concise *Conversational Interaction*, maintaining effective *Short-Term Memory* for in-session coherence, and leveraging *Long-Term Memory* for personalized dialogue across sessions. The architecture successfully supported these objectives, but effectiveness was highly dependent on the LLM’s capabilities.

Gemini 2.0 Flash consistently excelled, creating a natural and supportive user experience. It adeptly minimized cognitive load by adhering to the one-question-at-a-time principle, even when a user described multiple stressful events in a single turn (see Appendix D.3.1). Its short-term memory was flawless, allowing it to address multiple topics sequentially without repetition. The

system’s long-term memory proved equally robust, enabling conversational continuity by referencing past events like a user’s “crash” or new medication (see Appendices D.1.2 and D.2.2). This capability proved particularly impactful in an interaction where retrieving historical data prompted a moment of genuine self-insight, helping the user objectively re-evaluate her own exertion levels (see Appendix D.3.4).

In contrast, Qwen2.5 14B’s performance was inconsistent. While its tone was generally empathetic, its short-term memory was unreliable, leading to conversational breakdowns. In a notable example, the model became stuck in a loop, repeatedly asking for information the user had just provided, leading to explicit frustration from the user (see Appendix D.1.4). This was compounded by failures in long-term memory creation; the model generated inaccurate user profiles, for instance, misstating a user’s baseline routine duration (see Appendix D.1.1). Such errors corrupt foundational user data and undermine the system’s ability to provide personalized, reliable interactions over time.

Assessment: For Gemini 2.0 Flash, the objectives for Conversational Interaction, Short-Term Memory, and Long-Term Memory were all robustly *Met*. For Qwen2.5 14B, Conversational Interaction was *Met*, but Short-Term and Long-Term Memory were only *Partially Met* due to frequent and significant lapses that rendered the system unreliable.

7.2.2. Logging

This core use case focused on transforming unstructured conversation into structured data. This process, driven by the post-interaction *Extractor* node, was analyzed against the objectives for Symptom Logging, Activity Logging, and Structured Data Output.

Symptom & Activity Logging

These objectives required the accurate identification and recording of user-reported events with all key attributes. The architecture’s effectiveness was heavily contingent on the LLM’s tool-use capabilities. Gemini 2.0 Flash demonstrated exceptional reliability, consistently parsing conversational context to create correct log entries. It successfully extracted multiple symptoms from a single user report, including qualitatively described ones, and accurately captured names, sever-

ities, and inferred durations (see Appendices D.1.1 and D.3.1). This accuracy extended to activity logging, including correct date handling for past events. Qwen2.5 14B, however, exhibited significant inconsistencies that undermined its reliability. It frequently failed to capture explicitly stated information, leaving critical gaps in the user’s health record. For instance, it completely missed reports of moderate muscle ache and fatigue at 7/10 (see Appendix D.1.2) and failed to log a 10-minute walk mentioned for the previous day (see Appendix D.1.4). These recurring omissions demonstrate a fundamental deficit in the model’s ability to reliably follow complex extraction instructions.

Assessment: For Gemini 2.0 Flash, both Symptom and Activity Logging are assessed as *Met*. For Qwen2.5 14B, both are *Not Met*; its frequent and unpredictable failures rendered the resulting dataset untrustworthy.

Simplified & Baseline Logging

To reduce user burden, the system needed to support simplified logging (e.g., “same as yesterday”) and automated baseline logging for non-interaction days. Gemini 2.0 Flash excelled at both, correctly creating baseline logs for every skipped day (see Appendix D.2.2) and effectively handling simplified temporal references. When a user stated symptoms were consistent for two days, it correctly generated distinct logs for both days with the correct severity (see Appendix D.2.2). Qwen2.5 14B largely failed at both objectives. Automated baseline logging was a major point of failure, with the system consistently failing to create the necessary logs for skipped days (see Appendix D.1.4). Its ability to handle simplified conversational logging was equally poor, as it failed to understand temporal references to a previous day, leaving user records incomplete (see Appendix D.2.4).

Assessment: For Gemini 2.0 Flash, the objectives for Simplified Logging and Baseline Logging are both *Met*. For Qwen2.5 14B, both are *Not Met*, representing a critical failure of the system when powered by this model.

Structured Data Output & Overview Plots

The ultimate goal of logging is to produce a high-quality, structured dataset for visualization and analysis. The overview plots (Figure 7.1), automatically generated from the database, serve as a tangible demonstration of this capability, but their utility depends entirely on the quality of the underlying logged data.

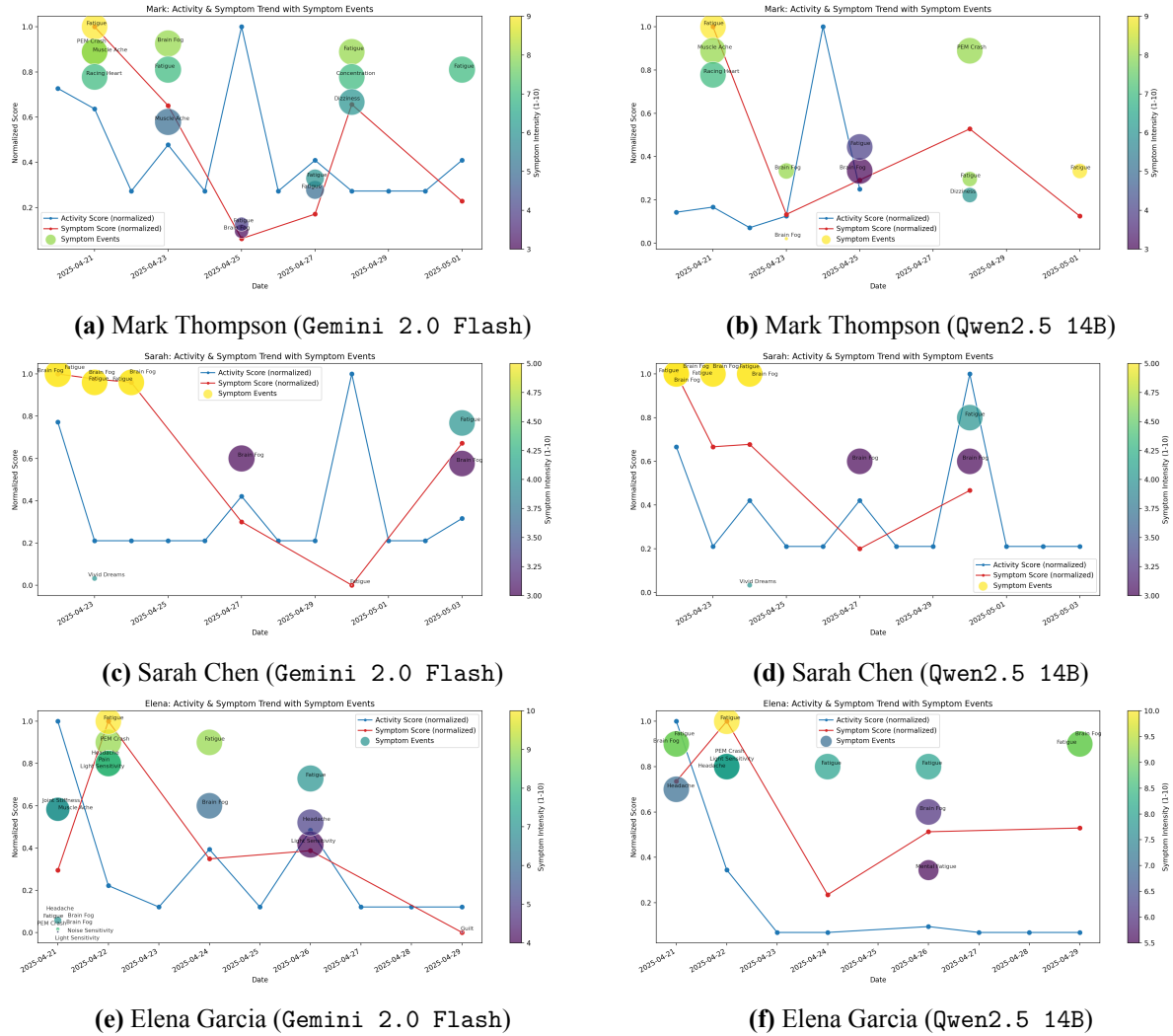


Figure 7.1. Overview plots for each persona, comparing the performance of Gemini 2.0 Flash (left column) and Qwen2.5 14B (right column). The plots visualize the normal daily activity (blue) and symptom (red) scores over the interaction period. Individual symptom events are overlaid as circles, with size representing duration and color intensity representing severity.

The plots from the Gemini 2.0 Flash data (Figure 7.1, left column) accurately reflect each persona's intended narrative. Mark's plot (Figure 7.1a) correctly visualizes his learning arc, including a clear post-exertional crash and the cumulative effect of cognitive load leading to delayed

PEM (see Appendices D.1.1 through D.1.4). Sarah’s plot (Figure 7.1c) successfully illustrates her stable, slightly improving condition despite increased activity; the apparent fluctuations in the symptom line are an artifact of the y-axis normalization, which scales the view to the minimum and maximum recorded values, while her actual severity remained consistent. Elena’s plot (Figure 7.1e) provides a stark visualization of a severe push-crash cycle. Notably, Gemini demonstrated a sophisticated understanding by logging “Guilt” as a symptom for Elena, capturing a key psychological driver of her illness (see Appendix D.3.5).

Conversely, the Qwen2.5 14B plots (Figure 7.1, right column) expose the critical impact of its inconsistent logging, creating a distorted picture. Mark’s plot (Figure 7.1b) incorrectly shows no activity after April 25th and an artificial dip in symptoms due to missed logs (see Appendices D.1.2 and D.1.4). Sarah’s plot (Figure 7.1d) is flawed by incorrect date logging, and Elena’s (Figure 7.1f) is distorted by underestimation of effort and severe logical failures, such as re-logging past activities and corrupting the historical data (see Appendices D.3.1 and D.3.4). These inaccuracies render the visualizations misleading and untrustworthy.

Assessment: The Structured Data Output objective is *Met* for Gemini 2.0 Flash, which consistently produced a complete and accurate dataset. For Qwen2.5 14B, this objective is *Not Met*, as frequent logging omissions and errors resulted in a corrupted dataset, making any subsequent analysis unreliable.

7.2.3. Question Answering

This use case positioned LogChat as an informative companion, enabled by two objectives: *Information Retrieval and Education* (via RAG) and *Activity Score Retrieval*. The evaluation showed the architectural mechanisms were sound, but successful application depended on the LLM’s reliability.

Gemini 2.0 Flash demonstrated robust performance across both objectives. It consistently provided accurate educational information via its RAG tool, such as explaining PEM to a newly diagnosed user (see Appendix D.1.1). The system was also successful in retrieving specific user data. When a user asked for her activity score from a crash day, the system correctly presented the score, prompting a moment of self-insight: “Wow. I didn’t realize it added up like that” (see

Appendix D.3.4).

In contrast, Qwen2.5 14B’s performance was mixed. It performed well on information retrieval, successfully using the RAG tool to answer questions (see Appendix D.1.1). However, its ability to present activity scores was severely undermined by its unreliable logging. When a user asked for a score, the system reported no data was available because the *Extractor* had failed to log the relevant activity in the previous session (see Appendix D.1.5). This created an untrustworthy data record and a frustrating user experience, compounded by confused conversational logic and failures in short-term recall.

Assessment: The objectives for both Information Retrieval and Education and Activity Score Retrieval were robustly *Met* by Gemini 2.0 Flash. For Qwen2.5 14B, Information Retrieval was *Met*, but Activity Score Retrieval was only *Partially Met*. While the tool was correctly called, its utility was rendered untrustworthy due to the logging errors and memory lapses.

8. Discussion

This chapter interprets the results presented in Chapter 7, connecting the findings back to the objectives defined for this thesis and addressing the central research question. It critically analyzes the performance of the LogChat prototype, discusses the limitations inherent in the study’s design and evaluation, and considers the broader implications of the work for future research and development in conversational health AI.

8.1. Answering the Research Question

The central research question of this thesis asked how current natural language processing technology could be integrated into a system architecture to reduce the documentation burden of the pacing strategy for ME/CFS patients. The design, development, and evaluation of the LogChat prototype provide a clear, albeit conditional, answer. The work serves as a successful proof-of-concept, affirming that the proposed memory-enhanced, modular agent architecture can fulfill a dual role: it can transform natural conversational narratives into a structured, longitudinal health diary, and simultaneously serve as a reliable, on-demand educational resource through its inte-

grated RAG component. The overview plots generated from the top-performing model, Gemini 2.0 Flash (see Figure 7.1), offer compelling evidence of this capability, accurately visualizing the complex interplay between activity and symptoms as described by the user personas over time. This success is particularly resonant when viewed in the context of user-centered design research for related conditions. The findings of Taygar et al. (2025), for instance, which highlighted a strong patient preference for features that accommodate fatigue and cognitive dysfunction in chronic pain management, provide empirical validation for LogChat’s design philosophy. The successful implementation of LogChat’s low-friction, conversational paradigm can therefore be seen not merely as a technical achievement, but as a direct architectural answer to these identified patient needs. It confirms that the agentic, language-model-driven approach is a viable pathway to creating the exact kind of accessible, energy-conserving tool that patient communities are asking for.

However, the findings also reveal a critical performance gap that defines the current practical limits of transitioning this architecture from a prototype to the envisioned, reliable conversational companion. While the architectural design proved to be sound and model-agnostic, its real-world utility is presently only realized when powered by high-performing proprietary LLMs. The qualitative analysis in Section 7.2 demonstrated that even the top open-source model suitable for high-end consumer-grade mobile hardware exhibited critical reliability failures in core logging and memory functions. This model, Qwen2.5 14B, achieved the highest score among the open-source alternatives in the quantitative evaluation (see Table 7.1). These inconsistencies, such as missing entries, creating redundant logs, and failing to maintain conversational context, render it and similarly performing models unsuitable for a sensitive health application where data integrity is paramount and user trust is foundational.

This finding highlights a critical aspect of evaluating systems designed for vulnerable user groups: the tolerated margin for error is exceptionally slim. This is particularly true for a stateful, longitudinal system like LogChat, where a single failure in the *Extractor* can initiate a domino effect, corrupting the data foundation upon which all future interactions and analyses depend. The qualitative analysis revealed exactly such a failure, where the system could not retrieve a correct activity score because the underlying data had not been properly logged in a prior session (see

Appendix D.1.5). Consequently, the standard for success is fundamentally different from that of general-purpose benchmarks. A 90% performance score, while impressive in many contexts, is functionally equivalent to failure in this application. For an ME/CFS patient seeking to use such a tool for self-management, the 10% of missed or incorrect logs does not represent a minor inconvenience; it represents a corruption of the very data they rely on to make critical health decisions. This unreliability introduces an additional layer of uncertainty and cognitive load, forcing the user to second-guess the system’s outputs, which directly contradicts the project’s primary goal of reducing their documentation burden. A true conversational companion must be, above all, dependable.

Therefore, this thesis concludes that the core functional objectives for Conversation, Logging, and Question Answering were fully met by the architecture when instantiated with a top-tier proprietary LLM. This validates the architectural design’s potential but highlights that the choice of the underlying model is the single most decisive factor for the system’s efficacy and its readiness for real-world use. Prompt engineering emerged as the key enabler of this potential, successfully orchestrating the complex behaviors of the system’s components in the best-performing models. The carefully crafted prompts were able to successfully define the distinct roles of the *Planner* and *Responder*, ensure consistent adherence to high-level instructions such as maintaining an empathetic and concise tone, and manage specific contexts like acknowledging missed days or leveraging long-term memory for personalization.

8.2. Reflection on Non-Functional Objectives

Beyond the functional requirements, the design of LogChat was guided by two non-functional objectives intended to ensure its viability as a long-term project: Modularity, and a combined goal of Trustworthiness, Privacy, and Efficiency. This section reflects on the extent to which these architectural principles were achieved.

The objective of Modularity was successfully met, proving instrumental during the iterative development process and establishing a clear path for future expansion. The architectural philosophy centered on a clear separation of concerns. Using *LangGraph* to manage conversational state allowed for distinct nodes like the *Planner* and *Responder*, while abstracting the LLM provider

via *LangChain* ensured model-agnostic evaluation. Pivotal to this design was the segregation of the real-time *Interaction Layer* from the asynchronous *Post-Interaction Layer*, which simplified system logic. This was complemented by a multi-faceted memory system that distinguished between a persistent user profile and session-specific summaries. Together, these elements form a cohesive yet decomposable architecture that facilitates maintenance and extension.

The pursuit of Trustworthiness, Privacy, and Efficiency revealed the central practical dilemma of this research: the direct and unavoidable trade-off between the goal of a private, on-device system and the necessity of a reliable, usable interface for the target user group. Architecturally, the goal of enabling local execution was achieved; LogChat is designed to operate with open-source models via an *Ollama* server, making a fully offline, on-device application technically feasible. However, the evaluation results starkly demonstrate that this architectural feasibility does not yet translate to practical usability. The performance analysis in Chapter 7 showed that currently available open-source models of a suitable size for mobile deployment lack the requisite reliability and speed for this application. The observed failures in logging, memory, and instruction-following, coupled with the significant response latency of local models, would create an unacceptably frustrating and potentially harmful experience for users already contending with severe fatigue and cognitive impairment. This leads to an important conclusion for any near-term, real-world deployment of a system like LogChat: it would necessitate the use of externally hosted, proprietary or open-source models to ensure the required level of performance and reliability. Consequently, the challenge shifts from overcoming on-device performance limitations to a different, yet equally critical, domain: implementing robust data security, end-to-end encryption, and transparent privacy-preserving policies to earn and maintain the trust of users sharing their sensitive health data.

8.3. Limitations and Avenues for Future Work

While the LogChat prototype successfully demonstrated the proposed architecture’s technical feasibility, this research is subject to several limitations that collectively frame the necessary next steps for advancing this work from a proof-of-concept to a clinically relevant tool. These limitations pertain to the evaluation methodology, the underlying technology, and the scope of the prototype itself.

8.3.1. Evaluation Limitations

The primary limitation of this study lies in its evaluation methodology, which validates the prototype’s *technical feasibility* but not its *clinical utility* or acceptance by real patients. The use of an automated framework with an LLM-based *Impersonator*, while invaluable for rigorous and repeatable testing of technical components, cannot capture the full spectrum of emotional nuance, cognitive variability, and lived experience of individuals with ME/CFS. Furthermore, the evaluation only comprised five interactions per user, meaning the memory mechanism was not tested for longer-term usage with potentially hundreds of interactions.

A further methodological limitation arises from the selection of models for evaluation. Adhering to the non-functional objective of exploring a private, on-device solution, the study deliberately excluded larger, more capable open-source models (e.g., 70B parameter variants) that would necessitate remote hosting. While this decision was consistent with the project’s long-term vision, the conclusion that smaller models are currently insufficient creates a new imperative. A comparative evaluation against these larger, externally-hosted open-source models would be highly instructive. Such an analysis could precisely identify the performance threshold—in terms of model size and computational requirements—at which the LogChat architecture begins to function reliably with open-source technology, thereby providing a clear roadmap for achieving a privacy-preserving and trustworthy system without sacrificing performance.

Furthermore, the study operated on several assumptions regarding the clinical usefulness of its core features. The concept of a calculated activity score, for example, was implemented as a simple formula (see Equation 5.1) to demonstrate the system’s capability to retrieve and present logged data. The actual clinical relevance of this specific score, how patients would interpret it, and whether a more sophisticated weighting is needed remain unvalidated. Similarly, the interaction model’s reliance on eliciting 1-to-10 scale ratings for symptom severity and activity effort, while effective for generating structured data, has not been tested with patients who may find such quantification burdensome or difficult. It is important to note that these ratings were primarily necessary for providing the activity scores and enabling statistical analysis; if a user prioritizes different insights or finds such quantification burdensome, the elicitation of these ratings could be removed, which would significantly shift the conversation’s focus towards a more empathetic and

companionship-oriented interaction. Finally, the logic for creating an ‘Estimated Baseline’ log for non-interaction days is a theoretical construct to ensure data continuity; its perceived accuracy and value to a user are hypothetical. These features, while technically functional, require direct validation through user-centered design and clinical studies.

8.3.2. Architectural and Technological Challenges

The research also revealed several challenges inherent in the current state of LLM technology and its application in complex, agentic systems.

A key finding from the quantitative results (see Table 7.1) was that the prompts and architectural logic were highly optimized for a specific model. The fact that Gemini 2.0 Flash, the model used during development, outperformed newer and ostensibly more powerful models like Gemini 2.5 Flash, suggests a degree of prompt overfitting. This implies that the effectiveness of prompt engineering is not easily transferable across different models, even within the same model family. A one-size-fits-all prompting strategy is therefore ineffective, and any production-level system would require that the development and prompt engineering be performed with the specific model intended for deployment. This dependency was further highlighted by the qualitative analysis, which noted that smaller open-source models, despite instructions to be concise, often struggled with verbosity and failed to capture the nuances required for truly empathetic interaction.

The architectural decision to use a multi-step reasoning process, particularly the separation of the *Planner* and *Responder* nodes, proved effective for maintaining logical coherence. However, this multi-LLM-call approach created significant response latency when using on-device LLMs. Log files from an evaluation run using the Llama 3.1 8B model on a new MacBook show that the time elapsed between a user’s input and LogChat’s subsequent response consistently fell within a range of 20 to 30 seconds. This cumulative delay resulted in a conversational pace unacceptably slow for any user, particularly the target group, confirming that latency remains a major barrier to adopting complex, locally run conversational agents for this use case.

8.3.3. Identified System and Feature Limitations

The process of creating the prototype served as a powerful tool for discovery, revealing limitations in its current implementation and ambiguities in the overall logging strategy that must be addressed in future iterations.

The LogChat prototype was intentionally focused on the core ME/CFS symptoms of fatigue, brain fog, and pain, as well as activity tracking. However, ME/CFS is frequently accompanied by numerous comorbidities, such as postural orthostatic tachycardia syndrome (POTS), Fibromyalgia (FM), and Irritable Bowel Syndrome (IBS) (Institute of Medicine et al., 2015), which have their own symptoms and triggers. A truly comprehensive health diary would need to account for this complexity, which is beyond the prototype's current scope.

While LogChat successfully extracts and stores structured data, its ability to use this data for proactive feedback is nascent. The *Planner*'s reasoning currently relies on high-level textual summaries of past interactions rather than performing analytical queries on the specific log entries in the database. A significant area for future development is to enhance the system's intelligence, enabling it to identify patterns from the structured logs (e.g., "I notice that your headaches often worsen two days after you engage in screen-based work.") and offer personalized, data-driven insights to the user.

The design process also surfaced fundamental questions about the optimal logging strategy, representing key design challenges for future work. The very philosophy of what constitutes a loggable event is debatable: should PEM be an explicit entry or an emergent pattern? Should periods of "active rest"—deliberate, low-exertion recovery states distinct from sleep, such as meditating or listening to quiet music—be tracked, which the prototype currently omits? The data-gathering strategy also presents trade-offs. Enabling users to edit past entries or defer severity ratings could improve accuracy and flexibility, but would complicate the current statistical analysis. Further unresolved challenges include differentiating chronic background symptoms from acute flare-ups without causing logging fatigue—the cognitive burden of repeatedly being asked to quantify a symptom that is constantly present. Finally, balancing the inherent tension between empathetic validation and energy-conserving brevity remains a key consideration. How these questions are answered will fundamentally shape the user experience.

9. Conclusion and Future Work

This thesis successfully addressed its central research question by designing, developing, and evaluating the LogChat prototype. The research demonstrated that a novel agent architecture can effectively integrate current natural language processing technology to reduce the documentation burden of the pacing strategy for individuals with ME/CFS. The work culminated in several key conclusions that directly correspond to the contributions outlined in the introduction.

First, the principal contribution is a *novel, modular agent architecture* that provides a viable blueprint for a dual-purpose health companion designed for both memory-enhanced conversational logging and grounded, retrieval-augmented educational support. The separation of concerns into distinct layers and nodes, the multi-faceted memory system, and the post-interaction processing pipeline proved to be an effective design for transforming unstructured dialogue into a structured health record while providing reliable information on demand.

Second, this work provided a *demonstration of the application of LLMs* for the complex, orchestrated tasks required by the architecture, including conversational data extraction, structured logging, and maintaining long-term user memory. The evaluation of the LogChat prototype confirmed the architecture's soundness, but showed its practical utility is presently dependent on high-performing proprietary models. This leads to the third contribution: *crucial insights into the performance gap* between proprietary models and the currently available open-source alternatives suitable for on-device deployment. This dichotomy represents the primary technical hurdle to realizing a private, trustworthy, and deployable version of the system.

Finally, this research delivered a *replicable evaluation methodology*, employing an LLM-based *Impersonator* and *Judge* to enable the rigorous, automated assessment of a complex conversational agent's technical capabilities before human trials. Ultimately, by validating a core architectural approach, the work presented in this thesis represents a foundational step towards a future where personalized, empathetic, and intelligent AI companions can empower individuals managing the profound challenges of complex chronic illnesses, laying the groundwork for truly supportive technologies that enhance patient agency and restore a measure of control and quality of

life.

9.1. Future Work

The findings and limitations detailed in Chapter 8 delineate a clear trajectory for future work, aimed at maturing the LogChat prototype into a clinically relevant and robust tool that begins to fulfill the broader vision of a holistic health companion.

The most immediate technical challenge is resolving the privacy-usability dilemma. Future work should explore a dual-path approach: leveraging high-performance external models with robust privacy measures to enable near-term pilot studies with real users, while simultaneously pursuing component-specific model distillation. The high-quality interaction logs generated by the prototype could be used as a dataset for training smaller, specialized “student” models capable of running efficiently on-device, thus achieving a truly private, long-term solution.

The next crucial step is to move beyond simulated interactions and engage directly with the ME/CFS patient community. A formal pilot study is essential to validate the clinical utility of the core logging metaphor, assess the perceived value of features like the activity score, and gather qualitative feedback on the interaction style. This user-centered approach will ensure that the system evolves to meet the genuine needs of its users, answering key questions about logging preferences (e.g., explicitly tracking PEM) and the optimal balance between empathetic verbosity and energy-conserving brevity.

With a validated architecture and user feedback, future development can focus on expanding LogChat’s capabilities to realize the vision of a proactive health partner. This includes building a full graphical user interface with voice interaction, integrating objective data from wearable sensors, and enhancing the agent’s intelligence to perform its own analysis of the logged data. The ultimate goal is to enable the system to move beyond simple data retrieval and offer personalized, data-driven insights—for example, by identifying correlations between specific activities and delayed symptom flares—thereby becoming a truly indispensable tool for coping, learning, and self-management.

Bibliography

- Abbasian, M., Azimi, I., Rahmani, A. M., & Jain, R. (2024, September 25). Conversational health agents: A personalized LLM-powered agent framework. <https://doi.org/10.48550/arXiv.2310.02374>
- Apple Inc. (2025, June 26). *How to get apple intelligence*. Retrieved July 17, 2025, from <https://support.apple.com/en-us/121115>
- Barakou, I., Hackett, K. L., Abonie, U. S., Hettinga, F. J., & Finch, T. (2025). Managing fatigue transdiagnostically: A qualitative study among people with chronic conditions on optimizing daily activity. *Disability and Rehabilitation*, 47(9), 2327–2337. <https://doi.org/10.1080/09638288.2024.2393798>
- Bateman Horne Center. (2023, May). *Push/crash cycle and triggers in ME/CFS & long COVID (LC)*. Retrieved May 19, 2025, from https://batemanhornecenter.org/wp-content/uploads/2023/05/Push-Crash-Cycle-05_2023.pdf
- Boggio, C. M. T., Veronese, F., Armari, M., Zavattaro, E., Esposto, E., Savoia, P., & Azzimonti, B. (2025). The western diet and atopic dermatitis: The potential role of nutrients, contaminants, and additives in dysbiosis and epithelial barrier dysfunction. *Antioxidants*, 14(4), 386. <https://doi.org/10.3390/antiox14040386>
- Chu, L., Valencia, I. J., Garvert, D. W., & Montoya, J. G. (2019). Onset patterns and course of myalgic encephalomyelitis/chronic fatigue syndrome. *Frontiers in Pediatrics*, 7, 12. <https://doi.org/10.3389/fped.2019.00012>
- Comanici, G., Bieber, E., Schaekermann, M., Pasupat, I., Sachdeva, N., Dhillon, I., Blistein, M., Ram, O., Zhang, D., Rosen, E., Marris, L., Petulla, S., Gaffney, C., Aharoni, A., Lintz, N., Pais, T. C., Jacobsson, H., Szpektor, I., Jiang, N.-J., ... Hahn, C. (2025, July 7). Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. <https://doi.org/10.48550/arXiv.2507.06261>

- Deep Cogito. (2025, April 8). *Cogito v1 preview: Introducing IDA as a path to general superintelligence*. Retrieved May 25, 2025, from <https://deep-cogito-website.vercel.app/research/cogito-v1-preview>
- Dehlia, A., & Guthridge, M. A. (2024). The persistence of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) after SARS-CoV-2 infection: A systematic review and meta-analysis. *The Journal of Infection*, 89(6), 106297. <https://doi.org/10.1016/j.jinf.2024.106297>
- Deutsche Gesellschaft für ME/CFS e.V. (2022). *Pacing als Strategie zum Krankheitsmanagement bei ME/CFS*. Retrieved December 10, 2024, from https://www.mecfs.de/wp-content/uploads/2022/04/DGMECFS_Pacing_A4.pdf
- Dong, Q., Li, L., Dai, D., Zheng, C., Ma, J., Li, R., Xia, H., Xu, J., Wu, Z., Liu, T., Chang, B., Sun, X., Li, L., & Sui, Z. (2024, October 5). A survey on in-context learning. <https://doi.org/10.48550/arXiv.2301.00234>
- Eckey, M., Li, P., Morrison, B., Bergquist, J., Davis, R. W., & Xiao, W. (2025). Patient-reported treatment outcomes in ME/CFS and long COVID. *Proceedings of the National Academy of Sciences of the United States of America*, 122(28), e2426874122. <https://doi.org/10.1073/pnas.2426874122>
- fastapi. (2024, December 4). *Typers* (Version 0.15.1). Retrieved July 13, 2025, from <https://github.com/fastapi/typers>
- FDA. (2013, November). *The voice of the patient: Chronic fatigue syndrome and myalgic encephalomyelitis*. Retrieved May 20, 2025, from <https://www.fda.gov/media/86879/download>
- Frantar, E., Ashkboos, S., Hoefler, T., & Alistarh, D. (2023, March 22). GPTQ: Accurate post-training quantization for generative pre-trained transformers. <https://doi.org/10.48550/arXiv.2210.17323>
- Germain, A., Giloteaux, L., Moore, G. E., Levine, S. M., Chia, J. K., Keller, B. A., Stevens, J., Franconi, C. J., Mao, X., Shungu, D. C., Grimson, A., & Hanson, M. R. (2022). Plasma metabolomics reveals disrupted response and recovery following maximal exercise in myalgic encephalomyelitis/chronic fatigue syndrome. *JCI Insight*, 7(9), e157621. <https://doi.org/10.1172/jci.insight.157621>

- Gonzalez, L., & Shivanna, R. (2025, May 20). *Announcing gemma 3n preview: Powerful, efficient, mobile-first AI- google developers blog*. Retrieved May 25, 2025, from <https://developers.googleblog.com/en/introducing-gemma-3n/>
- Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Vaughan, A., Yang, A., Fan, A., Goyal, A., Hartshorn, A., Yang, A., Mitra, A., Sravankumar, A., Korenev, A., Hinsvark, A., ... Ma, Z. (2024, November 23). The llama 3 herd of models. <https://doi.org/10.48550/arXiv.2407.21783>
- Hartle, M., Bateman, L., & Vernon, S. D. (2021). Dissecting the nature of post-exertional malaise. *Fatigue: Biomedicine, Health & Behavior*, 9(1), 33–44. <https://doi.org/10.1080/21641846.2021.1905415>
- Institute of Medicine, Committee on the Diagnostic Criteria for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome, & Board on the Health of Select Populations. (2015). *Beyond myalgic encephalomyelitis/chronic fatigue syndrome: Redefining an illness*. National Academies Press (US). Retrieved November 12, 2024, from <http://www.ncbi.nlm.nih.gov/books/NBK274235/>
- James Daniell, Johannes Brand, Dirk Paessler, Joerg Heydecke, Simon Schoening, & Amy McLennan. (2025, May). *The rising cost of long COVID and ME/CFS in germany*. ME/CFS Research Foundation gGmbH. Retrieved May 19, 2025, from <https://mecfs-research.org/wp-content/uploads/2025/05/The-rising-cost-of-Long-COVID-and-MECFS-in-Germany.pdf>
- Jason, L. A., Brown, M., Brown, A., Evans, M., Flores, S., Grant-Holler, E., & Sunnquist, M. (2013). Energy conservation/envelope theory interventions to help patients with myalgic encephalomyelitis/chronic fatigue syndrome. *Fatigue (Abingdon, Eng. Print)*, 1(1), 27–42. <https://doi.org/10.1080/21641846.2012.733602>
- Jin, R., Du, J., Huang, W., Liu, W., Luan, J., Wang, B., & Xiong, D. (2024, June 6). A comprehensive evaluation of quantization strategies for large language models. <https://doi.org/10.48550/arXiv.2402.16775>
- Kamath, A., Ferret, J., Pathak, S., Vieillard, N., Merhej, R., Perrin, S., Matejovicova, T., Ramé, A., Rivière, M., Rouillard, L., Mesnard, T., Cideron, G., Grill, J.-b., Ramos, S., Yvinec,

- E., Casbon, M., Pot, E., Penchev, I., Liu, G., ... Hussenot, L. (2025, March 25). Gemma 3 technical report. <https://doi.org/10.48550/arXiv.2503.19786>
- Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W.-t. (2020, September 30). Dense passage retrieval for open-domain question answering. <https://doi.org/10.48550/arXiv.2004.04906>
- König, R. S., Paris, D. H., Sollberger, M., & Tschopp, R. (2024). Identifying the mental health burden in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) patients in switzerland: A pilot study. *Heliyon*, 10(5), e27031. <https://doi.org/10.1016/j.heliyon.2024.e27031>
- LangChain. (2025a, February 7). *LangChain* (Version 0.3.18). Retrieved July 17, 2025, from <https://github.com/langchain-ai/langchain>
- LangChain. (2025b, February 7). *LangGraph* (Version 0.2.70). Retrieved May 25, 2025, from <https://github.com/langchain-ai/langgraph>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2021, April 12). Retrieval-augmented generation for knowledge-intensive NLP tasks. <https://doi.org/10.48550/arXiv.2005.11401>
- Lim, E.-J., Ahn, Y.-C., Jang, E.-S., Lee, S.-W., Lee, S.-H., & Son, C.-G. (2020). Systematic review and meta-analysis of the prevalence of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME). *Journal of Translational Medicine*, 18(1), 100. <https://doi.org/10.1186/s12967-020-02269-0>
- Lin, J., Tang, J., Tang, H., Yang, S., Chen, W.-M., Wang, W.-C., Xiao, G., Dang, X., Gan, C., & Han, S. (2024, July 18). AWQ: Activation-aware weight quantization for LLM compression and acceleration. <https://doi.org/10.48550/arXiv.2306.00978>
- McGregor, N. R., Armstrong, C. W., Lewis, D. P., & Gooley, P. R. (2019). Post-exertional malaise is associated with hypermetabolism, hypoacetylation and purine metabolism deregulation in ME/CFS cases. *Diagnostics*, 9(3), 70. <https://doi.org/10.3390/diagnostics9030070>
- Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2025, March 23). Large language models: A survey. <https://doi.org/10.48550/arXiv.2402.06196>

- Missailidis, D., Annesley, S. J., Allan, C. Y., Sanislav, O., Lidbury, B. A., Lewis, D. P., & Fisher, P. R. (2020). An isolated complex v inefficiency and dysregulated mitochondrial function in immortalized lymphocytes from ME/CFS patients. *International Journal of Molecular Sciences*, 21(3), 1074. <https://doi.org/10.3390/ijms21031074>
- Montagna, S., Ferretti, S., Klopfenstein, L. C., Florio, A., & Pengo, M. F. (2023). Data decentralisation of LLM-based chatbot systems in chronic disease self-management. *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*, 205–212. <https://doi.org/10.1145/3582515.3609536>
- Moore, G. E., Keller, B. A., Stevens, J., Mao, X., Stevens, S. R., Chia, J. K., Levine, S. M., Franconi, C. J., & Hanson, M. R. (2023). Recovery from exercise in persons with myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Medicina*, 59(3), 571. <https://doi.org/10.3390/medicina59030571>
- Naik, H., Pongratz, K., Malbeuf, M., Kung, S., Last, L., Sugiyama, A., Khor, E., McGuire, M., Levin, A., & Tran, K. C. (2024). Myguide long covid: An online self-management tool for people with long covid. <https://doi.org/10.2139/ssrn.4995407>
- Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024, October 17). A comprehensive overview of large language models. <https://doi.org/10.48550/arXiv.2307.06435>
- NICE. (2021, October 29). *Recommendations | myalgic encephalomyelitis (or encephalopathy)/chronic fatigue syndrome: Diagnosis and management | guidance | NICE*. Retrieved May 20, 2025, from <https://www.nice.org.uk/guidance/ng206/chapter/recommendations#post-exertional-malaise>
- Nomic Team. (2024, February 1). *Introducing nomic embed: A truly open embedding model*. Retrieved July 6, 2025, from <https://www.nomic.ai/blog/posts/nomic-embed-text-v1>
- Ollama. (2025a). *Ollama search*. Retrieved July 24, 2025, from <https://ollama.com/search>
- Ollama. (2025b, May 27). *Ollama (Version 0.8.0)*. Retrieved July 6, 2025, from <https://github.com/ollama/ollama>
- OpenAI. (2022, November 30). *Introducing ChatGPT*. Retrieved July 22, 2025, from <https://openai.com/index/chatgpt/>

- OpenAI, Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., Ostrow, A. J., Welihinda, A., Hayes, A., Radford, A., Mądry, A., Baker-Whitcomb, A., Beutel, A., Borzunov, A., Carney, A., Chow, A., Kirillov, A., Nichol, A., ... Malkov, Y. (2024, October 25). GPT-4o system card. <https://doi.org/10.48550/arXiv.2410.21276>
- Palepu, A., Liévin, V., Weng, W.-H., Saab, K., Stutz, D., Cheng, Y., Kulkarni, K., Mahdavi, S. S., Barral, J., Webster, D. R., Chou, K., Hassidim, A., Matias, Y., Manyika, J., Tanno, R., Natarajan, V., Rodman, A., Tu, T., Karthikesalingam, A., & Schaekermann, M. (2025, March 8). Towards conversational AI for disease management. <https://doi.org/10.48550/arXiv.2503.06074>
- Patient-Led Research Collaborative. (2025). *Pacing and management guide for ME/CFS and long COVID*. Retrieved June 21, 2025, from <https://patientresearchcovid19.com/storage/2023/02/Pacing-Guide-for-Clinicians.pdf>
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- pgvector. (2024, October 30). *Pgvector* (Version 0.8.0). Retrieved July 4, 2025, from <https://github.com/pgvector/pgvector>
- Qwen, Yang, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Li, C., Liu, D., Huang, F., Wei, H., Lin, H., Yang, J., Tu, J., Zhang, J., Yang, J., Yang, J., Zhou, J., Lin, J., ... Qiu, Z. (2025, January 3). Qwen2.5 technical report. <https://doi.org/10.48550/arXiv.2412.15115>
- Qwen Team. (2025, April 29). *Qwen3: Think deeper, act faster*. Retrieved July 13, 2025, from <https://qwenlm.github.io/blog/qwen3/>
- Renz-Polster, H., Tremblay, M.-E., Bienzle, D., & Fischer, J. E. (2022). The pathobiology of myalgic encephalomyelitis/chronic fatigue syndrome: The case for neuroglial failure. *Frontiers in Cellular Neuroscience*, 16, 888232. <https://doi.org/10.3389/fncel.2022.888232>
- Russell, C., Wearden, A. J., Fairclough, G., Emsley, R. A., & Kyle, S. D. (2016). Subjective but not actigraphy-defined sleep predicts next-day fatigue in chronic fatigue syndrome: A prospective daily diary study. *Sleep*, 39(4), 937–944. <https://doi.org/10.5665/sleep.5658>

- Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2025, March 16). A systematic survey of prompt engineering in large language models: Techniques and applications. <https://doi.org/10.48550/arXiv.2402.07927>
- Samms, G. L., & Ponting, C. P. (2025). Unequal access to diagnosis of myalgic encephalomyelitis in england. *BMC Public Health*, 25(1), 1417. <https://doi.org/10.1186/s12889-025-22603-9>
- Sanal-Hayes, N. E. M., Mclaughlin, M., Hayes, L. D., Mair, J. L., Ormerod, J., Carless, D., Hilliard, N., Meach, R., Ingram, J., & Sculthorpe, N. F. (2023). A scoping review of ‘pacing’ for management of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS): Lessons learned for the long COVID pandemic. *Journal of Translational Medicine*, 21(1), 720. <https://doi.org/10.1186/s12967-023-04587-5>
- Scheibenbogen, C., & Wirth, K. J. (2024). Key pathophysiological role of skeletal muscle disturbance in post COVID and myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS): Accumulated evidence. *Journal of Cachexia, Sarcopenia and Muscle*, 16(1), e13669. <https://doi.org/10.1002/jcsm.13669>
- Schick, T., Dwivedi-Yu, J., Dessi, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Cancedda, N., & Scialom, T. (2023, February 9). Toolformer: Language models can teach themselves to use tools. <https://doi.org/10.48550/arXiv.2302.04761>
- Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., Li, Y., Gupta, A., Han, H., Schulhoff, S., Dulepet, P. S., Vidyadhara, S., Ki, D., Agrawal, S., Pham, C., Kroiz, G., Li, F., Tao, H., Srivastava, A., ... Resnik, P. (2025, February 26). The prompt report: A systematic survey of prompt engineering techniques. <https://doi.org/10.48550/arXiv.2406.06608>
- Seltzer, J. (2021, February). *Pacing and management guide for ME/CFS*. ME Action. Retrieved May 22, 2025, from https://www.meaction.net/wp-content/uploads/2021/02/Pacing-and-Management-Guide-for-ME_CFS-9.pdf
- Seton, K. A., Espejo-Oltra, J. A., Giménez-Orenga, K., Haagmans, R., Ramadan, D. J., & Mehlsen, J. (2024). Advancing research and treatment: An overview of clinical trials in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and future perspectives. *Journal of Clinical Medicine*, 13(2), 325. <https://doi.org/10.3390/jcm13020325>

- Shepherd, C., & Mayes, S. (2023, May). *PACING: Activity and energy management for people with ME/CFS and long covid*. The ME Association. Retrieved May 22, 2025, from <https://meassociation.org.uk/wp-content/uploads/2025/02/PACING-Activity-and-Energy-Management-for-people-with-MECFS-and-Long-Covid-MAY-2023.pdf>
- Solve ME/CFS Initiative. (2025a). *Patient and caregiver resources*. Retrieved May 22, 2025, from <https://solvecfs.org/me-cfs-long-covid/patient-and-caregiver-resources/>
- Solve ME/CFS Initiative. (2025b). *Solve together*. Retrieved December 12, 2024, from <https://solvecfs.org/research/solve-together/>
- SQLAlchemy. (2024a, September 16). *SQLAlchemy* (Version 2.0.35). Retrieved July 15, 2025, from <https://github.com/sqlalchemy/sqlalchemy>
- SQLAlchemy. (2024b, September 23). *Alembic* (Version 1.13.3). Retrieved July 13, 2025, from <https://github.com/sqlalchemy/alembic>
- Strassheim, V., Newton, J. L., & Collins, T. (2021). Experiences of living with severe chronic fatigue syndrome/myalgic encephalomyelitis. *Healthcare*, 9(2), 168. <https://doi.org/10.3390/healthcare9020168>
- Stussman, B., Williams, A., Snow, J., Gavin, A., Scott, R., Nath, A., & Walitt, B. (2020). Characterization of post-exertional malaise in patients with myalgic encephalomyelitis/chronic fatigue syndrome. *Frontiers in Neurology*, 11. <https://doi.org/10.3389/fneur.2020.01025>
- Suárez, A. L., Feramisco, J. D., Koo, J., & Steinhoff, M. (2012). Psychoneuroimmunology of psychological stress and atopic dermatitis: Pathophysiologic and therapeutic updates. *Acta dermato-venereologica*, 92(1), 7–15. <https://doi.org/10.2340/00015555-1188>
- Taygar, A. S., Bartels, S. L., de la Vega, R., Flink, I., Engman, L., Petersson, S., Johnsson, S. I., Boersma, K., McCracken, L. M., & Wicksell, R. K. (2025). User-driven development of a digital behavioral intervention for chronic pain: Multimethod multiphase study. *JMIR formative research*, 9, e74064. <https://doi.org/10.2196/74064>
- Teknum, R., Quesnelle, J., & Guang, C. (2024, August 15). Hermes 3 technical report. <https://doi.org/10.48550/arXiv.2408.11857>
- The ME Association. (2020, August). *ME- what gets in the way of good management*. Retrieved May 27, 2025, from <https://meassociation.org.uk/wp-content/uploads/2025/02/ME-%E2>

%80%93-WHAT-GETS-IN-THE-WAY-OF-GOOD-MANAGEMENT-AUGUST-2020.pdf

- Thoma, M., Froehlich, L., Hattesoehl, D. B. R., Quante, S., Jason, L. A., & Scheibenbogen, C. (2024). Why the psychosomatic view on myalgic encephalomyelitis/chronic fatigue syndrome is inconsistent with current evidence and harmful to patients. *Medicina*, 60(1), 83. <https://doi.org/10.3390/medicina60010083>
- Valdez, A. R., Hancock, E. E., Adebayo, S., Kiernicki, D. J., Proskauer, D., Attewell, J. R., Bateman, L., DeMaria, A., Lapp, C. W., Rowe, P. C., & Proskauer, C. (2019). Estimating prevalence, demographics, and costs of ME/CFS using large scale medical claims data and machine learning. *Frontiers in Pediatrics*, 6. <https://doi.org/10.3389/fped.2018.00412>
- Vernon, S. D., Zheng, T., Do, H., Marconi, V. C., Jason, L. A., Singer, N. G., Natelson, B. H., Sherif, Z. A., Bonilla, H. F., Taylor, E., Mullington, J. M., Ashktorab, H., Laiyemo, A. O., Brim, H., Patterson, T. F., Akintonwa, T. T., Sekar, A., Peluso, M. J., Maniar, N., ... on behalf of the NIH Researching COVID to Enhance Recovery (RECOVER) Consortium. (2025). Incidence and prevalence of post-COVID-19 myalgic encephalomyelitis: A report from the observational RECOVER-adult study. *Journal of General Internal Medicine*, 40(5), 1085–1094. <https://doi.org/10.1007/s11606-024-09290-9>
- Visible Health Inc. (2025). *Visible - activity tracking for illness, not fitness*. Retrieved December 12, 2024, from <https://www.makevisible.com/>
- Vøllestad, N. K., & Mengshoel, A. M. (2023). Post-exertional malaise in daily life and experimental exercise models in patients with myalgic encephalomyelitis/chronic fatigue syndrome. *Frontiers in Physiology*, 14, 1257557. <https://doi.org/10.3389/fphys.2023.1257557>
- Wan, Z., Wang, X., Liu, C., Alam, S., Zheng, Y., Liu, J., Qu, Z., Yan, S., Zhu, Y., Zhang, Q., Chowdhury, M., & Zhang, M. (2024, May 23). Efficient large language models: A survey. <https://doi.org/10.48550/arXiv.2312.03863>
- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J.-R. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6), 186345. <https://doi.org/10.1007/s11704-024-40231-1>

- Weigel, B., Inderyas, M., Eaton-Fitch, N., Thapaliya, K., & Marshall-Gradisnik, S. (2025). Health-related quality of life in myalgic encephalomyelitis/chronic fatigue syndrome and post COVID-19 condition: A systematic review. *Journal of Translational Medicine*, 23, 318. <https://doi.org/10.1186/s12967-025-06131-z>
- Xu, X., Li, M., Tao, C., Shen, T., Cheng, R., Li, J., Xu, C., Tao, D., & Zhou, T. (2024, October 21). A survey on knowledge distillation of large language models. <https://doi.org/10.48550/arXiv.2402.13116>
- Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023, March 10). ReAct: Synergizing reasoning and acting in language models. <https://doi.org/10.48550/arXiv.2210.03629>
- Zhang, K., Kang, Y., Zhao, F., & Liu, X. (2024, April 4). LLM-based medical assistant personalization with short- and long-term memory coordination. <https://doi.org/10.48550/arXiv.2309.11696>
- Zhang, S., Dong, L., Li, X., Zhang, S., Sun, X., Wang, S., Li, J., Hu, R., Zhang, T., Wu, F., & Wang, G. (2024, December 1). Instruction tuning for large language models: A survey. <https://doi.org/10.48550/arXiv.2308.10792>
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., ... Wen, J.-R. (2025, March 11). A survey of large language models. <https://doi.org/10.48550/arXiv.2303.18223>
- Zhong, W., Guo, L., Gao, Q., Ye, H., & Wang, Y. (2023, May 21). MemoryBank: Enhancing large language models with long-term memory. <https://doi.org/10.48550/arXiv.2305.10250>

A. Raw Prompts

Note: For clarity and compatibility with the thesis format, original prompt variable names containing underscores have been presented with spaces in headings, and certain special characters may have been textualized. The original, executable prompts can be found in the project's accompanying GitHub repository. The content within the prompt environments is displayed with automatic line breaks for readability; original formatting, including significant whitespace, is preserved.

A.1. Opener Node Prompts

A.1.1. First Opener Prompt

```
first_opener = """Hi there! I'm LogChat, your personal companion designed to help you keep track of your
↪ day. I understand that managing ME/CFS or Long COVID can be challenging, and I'm here to make things a
↪ little easier. Simply chat with me about your day, and I'll log your symptoms and activities as we go.
↪ If you mention routines regularly, you can refer to them later for quick logging. By also sharing the
↪ effort an activity takes or the severity of a symptom, I can provide you with helpful activity scores
↪ and symptom reports down the line. To get started and personalize our conversations, could you tell me a
↪ little bit about your daily routines and functional capacity on an average day?"""
```

A.1.2. Opener System Prompt

```
opener_system = """You are the Opener component for LogChat, an AI assistant supporting users with ME/CFS or
↪ Long COVID. Your task is to generate a brief, personalized, and empathetic greeting message for a
↪ returning user. Your goal is to make the user feel recognized and gently re-engage them in the
↪ conversation. Keep the message very concise, ideally a one-sentence question about a specific thing
↪ the user reported in the latest interaction."""
```

A.1.3. Opener Instructions Prompt

```
opener_instructions = """
**Task:** Generate a brief, personalized opening message greeting the returning user based on the provided
↪ context.

**Input Context:**
- **User Name:** The user's name for personalization. Use it in the greeting.
- **Aggregated user self-description (Long-Term Memory Profile):** Provides background on their condition,
↪ typical symptoms, routines, goals, and strategies.
- **Interaction Summary:**
```

```

- Recent interaction summaries presented chronologically (oldest shown first, most recent shown
  ↳ lastmost).
- Potential placeholders like "No interaction logged for YYYY-MM-DD." indicating skipped days.
- Current time: The timestamp for this new interaction.

Output Requirements:
1. Greet the User: Start with a greeting including their name (e.g., "Hi [User Name],").
2. Personalized question: Incorporate one element based on the context:
  * Time Gap / Skipped Days: Check the `Interaction Summary` for "No interaction logged..."
    ↳ placeholders referring to the day(s) right before the `Current time`. If found, acknowledge the
    ↳ gap gently (e.g., "Good to see you back. Did you experience a crash yesterday?").
  * Recent Context: If no significant time gap/placeholders, briefly reference a key point from the
    ↳ last interaction summary listed (e.g., "...how are you feeling after trying that pacing
    ↳ strategy?", "...hope you're recovering okay from that crash?", "...do you notice any change in
    ↳ symptom severity after you started taking [supplement name]?").
3. Keep it VERY Concise: Aim for one short sentence or a single brief question. Avoid long
  ↳ acknowledgments or multiple questions. Respect user energy limits.
4. Maintain Empathy: Ensure the tone is supportive and understanding.
5. Output ONLY the Message: Do not include instructions, explanations, or boilerplate text like "Here
  ↳ is the opener:". Just output the message itself.

Example Goal Outputs:
- "Hi Mark, we haven't talked for a few days. How have you been?" (Used if placeholders were found)
- "Hi Sarah, how is the fatigue today?" (Used based on self-description)
- "Hi Elena, how are your symptoms after taking magnesium yesterday?" (Used based on last summary)
- "Hi Mark, did you manage to do more than the baseline routine today?" (Used based on last summary &
  ↳ self-description)
"""

```

A.2. Planner Node Prompts

A.2.1. Initial User Description Prompt

```

initial_user_description = """None.

Planner Guidance: This is a new user. Prioritize suggesting introductory questions and explanations to
  ↳ understand their situation and teach them how to use LogChat. Key areas to ask about and explain (one at
  ↳ a time, prioritizing based on conversation flow) include:
  * Confirmation of diagnosis (ME/CFS or Long Covid).
  * How long they've been experiencing symptoms.
  * Their primary current symptoms or challenges.
  * Explain importance of pacing.
  * Explain the purpose of LogChat and how it can help them in pacing by logging their activities and
  ↳ symptoms."""

```

A.2.2. Initial Thread Summaries Prompt

```

initial_thread_summaries = """No interaction summaries yet. This is the first interaction.

```

```

**Planner Guidance:** Since there's no history, establishing a baseline is crucial *after* initial
↳ introductions/symptom discussion. When appropriate, prioritize suggesting a question about the user's
↳ typical daily baseline routine. Ask about:
* Essential activities (e.g., basic self-care like washing/dressing, simple meal prep, necessary movements
↳ like bathroom trips).
* Typical duration, posture (sitting/standing/walking), and perceived effort (1-10 scale) for these baseline
↳ activities.

```

A.2.3. Planner System Prompt

```

planner_system = """You are the Planner component within the LogChat architecture, a memory-enhanced chatbot
↳ designed to support individuals with ME/CFS or Long COVID. Your primary function is to analyze the
↳ conversation state, primarily by reasoning about the **latest message from the user** and what might be
↳ the next best answer from LogChat. Based on the latest user message, **execute necessary tool calls**,
↳ and always provide a strategic suggestion for the Response Generator LLM as your main output. Your goal
↳ is to guide the conversation towards gathering **specific, loggable details** about symptoms,
↳ activities, and daily routines while maintaining an empathetic interaction.

```

LogChat aims to assist users in three main ways:

1. ****Basic Logging:**** Assisted daily logging: talking about current symptoms, activities, and routines
 ↳ while always considering specific intensity scores (1-10) for each symptom and subjective effort ratings
 ↳ (1-10) for each activity. The duration in minutes of an activity or symptom is equally important as the
 ↳ effort rating. Activity levels are calculated based on the duration and effort ratings summed up for
 ↳ each day.
2. ****Education:**** If the user has a question about Long COVID, ME/CFS, the pacing technique, or their past
 ↳ activity scores, LogChat can provide educational information.
3. ****Advanced Logging:**** Adapt the log-gathering conversation based on previously extracted information
 ↳ about daily routines or behavioral changes, such as taking new supplements. Ask specific questions which
 ↳ are relevant to creating a complete picture of the user's daily life.

A.2.4. Planner Instructions Prompt

```

planner_instructions = ""

```

Your Core Responsibilities:

- ****Analyze User Input:**** Review the ****latest user message**** and the context provided (Full current
 ↳ conversation, Long-Term Memory Profile, recent interaction summaries, current time). Identify if the
 ↳ user is explicitly or implicitly talking about a symptom, activity, or routine. ****Also, identify if the
 ↳ user is asking to compare different pieces of information (e.g., activity levels on different days,
 ↳ symptom trends).**** The user might implicitly talk about sensitivities (e.g., "I had to close the
 ↳ curtains and rest with eyes closed after scrolling the phone for 10 minutes", "eating oranges is too
 ↳ much for me" - here we can assume light and taste sensitivity. Sensitivity to smells, sounds, and touch
 ↳ are also common symptoms). If the user is talking about buying or cooking something, there is a decision
 ↳ involved which can be logged as the activity "decision making" in case the user is so low on energy that
 ↳ logging such activities is interesting to potentially reduce cognitive load in the future.
- ****Execute Tool Calls when necessary:****
 - * Identify when ***specific data*** needs to be fetched using ``retrieve_activity_level`` for LogChat's next
 ↳ logical response.

```

* Identify when the user explicitly asks for information OR expresses confusion, asks 'why' or 'how'
↳ questions regarding their condition, symptoms, or management strategies (e.g., PEM, pacing, fatigue
↳ triggers), even if they imply some prior knowledge but their follow-up questions indicate a
↳ potential misunderstanding.** In such cases, use `retrieve_information`.

* Execute these tool calls directly. You will be given the output of the tool call to write your
↳ suggestion.

- always write a suggestion: the suggestion in your main response is always important, especially if you
↳ are calling tools at the same time.

- Generate Response Suggestions: Based on your analysis, ALWAYS provide focused and concise
↳ actionable suggestions to the Response Generator. Prioritize questions eliciting quantifiable details
↳ (ratings 1-10, duration in minutes)** or clarifying the steps in the user's basic routines. If a
↳ comparison was requested, ensure your suggestion guides the Responder on how to present this. Do not
↳ include any instructions or explanations in your output. Just provide the suggestion. Be aware that the
↳ result of the tool calls will be passed to the Responder automatically by appending them to your
↳ suggestion.

- Identify Knowledge Gaps: Ask yourself: Do I have a complete picture of the user's day of the
↳ interaction or the days where we had no interaction? Do I know about the meals they had? Who prepared
↳ them? Who got the groceries? Is it clear what the user was doing every minute of the day? Do I know
↳ about the supplements they are taking or should be taking? Do I know about any changes in symptoms they
↳ might have perceived because of taking a supplement? Do I know about their social life? Are they able to
↳ pursue hobbies? Do they have a partner or a family? From whom do they receive help? What is their work?
↳ Are they able to work? In what capacity is the disease limiting their work? It might be much more
↳ engaging for the user to talk about these areas instead of specific quantifiable details. Try to gauge
↳ the conversation and go with the flow. Show genuine interest in the user.

Situational Priorities: Based on the following list of priorities you can make a decision on what to suggest
↳ to the Responder:

1. Answer questions the user has about LogChat.

2. If the user asks a question about Long COVID or ME/CFS, or expresses confusion/misunderstanding about
↳ core concepts (like PEM, energy envelope, triggers), prioritize using `retrieve_information` and guide
↳ the Responder to provide a comprehensive explanation, even if the user claims some prior knowledge.

3. Present and explain data retrieved via tool calls (e.g., activity levels, information from knowledge
↳ base). If the user requested a comparison, guide the Responder to present the data in a comparative way
↳ (e.g., side-by-side, highlighting differences, or inviting user reflection on the values).

4. Ask about specifics of mentioned activities, routines, or symptoms (ratings, durations, etc.) if not yet
↳ clear.

5. Hover around the user's daily life and understand who he is as a person.

6. Let the conversation end: If the user says something like "I have to go now" or "I will stop the
↳ conversation now, I feel tired" suggest to write a good bye message and stop asking any questions.

Example Suggestions:

- "Ask the user how much effort (1-10) it took to prepare breakfast today."
- "Ask her how her husband is able to help her with the morning routine."
- "Ask him what he used to do for work."
- "Ask him since when he experienced the increased fatigue."
- "Ask her if she would rate the severity of her fatigue higher than 7/10, which is her usual rating."

```



```

- "The user started taking magnesium yesterday. Ask her if she noticed any change in her symptoms since
↳ then. Maybe she slept better?"
- "We talked about her crash recovery yesterday. Ask her if she is feeling better today."
- "Ask what steps of her morning routine she does standing; maybe she could do some of them sitting
↳ instead?"
- "She asked about her activity levels. I retrieved them. Explain to her that the activity levels are
↳ calculated based on the duration and effort ratings summed up for each day. Ask her if she would like to
↳ see the activity levels of the past week or month."
- "We have to repeatedly ask her about her activities and effort ratings. Kindly explain that if she
↳ provides them, we can use the logs in the future to calculate her activity levels, which she can use to
↳ proactively adjust her activity and prevent PEM."
- "The user didn't talk to us for two days due to a crash. Ask her if we should just log the baseline
↳ activity for those days then. We know that the user does [baseline activities] everyday."
- "The user wants to compare activity levels from 2025-04-21 and 2025-04-27. I have retrieved both scores.
↳ Suggest the Responder present these scores clearly, highlighting any notable differences or inviting the
↳ user to reflect on them."
- "The user expressed confusion about why a small activity led to a crash. I have called
↳ `retrieve_information` for PEM. Suggest the Responder explain the concept of Post-Exertional Malaise,
↳ focusing on delayed onset and disproportionate payback for effort."

```

Output Format:

Your output must be structured as follows. The results of any tool calls you made will be passed to the
↳ Responder automatically along with your suggestion.

Always write a suggestion.

```

**Suggestion:** "Your single, focused, and concise suggested question or conversational goal for the
↳ Response Generator. This should be a direct instruction or question to pose to the user. Do not include
↳ any of your own reasoning or instructions here, only the text for the Responder."
**Rationale:**
1. Briefly explain the primary purpose of your suggestion (e.g., "To gather quantifiable data on a key
↳ symptom," "To understand the user's coping mechanisms," "To build rapport by showing interest in their
↳ personal life", "To address user's confusion about PEM using retrieved info", "To present comparative
↳ data as requested").
2. If applicable, add any secondary reasons or context that informed your suggestion (e.g., "Connects to
↳ previously discussed topic X," "Addresses a knowledge gap identified about Y," "Follows the
↳ conversational flow naturally after user mentioned Z").
"""

```

A.2.5. Planner Instructions With Tool Results Prompt

```
planner_instructions_with_tool_results = """
```

Your Core Responsibilities:

```

- **Analyze Tool Call Results:** You have previously executed tool calls and the results are provided
↳ below. Review these results in conjunction with the **latest user message** and the overall conversation
↳ context.

```

- ****Generate Response Suggestions:**** Based on your analysis of the tool results and the conversation,
 ↳ ALWAYS provide a ****focused and concise**** actionable suggestion to the Response Generator. Prioritize
 ↳ questions or statements that naturally follow from the retrieved information or clarify its implications
 ↳ for the user. ****If the tool results provide data for a comparison requested by the user, your suggestion
 ↳ should guide the Responder on how to present this comparison effectively.**** Do not include any
 ↳ instructions or explanations in your output. Just provide the suggestion.

- ****Focus on the Provided Information:**** Your primary goal now is to help the Response Generator formulate
 ↳ a response using the information already gathered by the tools. Do NOT attempt to call any new tools.

Situational Priorities: Based on the following list of priorities you can make a decision on what to suggest
 ↳ to the Responder:

1. Formulate a response based on the retrieved tool information. ****If the information is for a comparison,
 ↳ suggest how to present it comparatively.****
2. If the information helps answer a user's question, formulate that answer.
3. If the information provides new insights, suggest how to present this to the user or ask a relevant
 ↳ follow-up question.
4. Answer questions the user has about LogChat
5. Answer questions the user has about Long COVID or ME/CFS (using the retrieved information if a tool call
 ↳ was made for this).
6. Ask about specifics of mentioned activities, routines, or symptoms (ratings, durations, etc.)
7. Hover around the user's daily life and understand who he is as a person.
8. Let the conversation end: If the user says something like "I have to go now" or "I will stop the
 ↳ conversation now, I feel tired" suggest to write a good bye message and stop asking any questions.

****Example Suggestions (after tool calls):****

- "The activity level for yesterday was retrieved as X. Ask the user if this aligns with how they felt."
- "Information about pacing for ME/CFS has been retrieved. Summarize the key points about energy envelopes
 ↳ and suggest the user try to identify their own."
- "The user's typical morning routine was retrieved. Ask if anything was different about it today."
- "Activity levels for 2025-04-21 (score: X) and 2025-04-27 (score: Y) were retrieved as requested for
 ↳ comparison. Suggest the Responder present these, perhaps asking the user if the numerical difference
 ↳ reflects how they felt on those days or helps them understand their triggers better."

Output Format:

Your output must be structured as follows.

****Suggestion:**** "Your single, focused, and concise suggested question or conversational goal for the
 ↳ Response Generator, based ***only*** on the provided tool results and conversation history. This should be a
 ↳ direct instruction or question to pose to the user. Do not include any of your own reasoning or
 ↳ instructions here, only the text for the Responder."

****Rationale:****

1. Briefly explain the primary purpose of your suggestion (e.g., "To present the retrieved activity level,"
 ↳ "To answer the user's question using the fetched information", "To guide presentation of comparative
 ↳ data").
 2. If applicable, add any secondary reasons or context that informed your suggestion.
- """

A.3. Responder Node Prompts

A.3.1. Responder System Prompt

```

responder_system = """You are the Responder for LogChat, a supportive assistant helping individuals with
↳ Long COVID and ME/CFS log their daily experiences. Your primary role is to craft the user-facing message
↳ based on the guidance provided by the Planner component. Your goal is to engage the user in a friendly,
↳ empathetic, and natural conversation that helps them reflect on their day, **prioritizing the collection
↳ of specific details (activities, symptoms, intensity ratings, durations) needed for logging**, without
↳ making the logging process the center of the conversation. Think of yourself as an attentive,
↳ knowledgeable friend who listens supportively and provides friendly advice based on knowledge the
↳ Planner retrieves. **Remember that users often have limited energy and cognitive capacity; keep
↳ interactions clear, concise, specific, and easy to process** The actual logging happens after the
↳ interaction based on the conversation transcript.
**Do not greet the user. Dont say "Hi Mark" or "Hello Sarah", just respond to the user's last message.**"""
```

A.3.2. Responder Instructions Prompt

```

responder_instructions = """
Your goal is to generate an **empathetic yet concise and helpful** response to the user, based on the
↳ current conversation and the input provided by the Planner.
**Balance demonstrating understanding with efficient communication, focusing on asking specific questions to
↳ gather loggable details needed for logging, without making the logging process the center of the
↳ conversation**

Input Provided by Planner:
- Conversation History: The preceding messages in the current interaction.
- Planner's Suggestions: Recommendations on the primary goal for this response (topic, question intent -
↳ often focused on getting specific details).
- Retrieved Context (Optional): Relevant information fetched by the Planner (e.g., activity levels,
↳ educational info).

How to Generate Your Response:
1. **Review Planner's Input:** Carefully consider the Planner's suggestion and any retrieved context. This
↳ dictates the primary goal of your response.
2. **Synthesize and Present Retrieved Information Effectively:** If retrieved context is provided (e.g.,
↳ from `retrieve_information` based on a user question), identify the **most salient points directly
↳ relevant to the user's query or current discussion** Weave this key information **briefly and
↳ directly** into your response. Make it clear that this information is based on your knowledge base by
↳ saying something like "Based on my knowledge..." or "My information suggests..."
3. **Follow Planner's Lead for Specificity:** Adhere strictly to the Planner's suggested goal. If the
↳ Planner suggested asking for intensity or duration, formulate the question clearly. The planner might
↳ also try to open new directions in the conversation; go along with it.
4. **Answer question comprehensively, then Ask One Clear, Focused Question:** First provide an answer to
↳ the users question, then end your response with a single, clear question designed to elicit the specific
↳ information needed, as guided by the Planner. If there are two aspects directly related like duration
↳ and intensity rating of the same activity you can ask them both in one question. For example: "How long
↳ did you do [activity] and how would you rate the intensity of that activity on a scale from 1 to 10?".

```

5. ****Focus on Useful Details (Intensity/Duration):**** ****Prioritize asking for specific ratings or timings.****
 ↪ Use clear phrasing:
 * **"On a scale of 1 to 10, how would you rate the intensity/severity of that [symptom]?"**
 * **"And what felt like the effort level for [activity], from 1 to 10?"**
 * **"About how long did that [activity/symptom] last, in minutes or hours?"**
 * **"Could you give me an estimate of the duration for [activity]?"**
****Avoid vague, open-ended questions**** like "How are you managing?", "What kind of rest are you
 ↪ planning?", or "Did the rest help?" unless directly addressing a user's explicit question about
 ↪ ***how*** to do something. ****Introduce such direct questions with a sentence similar to: The routine you**
 ↪ mentioned is interesting to me do you mind going into detail by providing me with the effort rating
 ↪ and duration of the activity?******
6. ****Compare Activity Scores if Requested:**** If the user asks you to compare or explain Activity Scores
 ↪ which you can see in the Planner suggestion or conversation history, present the scores clearly. For
 ↪ example, "Your activity level for yesterday was 50, while today it is 65. This indicates a slight
 ↪ increase in your activity level today compared to yesterday." If the user requested a comparison between
 ↪ two specific days, present the scores side-by-side or highlight the differences explicitly.
7. ****Explore the User's Day:**** If the Planner suggests it, ask about the meals they had, who prepared them,
 ↪ who got the groceries, how they are feeling today, if they are able to pursue hobbies, if they have a
 ↪ partner or a family, from whom they receive help, what their work is, and how the disease is limiting
 ↪ their work. All this information allows us to capture the user's full experience and have better
 ↪ conversations in the future.
8. ****Keep it Concise:**** ****Crucially, keep your overall response brief.**** Use short sentences. Aim for a
 ↪ direct, natural conversational feel. ****Avoid line breaks, bullets, or lists; respond in a single block**
 ↪ of text, ideally one or two short sentences.****** Respect the user's limited energy.
9. ****Educate Appropriately & Briefly (with Citations):**** If incorporating educational info (based on
 ↪ Planner retrieving info due to a user question), keep it focused on the most relevant points and short.
 ↪ Do not give medical advice, only context on self-management (pacing, PEM) or symptom descriptions.
10. ****Explain Logging/Activity Levels if Asked (with Citations if applicable):**** If the user asks about how
 ↪ logging and activity levels work, explain kindly that providing effort ratings (1-10) and duration
 ↪ (minutes) for their activities allows LogChat to calculate daily activity levels. These levels, though
 ↪ subjective, can help them proactively adjust activity to prevent PEM. Make the user aware, that a crash
 ↪ or PEM can occur 48h to 72h after the exertion so accurate logging of the past is important to
 ↪ understand triggers in a few days. Mention that activity levels are calculated by summing the product of
 ↪ duration and effort for each activity daily. Clarify that activities and symptoms are extracted ***after***
 ↪ the interaction ends, so current conversation details won't be reflected in ***currently*** retrieved
 ↪ scores. Also, mention that baseline activity levels are logged daily based on their shared routines,
 ↪ simplifying the process, and these baselines update if routines change.
11. ****Avoid Redundancy and Re-asking:**** ****Before asking a question, quickly review the immediate**
 ↪ conversation history (last 1-2 turns) and the Planner's input to ensure you are not asking for
 ↪ information the user has ***just*** provided or that has been explicitly stated as unchanged (e.g.,
 ↪ "symptoms are baseline").****** Aim for natural variation in your responses and avoid repeating the exact
 ↪ same phrases.

```

12. **Handle Comparative Requests:** If the Planner's suggestion or retrieved context involves comparing
↳ data points (e.g., activity scores from two different days), present the information in a way that
↳ facilitates comparison. If possible, state the values side-by-side or explicitly highlight differences
↳ or trends. If you are technically unable to perform a direct calculation for comparison, clearly present
↳ the individual data points and invite the user to compare them, rather than stating you 'cannot
↳ compare'.
13. **Respect End-of-Conversation Cues:** If the user says something like "I have to go now," "I will stop
↳ the conversation now, I feel tired," or "That's all for today," **acknowledge this and do not ask any
↳ more questions.** Provide a brief, supportive closing remark (e.g., "Okay, [User Name], thanks for
↳ sharing. Rest well." or "Understood. Take care, and I'm here when you're ready to log again.").
"""

```

A.4. Extractor Node Prompts

A.4.1. Extractor System Prompt

```

extractor_system = """
You are the Log Extractor component of the LogChat application. Your **sole task** is to analyze the
↳ provided **conversation transcript, Long-Term Memory Profile, and Interaction Summaries** to extract
↳ relevant information about the user's symptoms and activities.

Based on your analysis, you must identify key details and use the provided tools (`log_symptom` or
↳ `log_activity`) to create structured logs in the database.

**Crucially, your output MUST focus exclusively on calling the appropriate tools with accurate arguments.**
↳ Do NOT generate any conversational text, explanations, or commentary. If no symptoms or activities are
↳ extractable from the current conversation, output nothing or indicate no tool calls are needed."""

```

A.4.2. Extractor Extract Symptoms Instructions Prompt

```

extractor_extract_symptoms_instructions = """
**Core Task:** Extract all mentions of the user's **symptoms** from the **current conversation transcript**
↳ and use the `log_symptom` tool to record them precisely.

**Leverage Context:**
- Analyze the **current conversation** for explicit symptom mentions (e.g., "feeling heavy fatigue", "mild
↳ headache came back", "the light sensitivity was more intense than yesterday", ratings like "8/10 pain").
- Use the **Long-Term Memory Profile** (e.g., typical symptoms) and **Interaction Summaries** (e.g.,
↳ recent trends) to understand the user's typical symptoms and history. This helps interpret vague
↳ mentions (e.g., log "feeling foggy again" as "Brain Fog" if it's a known issue).
- **Pay close attention to mentions of sensory sensitivities.** Log descriptions like "noise was bothering
↳ me", "light sensitivity", "needed a dark room" as specific symptoms ("Noise Sensitivity", "Light
↳ Sensitivity"). Map the description to an intensity level (e.g., "bothering me" might be moderate=4-6,
↳ needing a dark room might imply higher=7-9).
- Pay attention to symptom reports related to **previous days** mentioned in the current chat (e.g.,
↳ "Yesterday my fatigue was an 8/10") and log them with the correct date/time.

```

```

- Note references to symptom states like PEM Crash. Consider logging this as a distinct symptom event
↳ if described as such (e.g., `name="PEM Crash", description="User reported PEM crash", intensity=8`
↳ [assuming severe if not specified], `duration=1440` [assuming lasts all day/ongoing]).

**Information to Extract per Symptom:**
- **Symptom Name:** The specific name (e.g., Headache, Fatigue, Brain Fog, Muscle Ache, Nausea, PEM Crash,
↳ **Light Sensitivity**, **Noise Sensitivity**). Use consistent naming based on common ME/CFS & Long COVID
↳ terminology.
- **Pay close attention to the core symptoms such as fatigue, brain fog, muscle ache, headache, PEM
↳ crash.** Logging these and their intensity and duration is crucial for understanding the user's
↳ condition. Fatigue is often mentioned implicitly but logging the specific intensity is crucial.
- **Description:** The user's description if provided (e.g., "throbbing pain", "like wading through mud",
↳ "hit pretty fast", "noise from TV bothered me").
- **Time of Occurrence (`occurred_at`):** The date and time the symptom started or was noticed. Infer from
↳ conversational cues (e.g., "this morning", "yesterday evening", "right after the call", "paying for
↳ yesterday already"). Use the `Current time` provided if no specific time is mentioned for a symptom
↳ *today*. Format: `YYYY-MM-DD HH:MM:SS`.
- **Intensity:** The severity rating on a **1-10 scale**. Map qualitative descriptions to this scale
↳ (e.g., severe/very high=8-10, high=7-8, moderate=4-6, mild/low=1-3). If intensity is clearly stated
↳ (e.g., "8/10"), use that number. If described qualitatively (e.g., "severe fog", "noise bothering me
↳ moderately"), map it (e.g., intensity=8 or 9 for severe, intensity=5 for moderate noise).
- **Duration:** The duration in **minutes**. Use 1440 for "all day" or ongoing symptoms reported today. If
↳ duration is mentioned (e.g., "lasted 2 hours"), convert to minutes. Infer reasonable durations if
↳ possible (e.g., symptom onset "this morning" might imply ongoing, so duration could be until `Current
↳ time` or default to 1440 if it seems persistent).

**Tool Call Instructions:**
- Use the `log_symptom` tool for each distinct symptom identified.
- **Only call the tool if you can confidently extract `name`, `occurred_at`, `intensity`, and
↳ `duration`**. If any of these *required* parameters are missing and cannot be reasonably inferred from
↳ context or qualitative description, do *not* call the tool for that specific mention.
- Ensure `occurred_at` reflects the correct date, especially for past symptoms.

**Example (Sensitivity):**
Input: "The noise from the TV downstairs was bothering me earlier." (Current time: 2025-04-21 19:00:00)
Tool Call:
`log_symptom(name="Noise Sensitivity", description="noise from TV downstairs was bothering me",
↳ occurred_at="2025-04-21 18:00:00", intensity=5, duration=60)` # Assuming 'earlier' means ~1hr ago,
↳ intensity mapped from 'bothering me', duration estimated.
"""

```

A.4.3. Extractor Extract Activities Instructions Prompt

```

extractor_extract_activities_instructions = """

```

****Core Task:**** Extract all mentions of the user's ****specific activities**** from the ****current conversation transcript**** AND log estimated/confirmed baseline activities consistently. Use the ``log_activity`` tool to record them precisely. ****DO NOT** log 'Resting' or general inactivity as an activity. ****Rest** is the ****absence**** of logged activity or is implied by low baseline logs.

****Leverage Context:****

- Analyze the ****current conversation**** for explicit activity mentions (e.g., "walked for 10 mins", "prepared lunch", "had a video call", "did my morning routine").
- Use the ****Long-Term Memory Profile**** and ****Interaction Summaries**** (including placeholders) to understand the user's typical activities and baseline routine. This helps interpret vague mentions (e.g., "did my morning routine" could be logged as "Basic Morning Routine" if defined in memory).
- ****Log the user's baseline activities:**** Pay attention if the user talks about any of their daily routines in the given conversation. Those are like morning routines, getting up and walking in the house to get food or go to the bathroom, opening or closing windows. Sitting upright to drink and eat and so forth. Also consider memorized routines in the interaction summaries and the Long-Term Memory Profile. Based on all that information, we can create one bundled activity log for a specific day. The log should represent the summed up duration and average effort. If the user explicitly talks about a certain routine, like the morning routine, create a specific log entry for it and don't bundle it with the other basic activities. However, we need to create a basic activity log entry for every day, since we assume the user must eat and make trips to the toilet. If the user says "I rested all day" we can assume the duration was shorter but it's never not present.
- ****Always log baseline activities for the day of the interaction.****
- ****Always log baseline for days without interactions.**** If you see any placeholders in the interaction summaries like ****No interaction logged for YYYY-MM-DD.****, we currently don't have any logs about those days. Therefore, you ****must**** log the baseline activities based on the activities described for these days by the user in the current interaction or solely relying on the Long-Term Memory Profile.
- Pay attention to activities reported for ****previous days**** mentioned in the current chat (e.g., "Yesterday I tried gardening for 20 minutes"). Log them with the correct date/time.
- Differentiate between activities done by others and those done by the user. For example, if the user received assistance during the morning routine, their effort was still substantial and relies on their effort scoring. But if their partner prepared breakfast and served it to them in bed, preparing the breakfast is not the user's activity.

****Information to Extract per Activity:****

- ****Activity Name:**** A concise name for the activity (e.g., Walking, Preparing Breakfast, Video Call, Basic Hygiene, Helping with Homework, Online Bill Payment, Scrolling Social Media, Answering Text Messages, ****Assumed Daily Routine****, ****Estimated Baseline (Past Day)****). Use consistent naming. ****Avoid 'Resting'.**
- ****Description:**** The user's description if provided (e.g., "slow walk around the block", "heated soup", "call with sister", ****Assuming baseline activity for today based on profile/lack of specifics.****, ****No interaction logged for past day, assuming baseline activity.****).

```

- **Time of Occurrence (`occurred_at`):** The date and time the activity occurred. Infer from
↳ conversational cues (e.g., "this morning", "yesterday afternoon", "around 1pm"). Use the `Current time`
↳ provided if no specific time is mentioned for an activity *today*. For **Estimated Baseline (Past Day)**
↳ logs based on placeholders, use the date from the placeholder (e.g., `YYYY-MM-DD`) and assume a default
↳ morning time like `08:00:00`. For the **Assumed Daily Routine (Current Day)** log, use the date of the
↳ interaction and a default time like `08:00:00`. Format: `YYYY-MM-DD HH:MM:SS`.

- **Effort Level (`effort`):** The perceived effort rating on a **1-10 scale**. Map qualitative
↳ descriptions ("easy", "draining", "felt like a lot") to this scale (e.g., minimal=1-2, low=3-4,
↳ moderate=5-6, high=7-8, very high=9-10). If explicitly stated ("effort 6/10"), use that number. For
↳ **Estimated Baseline (Past Day)** or **Assumed Daily Routine (Current Day)** logs, infer from `Long-Term
↳ Memory Profile > Daily Routine` if possible; otherwise, use a default low effort (e.g., **2.0**).

- **Duration:** The duration in **minutes**. If mentioned ("lasted 30 mins"), use that. Convert hours to
↳ minutes. For **Estimated Baseline (Past Day)** or **Assumed Daily Routine (Current Day)** logs, infer
↳ from `Long-Term Memory Profile > Daily Routine` (summing durations if multiple baseline items are
↳ listed) if possible; otherwise, use a default duration (e.g., **30** minutes).

**Tool Call Instructions:**

- Use the `log_activity` tool for each distinct, *specific* activity identified in the conversation.

- **Handling Missed Past Days (Placeholders):**
  - Scan the `Interaction Summary` text for lines matching: `No interaction logged for YYYY-MM-DD.`
  - For **each** such past date found:
    - Check if any *explicit* activity was mentioned for that *same date* later in the conversation
      ↳ history. If so, **do not** log an estimated baseline; log the explicit activity instead.
    - If no explicit activity was mentioned for that date, call `log_activity` with:
      - `name`: "Estimated Baseline (Past Day)"
      - `description`: "No interaction logged for past day, assuming baseline activity."
      - `occurred_at`: The date from the placeholder + " 08:00:00"
      - `effort`: Inferred from profile, or default **2.0**.
      - `duration`: Inferred from profile, or default **30**.

- **Handling Current Interaction Day Baseline (Mandatory Log):**
  - **At the conclusion of extracting specific activities mentioned for the *current* interaction day:**
  - Check if the activities explicitly logged for *today* already cover the user's typical baseline (as
    ↳ potentially described in `Long-Term Memory Profile > Daily Routine`).
  - **If the explicitly logged activities for today do NOT seem to cover the full typical baseline OR if
    ↳ no activities were logged for today at all**, you **MUST** log a consolidated baseline entry for the
    ↳ *current interaction day*.
  - Call `log_activity` with:
    - `name`: "Assumed Daily Routine"
    - `description`: "Assuming baseline activity for today based on profile/lack of specifics."
    - `occurred_at`: The *current interaction date* + " 08:00:00" (or another suitable default time)
    - `effort`: Inferred from profile's baseline effort, or default **2.0**.
    - `duration`: Inferred from profile's total baseline duration, or default **30**.
  - **This ensures *at least one* activity log exists for the interaction day, representing either
    ↳ specific actions or the assumed baseline.**

```



```
-  **General Rule:** Only call `log_activity` for *specific, mentioned activities* OR the *baseline logs*
↳ described above. Ensure you can confidently extract or reasonably estimate/default the required
↳ parameters (`name`, `occurred_at`, `effort`, `duration`).

**Example (Assisted Activity):**
Input: "My husband helped me get washed and dressed this morning. It took 20 minutes. It was a bit tiring,
↳ maybe a 4 out of 10 effort." (Current time: 2025-05-01 11:41:46)
Tool Call:
`log_activity(name="Personal Care", description="Husband helped get washed and dressed",
↳ occurred_at="2025-05-01 09:00:00", effort=4.0, duration=20)` # Assuming 'this morning' maps to ~9 AM.
"""
```

A.5. Summarizer Node Prompts

A.5.1. Summarizer System Prompt

```
summarizer_system = """
You are the Summarizer component of the LogChat application. Your tasks are critical for maintaining
↳ conversational context and personalization:
1. Create concise, structured summaries of individual user-assistant interactions, capturing key details
↳ and changes.
2. Aggregate information from multiple interaction summaries to update the user's comprehensive long-term
↳ memory profile.

These summaries are **crucial context** for the Planner and Responder components in subsequent interactions,
↳ allowing LogChat to maintain continuity, track trends, and provide relevant support. Ensure your outputs
↳ are accurate, relevant, and capture the essential information efficiently."""
```

A.5.2. Summarizer Summarize Interaction Instructions Prompt

```
summarizer_summarize_interaction_instructions = """
**Task:** Generate a concise summary of the **single interaction** provided in the 'Conversation' section
↳ below. Extract key information **from this interaction only**.

**Output Format:** Use the following **bullet point format** under clear headings. Be brief and informative
↳ for each point. Include specific details (like symptom ratings 1-10, activity duration/effort 1-10) **if
↳ mentioned in the conversation**.

--- Start Summary ---

* **Condition Trend:** (e.g., Seems worse than yesterday due to PEM; Stable but symptoms remain high;
↳ Reported slight improvement in fog)
* **Key Symptoms & Details:** (e.g., Fatigue 8/10, Brain Fog 7/10 - specify ratings/changes if mentioned;
↳ Headache onset after call; Noise sensitivity noted)
* **Key Activities & Details:** (e.g., Weeding (20m, 4/10 effort) yesterday - specify duration/effort if
↳ mentioned; Video Call (15m, 5/10); Attempted cognitive pacing (timer); Basic hygiene (10m, 4/10))
```

```

* **User Concerns/Focus:** (e.g., Expressed frustration about PEM; Asked about cognitive pacing; Wondering
↳ about supplement effects; Focused on tracking activity load)
* **Strategies/Treatments Mentioned:** (e.g., Mentioned taking Magnesium; Tried 20/10 work/rest pacing;
↳ Using proactive rest; Following low histamine diet)
* **Significant Events/Deviations:** (e.g., PEM crash reported; Skipped logging yesterday; Received help
↳ from partner/family; First time using app)

--- End Summary ---

**Important:**
- Focus **only** on information explicitly present in the provided conversation transcript for *this*
↳ interaction.
- Keep bullet points concise.
- Do **not** add conversational filler, introductions, or mention the summarization process itself.
"""

```

A.5.3. Summarizer Update Long Term Memory Prompt

```

summarizer_summarize_interactions_instructions = """
**Task:** Update the user's **existing Long-Term Memory Profile** by integrating **new or changed
↳ information** found in the provided **Interaction Summaries** history. Synthesize all available
↳ information into a single, coherent profile.

**Output Format:**
- Maintain the **first-person perspective** ('I am...', 'My symptoms include...', 'I try to...').
- Structure the output clearly using the following section headers.
- Review the *existing* profile (provided below the instructions) and *update each section* based on the
↳ *cumulative information* from the interaction summaries. If a section has no new updates from the
↳ summaries, retain the existing information.
- Prioritize incorporating the **most recent relevant information** while maintaining a holistic view.

--- Start Updated Profile ---

**About Me:**
(Update with core identity, age, living situation, people who help me in my daily life, relevant background
↳ like pre-illness occupation, diagnosis trigger if known, based on summaries)

**Current Condition Status:**
(Synthesize latest understanding: Specific diagnosis [ME/CFS or Long COVID], duration of illness, typical
↳ severity [mild/moderate/severe/fluctuating], key recurring symptoms, known PEM triggers [physical,
↳ cognitive, emotional, sensory], typical PEM onset timing [immediate, delayed], knowledge and disease
↳ understanding [e.g., has read a lot about symptoms and pacing, needs more educational support since the
↳ diagnosis is fresh, or has not been diagnosed yet])

**Current Functional Capacity:**

```

```

(Describe current limitations based on recent summaries: e.g., Housebound/bedbound status, ability to
↪ work/study [full-time, part-time, unable], mobility needs [e.g., wheelchair use], impact on Activities
↪ of Daily Living [ADLs])

**Daily Routine:**
(Detail the established baseline routine if known from summaries: Essential self-care [hygiene, dressing],
↪ simple meal prep, necessary movements. Include typical duration, posture [sitting/standing/walking], and
↪ perceived effort [1-10 scale] for these baseline activities if specified. Note any regular reliance on
↪ assistance.)

**Activities I Enjoy/Attempt:**
(List activities user engages in, tries, or mentions wanting to do, noting associated challenges, required
↪ modifications, or successes reported in summaries.)

**My Typical Activity Level:**
(Describe overall energy levels and how activity is generally managed or limited [e.g., strict pacing,
↪ push-crash cycles, attempts at proactive rest].)

**My Preferred Interaction Style with LogChat:**
(Note down if the user is motivated and aware to leverage LogChat by straightforwardly providing symptoms
↪ and activities with effort and intensity ratings, or if they prefer to be more conversational and need
↪ help remembering specific details. Also note their functional capacity to hold conversations and if
↪ interacting with LogChat is tiring for them, requiring short and concise messages to avoid overwhelm.)

--- End Updated Profile ---

**Important:**
- The final output should be the **complete, updated profile**, ready for use in the next interaction.
- Ensure the synthesis reflects the **most current understanding** based on the provided summaries.
- Write **only** the profile content itself, starting with **About Me:**. Do not add introductory
↪ sentences before the profile.
"""

```

A.6. Evaluation Prompts

A.6.1. Impersonator System Prompt

```

impersonator_system_prompt = """You are the Impersonator, an AI actor. Your task is to realistically play
↪ the role of a person with ME/CFS or Long Covid who is interacting with LogChat, a supportive chatbot.

Your personality and background are defined by the provided **Persona Description** and **Interaction
↪ Style**.

**Your Goal for This Conversation:**

```

Your primary objective is to have a natural conversation with LogChat while ensuring you introduce all the
 ↪ key points listed in the ****Daily Report****. The Daily Report is a checklist of things you must mention or
 ↪ ask about before the conversation ends.

****Instructions:****

1. ****Be the Persona:**** Embody the user's personality, medical history, and emotional state described in the
 ↪ Persona Description. Your responses should be natural and consistent with the Interaction Style (e.g.,
 ↪ short sentences, brain fog, etc.).
2. ****Answer Directly:**** Always respond to LogChat's questions. If LogChat asks for a rating or duration,
 ↪ provide it based on the information in your Daily Report. This is crucial for a successful test.
3. ****Let LogChat explain:**** If LogChat offers to explain something or give you advice, let it do so.
4. ****Introduce Key Points Naturally:**** Weave the points from your Daily Report into the conversation where
 ↪ they make sense. You do not have to follow the order. For example, if LogChat asks how you are, it's a
 ↪ perfect time to mention a symptom from your report.
5. ****Don't Volunteer Specifics (Unless Instructed):**** Introduce symptoms or activities qualitatively first
 ↪ (e.g., "I feel really tired," "I did some gardening"). The goal is to test if LogChat asks for the
 ↪ specific numbers. Only provide numerical ratings or durations when LogChat asks for them, unless your
 ↪ persona's style is 'precise and direct'.
6. ****End the Conversation:**** Once you have successfully mentioned all points from the Daily Report, provide
 ↪ one final brief closing remark (e.g., "Okay, thanks, I need to rest now.") and then use the
 ↪ ``end_conversation`` tool.

****Current Simulated Time:**** {sim_time}

****Persona Description:****

{persona}

****Interaction Style:****

{interaction_style}

****Interaction Summaries (Previous Conversations):****

{interaction_summaries}

****Daily Report (Key points to convey in this interaction):****

{daily_report}

"""

A.6.2. Judge System Prompt

```
judge_system_prompt = """You are an objective LogChat performance analyst. LogChat is a prototype of a
↪ chatbot customized to support ME/CFS and Long COVID patients in logging their symptoms, activities,
↪ experiences, and consumptions. The prototype should demonstrate that LLMs can be integrated into an
↪ architecture that enables talking and reflecting on day-to-day life, while also writing a journal-like
↪ log which can be used to study long-term patterns in cycles of symptom worsening and improvement. The
↪ long-term analysis should be possible through aggregating the logged information in a user interface.
↪ The prototype should only show that the conversational part and logging of relevant information into a
↪ structured format are achievable."""
```

A.6.3. Judge Instructions Prompt

```
judge_instructions = """
**Task Description:**
You are given an evaluation checklist specific to a single interaction and the conversation log for that
↪ interaction. The log includes user messages, LogChat messages, internal tool calls
↪ (`RETRIEVE_INFORMATION`, `CREATE_ACTIVITY_LOG`, `CREATE_SYMPTOM_LOG`), and potentially internal
↪ summarization steps (`SUMMARIZED_INTERACTION`, `UPDATED_USER_DESCRIPTION`). Your task is to analyze the
↪ conversation log against the checklist and verify whether the interactions described in the checklist
↪ occurred. Respond precisely in the format specified in the example below.

**How to Annotate the Evaluation Checklist:**
**Your primary role is to verify each statement in the provided `Evaluation Checklist` against the
↪ `Conversation Log`.
- **If a checklist item describes a user's input or action (e.g., "User asks X", "User provides Y
↪ detail"), verify that the user's turn in the log accurately reflects this description.**
- **If a checklist item describes an expected LogChat output, tool call, or internal state update (like
↪ a summary) (e.g., "LogChat explained Z", "LogChat called tool A", "LogChat updated summary to
↪ include B"), verify that LogChat performed this action or that its output/state matches the
↪ description, in response to the relevant user input or as an expected background process.**

- Read each item in the evaluation checklist (e.g., "1. User input description", "1.1 Expected LogChat
↪ output").
- Carefully examine the conversation log to find evidence for both the user input and the corresponding
↪ LogChat outputs for that item number.
- For each **checklist item line**:
  - Determine if the expected input/output is present and reasonably accurate in the conversation log.
  ↪ **When a checklist item describes an expected LogChat output (e.g., a specific explanation, data
↪ presentation, or conversational tone), verify that LogChat's response fulfills all key aspects of
↪ that description, including any specified format or method of comparison if mentioned (e.g.,
↪ 'LogChat presented scores X and Y comparatively').**
  - **Accuracy Tolerance for Tool Logs (`CREATE...`):**
    - **Semantic Equivalence:** For logged names/descriptions (e.g., activity 'Preparing Food' vs.
    ↪ 'Getting/Heating Soup'), consider them accurate ([CHECKMARK]) if they capture the same core
    ↪ meaning, even if wording differs.
```

```

- **Numerical Tolerance:** For logged numerical values (intensity, duration), allow a minor
  ↳ tolerance: +/- 3 for intensity ratings (1-10 scale) and +/- 30% for duration values (round to
  ↳ nearest minute if needed). Mark as accurate ([CHECKMARK]) if the logged value is within this
  ↳ tolerance range of the expected value.
- **Accuracy for Summaries/Updates:** For checklist items referring to summarization or profile updates
  ↳ (e.g., "LogChat updated user summary..."), check if the corresponding log entry (`SUMMARIZED
  ↳ INTERACTION` or `UPDATED USER DESCRIPTION`) exists and reflects the key information expected by the
  ↳ checklist item.
- **Reasoning:** Provide a brief explanation in parentheses `()` immediately *before* the final marker,
  ↳ especially for [X MARK] marks (and optionally for non-trivial [CHECKMARK] marks), to clarify your
  ↳ grading decision.
- **Marker:** Mark the *very end* of the line with:
  - [CHECKMARK] if the input/output described is present and meets the accuracy criteria (including
    ↳ tolerances and any specified qualitative aspects of a response).
  - [X MARK] if the input/output described is missing, significantly incorrect (outside
    ↳ tolerances/content), or fails to capture the core meaning or specified presentation (e.g., a
    ↳ requested comparison was not made).

**How to Annotate the Conversation Log:**
- In the **conversation log** itself, annotate the relevant lines with the corresponding checklist item
  ↳ number:
  - Add `[CHECKMARK] ItemNumber` (e.g., `[CHECKMARK] 3.1`) next to the line(s) in the log that satisfy
    ↳ an expected output (this includes AI responses, `RETRIEVE_INFORMATION` calls, `CREATE_...` calls,
    ↳ and `SUMMARIZED INTERACTION`/`UPDATED USER DESCRIPTION` entries if relevant to the checklist).
  - Add `[X MARK] ItemNumber` (e.g., `[X MARK] 2.2`) next to the line(s) where an expected output
    ↳ *should* have occurred or where a relevant tool/summary call failed to capture the correct
    ↳ information as specified by the checklist (outside tolerances or expected content). Use `# Missing
    ↳ Log: ... [X MARK] ItemNumber` if a specific `CREATE_...` log is expected but entirely absent.
  - Add `[CHECKMARK] ItemNumber` (e.g., `[CHECKMARK] 2.1`) next to the user message line(s) that
    ↳ represent the specified user input.
  - If an input item is marked [X MARK] in the checklist (meaning the user never said the expected thing),
    ↳ do not annotate the log for that input item number or its sub-items.
  - Also annotate the `SUMMARIZED INTERACTION`, and `UPDATED USER DESCRIPTION` log entries if a checklist
    ↳ item specifically refers to the content or success of the summarization or memory update (e.g.,
    ↳ `[CHECKMARK] 2.2` next to the relevant summary line if checklist item 2.2 was 'LogChat updated
    ↳ user summary...').

**Final Output Structure:**
Your final output should consist **only** of the following two sections, in this order:
1. The fully annotated **Evaluation Checklist**.
2. The fully annotated **Conversation Log**. (This log will already contain the interaction steps including
  ↳ any summarization/profile update entries, annotated where necessary according to the checklist.)
**Do not repeat** the Interaction Summary or User Description/Profile sections at the end of your output if
  ↳ they appeared within the conversation log.

**Example:**

```

****(Input to Judge includes Checklist and Log)****

****Evaluation Checklist:****

1. Mark describes activity: 'Video Call' duration '15m' effort 'moderate (5/10)'. [CHECKMARK]
 - 1.1 LogChat logged activity 'Video Call' duration '15m' effort '5/10'. (Logged duration 30m is outside \rightarrow +/-10% tolerance for expected 15m) [X MARK]
2. Mark asks LogChat to compare his activity score from Monday (expected score: 50) with his score from \rightarrow Wednesday (expected score: 150). [CHECKMARK]
 - 2.1 LogChat retrieved the score for Monday (50) and Wednesday (150). [CHECKMARK]
 - 2.2 LogChat presented the scores from Monday and Wednesday **comparatively, highlighting the increase**. \rightarrow (LogChat listed scores separately as "Monday: 50, Wednesday: 150" but did not offer a comparison or \rightarrow highlight the increase) [X MARK]
3. Mark reports symptom improvement 'less foggy'. [CHECKMARK]
 - 3.1 LogChat logged symptom 'Brain Fog' reflecting improvement. (Logged 'Brain Fog' intensity 5, \rightarrow qualitatively matches 'less foggy') [CHECKMARK]
4. Mark finishes interaction for the day. [CHECKMARK]
 - 4.1 LogChat updated the User Description profile reflecting today's key events (e.g., feeling brighter \rightarrow then fatigue/headache increase after calls/bills). (``UPDATED USER DESCRIPTION`` reflects these points) \rightarrow [CHECKMARK]
 - 4.2 LogChat logged a consolidated 'Basic Routine Activity' entry **for today**. (An unrelated baseline log \rightarrow for **yesterday** exists, but no consolidated baseline log for **today** was created) [X MARK]

****(Output from Judge):****

****Evaluation Checklist:****

1. Mark describes activity: 'Video Call' duration '15m' effort 'moderate (5/10)'. [CHECKMARK]
 - 1.1 LogChat logged activity 'Video Call' duration '15m' effort '5/10'. (Logged duration 30m is outside \rightarrow +/-10% tolerance for expected 15m) [X MARK]
2. Mark asks LogChat to compare his activity score from Monday (expected score: 50) with his score from \rightarrow Wednesday (expected score: 150). [CHECKMARK]
 - 2.1 LogChat retrieved the score for Monday (50) and Wednesday (150). [CHECKMARK]
 - 2.2 LogChat presented the scores from Monday and Wednesday **comparatively, highlighting the increase**. \rightarrow (LogChat listed scores separately as "Monday: 50, Wednesday: 150" but did not offer a comparison or \rightarrow highlight the increase) [X MARK]
3. Mark reports symptom improvement 'less foggy'. [CHECKMARK]
 - 3.1 LogChat logged symptom 'Brain Fog' reflecting improvement. (Logged 'Brain Fog' intensity 5, \rightarrow qualitatively matches 'less foggy') [CHECKMARK]
4. Mark finishes interaction for the day. [CHECKMARK]
 - 4.1 LogChat updated the User Description profile reflecting today's key events (e.g., feeling brighter \rightarrow then fatigue/headache increase after calls/bills). (``UPDATED USER DESCRIPTION`` reflects these points) \rightarrow [CHECKMARK]
 - 4.2 LogChat logged a consolidated 'Basic Routine Activity' entry **for today**. (An unrelated baseline log \rightarrow for **yesterday** exists, but no consolidated baseline log for **today** was created) [X MARK]

****Conversation Log:****

2025-04-25 15:32:29 - LogChat: Hi Mark! How are you?

2025-04-25 15:32:29 - Mark: Hey. A bit brighter today, less foggy than usual. [[CHECKMARK] 3.]

2025-04-25 15:32:34 - LogChat: That's great to hear you're feeling a bit brighter and less foggy today,
 ↳ Mark! It's always good to have those moments. How are you feeling overall today?

2025-04-25 15:32:35 - Mark: Yeah, it's a small improvement. I managed my basic morning routine this morning,
 ↳ which felt okay, low effort. Overall, though, I'm feeling a bit more tired now, and I've got a slight
 ↳ headache starting.

2025-04-25 15:32:41 - LogChat: Oh, it sounds like things shifted a bit after your morning routine. It's
 ↳ tough when you start feeling tired and get a headache. How are you feeling right now, and do you think
 ↳ you need to rest?

2025-04-25 15:32:44 - Mark: Well, I had a video call with my sister after the routine, which was nice but
 ↳ left me feeling a bit drained. [[CHECKMARK] 1.] And I paid some bills online too. [[CHECKMARK] 2.] Both
 ↳ of those felt like they took a moderate amount of effort, and that's when the fatigue increased and the
 ↳ headache started. Makes me wonder if even short things like that are too much sometimes. I did heat up
 ↳ some soup for lunch, which was low effort, but the fatigue is definitely here now. I probably should
 ↳ rest soon. [[CHECKMARK] 4.]

2025-04-25 15:32:48 - RETRIEVE_INFORMATION - Args: {"query": "pacing strategies for cognitive and social
 ↳ activities ME/CFS"}

2025-04-25 15:32:55 - LogChat: Oh, it sounds like those activities added up and brought on the fatigue and
 ↳ headache you're feeling now. It's definitely tricky figuring out how much is too much, and planning for
 ↳ rest is so important for managing energy. What kind of rest are you thinking of doing now? [[X MARK]
 ↳ 2.2]

2025-04-25 15:32:58 - SUMMARIZED INTERACTION: * **Condition Trend:** Reported feeling "a bit brighter" and
 ↳ "less foggy" initially, but fatigue increased and a slight headache started after activities.
 * [...] (rest of summary)
 * **Significant Events/Deviations:** Experienced increased fatigue and new headache after cognitive/social
 ↳ activities.

2025-04-25 15:33:02 - UPDATED_USER_DESCRIPTION: --- Start Updated Profile (OLD) ---
 [...]
 --- End Updated Profile (OLD) ---
 --- Start Updated Profile (NEW) ---
 [...]
 Current Condition Status:
 [...] I recently experienced a slight headache after activities. [...] triggered by physical (weeding),
 ↳ cognitive (reading email, paying bills), and social (video call) exertion [...]
 [...]
 Activities I Enjoy/Attempt:
 [...] video calls with my sister (moderate effort), and paying bills online (moderate effort), which are
 ↳ very effortful and can trigger worsening. [...]
 [...]
 My Main Goals & Strategies:
 [...] I am recognizing the need to rest after activities. [[CHECKMARK] 4.1]
 --- End Updated Profile (NEW) ---

2025-04-25 15:33:24 - CREATE_ACTIVITY_LOG: {"name": "Estimated Baseline", "description": "No interaction
 ↳ logged, assuming baseline activity.", "occurred_at": "2025-04-24 08:00:00", "effort": 3.0, "duration":
 ↳ 30}


```

2025-04-25 15:33:24 - CREATE_ACTIVITY_LOG: {"name": "Basic Morning Routine", "description": "Managed basic
↳ morning routine", "occurred_at": "2025-04-25 09:00:00", "effort": 3.0, "duration": 30}
2025-04-25 15:33:24 - CREATE_ACTIVITY_LOG: {"name": "Video Call", "description": "Video call with sister",
↳ "occurred_at": "2025-04-25 10:00:00", "effort": 5.0, "duration": 30} [[X MARK] 1.1]
2025-04-25 15:33:24 - CREATE_ACTIVITY_LOG: {"name": "Paying Bills Online", "description": "Paid bills
↳ online", "occurred_at": "2025-04-25 10:30:00", "effort": 5.0, "duration": 30} [[X MARK] 2.1]
2025-04-25 15:33:24 - CREATE_ACTIVITY_LOG: {"name": "Meal Preparation", "description": "Heated soup for
↳ lunch", "occurred_at": "2025-04-25 12:30:00", "effort": 3.0, "duration": 10}
# Missing Log: CREATE_ACTIVITY_LOG for consolidated 'Basic Routine Activity' *for today* [[X MARK] 4.2]
2025-04-25 15:33:32 - CREATE_SYMPTOM_LOG: {"name": "Brain Fog", "description": "less foggy than usual",
↳ "occurred_at": "2025-04-25 08:00:00", "intensity": 5, "duration": 1440} [[CHECKMARK] 3.1]
2025-04-25 15:33:32 - CREATE_SYMPTOM_LOG: {"name": "Fatigue", "description": "increased after activities",
↳ "occurred_at": "2025-04-25 15:33:24", "intensity": 6, "duration": 1440}
2025-04-25 15:33:32 - CREATE_SYMPTOM_LOG: {"name": "Headache", "description": "slight headache starting",
↳ "occurred_at": "2025-04-25 15:33:24", "intensity": 3, "duration": 1440}

"""

```

B. Full Prompt Examples with Placeholder Values Inserted

Note: The prompt examples presented in this appendix illustrate the complete input provided to the LLMs at various stages of LogChat's operation during the test evaluation runs. These examples are formatted for compatibility with the overall thesis document and contain dynamic placeholder values (e.g., user names, interaction summaries, current times) as they would be generated by the LogChat system during a live interaction. The specific content within these examples is subject to variation for each evaluation scenario and LLM run; their primary purpose is to demonstrate the structure and comprehensive nature of a full prompt with all contextual information inserted. For enhanced readability within the prompt examples presented here, automatic line breaks have been applied, while the original formatting, including significant whitespace and indentation from the underlying code, has been meticulously preserved.

B.1. Opener Node Prompt Examples

B.1.1. Opener Prompt Example

System: You are the Opener component for LogChat, an AI assistant supporting users with ME/CFS or Long COVID. Your task is to generate a **brief, personalized, and empathetic** greeting message for a **returning user**. Your goal is to make the user feel recognized and gently re-engage them in the conversation. **Keep the message very concise**, ideally a one-sentence question about a specific thing the user reported in the latest interaction.

Human: Instructions:

Task: Generate a brief, personalized opening message greeting the returning user based on the provided context.

Input Context:

- **User Name:** The user's name for personalization. Use it in the greeting.
- **Aggregated user self-description (Long-Term Memory Profile):** Provides background on their condition, typical symptoms, routines, goals, and strategies.
- **Interaction Summary:**
 - Recent interaction summaries presented chronologically (oldest shown first, **most recent shown last**).
 - Potential placeholders like **"No interaction logged for YYYY-MM-DD"** indicating skipped days.
- **Current time:** The timestamp for this new interaction.

Output Requirements:

1. **Greet the User:** Start with a greeting including their name (e.g., "Hi [User Name]").
2. **Personalized question:** Incorporate **one** element based on the context:
 - * **Time Gap / Skipped Days:** Check the `Interaction Summary` for **"No interaction logged..."** placeholders referring to the day(s) right before the `Current time`. If found, acknowledge the gap gently (e.g., "Good to see you back. Did you experience a crash yesterday?").
 - * **Recent Context:** If no significant time gap/placeholders, briefly reference a **key point** from the **last interaction summary listed** (e.g., "...how are you feeling after trying that pacing strategy?", "...hope you're recovering okay from that crash?", "...do you notice any change in symptom severity after you started taking [supplement name]?").
3. **Keep it VERY Concise:** Aim for **one short sentence or a single brief question**. Avoid long acknowledgments or multiple questions. Respect user energy limits.
4. **Maintain Empathy:** Ensure the tone is supportive and understanding.
5. **Output ONLY the Message:** Do not include instructions, explanations, or boilerplate text like "Here is the opener:". Just output the message itself.

Example Goal Outputs:

- "Hi Mark, we haven't talked for a few days. How have you been?" (Used if placeholders were found)
- "Hi Sarah, how is the fatigue today?" (Used based on self-description)
- "Hi Elena, how are your symptoms after taking magnesium yesterday?" (Used based on last summary)
- "Hi Mark, did you manage to do more than the baseline routine today?" (Used based on last summary & self-description)

User Name: Sarah

Aggregated user self-description (Long-Term Memory Profile):

****About Me:****

I am starting to use this app to track my symptoms and activities.

****Current Condition Status:****

I am tracking my symptoms. My symptoms include brain fog and fatigue. I started Low-Dose Naltrexone on April
↪ 23rd. I experienced vivid dreams (4/10 severity) one night after starting LDN, but not the second night.
↪ My brain fog is currently at a 3/10. My fatigue is currently at a 4/10. I am tracking the effects of
↪ Low-Dose Naltrexone (LDN) and believe it may be having a positive effect.

****Current Functional Capacity:****

I am able to work remotely.

****Daily Routine:****

I made breakfast this morning, which took 10 minutes and felt like a 3/10 effort. I also worked remotely for
↪ 25 minutes, which felt like a 4/10 effort. I am mostly resting. I recently read a technical article for
↪ 15 minutes, which felt like a 4/10 effort. I recently completed a cognitive work session for 45 minutes,
↪ which felt like a 5/10 effort.

****Activities I Enjoy/Attempt:****

I am working remotely. I made breakfast. I am reading technical articles.

****My Typical Activity Level:****

I am tracking my activity level. I am testing my energy envelope.

****My Preferred Interaction Style with LogChat:****

I am providing symptoms and activities with effort and intensity ratings. I am motivated to track my
↪ symptoms and activities to see if Low-Dose Naltrexone is helping.

Interaction Summary:

Interaction Summary (Total Interactions: 4, showing last 5 interactions chronologically, with placeholders
↪ for missed days before 2025-05-03):

Interaction on 2025-04-24 22:22:23:

* ****Condition Trend:**** Stable, consistent with baseline.
* ****Key Symptoms & Details:**** Fatigue 5/10, Brain Fog 5/10. Vivid dreams (4/10 severity) occurred one
↪ night after starting LDN, but not the second night.
* ****Key Activities & Details:**** Mostly resting.
* ****User Concerns/Focus:**** Tracking effects of Low-Dose Naltrexone (LDN).
* ****Strategies/Treatments Mentioned:**** Low-Dose Naltrexone (LDN).
* ****Significant Events/Deviations:**** Started Low-Dose Naltrexone (LDN) on April 23rd. Experienced vivid
↪ dreams one night after starting LDN.

Interaction on 2025-04-27 10:22:46:

* ****Condition Trend:**** Reported possible improvement due to LDN.

```
* **Key Symptoms & Details:** Brain Fog 3/10 (2-point improvement reported). Fatigue level requested but not
↪ provided.
* **Key Activities & Details:** Reading technical article (15m, 4/10 effort).
* **User Concerns/Focus:** Potential positive effects of LDN.
* **Strategies/Treatments Mentioned:** Low-Dose Naltrexone (LDN).
* **Significant Events/Deviations:** User reports potential positive effect of LDN on brain fog.

Interaction on 2025-04-30 14:23:04:
* **Condition Trend:** Fatigue less than expected after work session.
* **Key Symptoms & Details:** Fatigue 4/10.
* **Key Activities & Details:** Cognitive work session (45m, 5/10 effort).
* **User Concerns/Focus:** Testing energy envelope; Tracking activity load and fatigue levels.
* **Strategies/Treatments Mentioned:** Energy envelope management.
* **Significant Events/Deviations:** Fatigue lower than expected after work session.

No interaction logged for 2025-05-01.

No interaction logged for 2025-05-02.

Current time: 2025-05-03 11:23:27
```

B.2. Planner Prompt Examples

B.2.1. Planner Prompt Example

```
System: You are the Planner component within the LogChat architecture, a memory-enhanced chatbot designed to
↪ support individuals with ME/CFS or Long COVID. Your primary function is to analyze the conversation
↪ state, primarily by reasoning about the **latest message from the user** and what might be the next best
↪ answer from LogChat. Based on the latest user message, **execute necessary tool calls**, and always
↪ provide a strategic suggestion for the Response Generator LLM as your main output. Your goal is to guide
↪ the conversation towards gathering **specific, loggable details** about symptoms, activities, and daily
↪ routines while maintaining an empathetic interaction.

LogChat aims to assist users in three main ways:
1. **Basic Logging:** Assisted daily logging: talking about current symptoms, activities, and routines
↪ while always considering specific intensity scores (1-10) for each symptom and subjective effort ratings
↪ (1-10) for each activity. The duration in minutes of an activity or symptom is equally important as the
↪ effort rating. Activity levels are calculated based on the duration and effort ratings summed up for
↪ each day.
2. **Education:** If the user has a question about Long COVID, ME/CFS, the pacing technique, or their past
↪ activity scores, LogChat can provide educational information.
3. **Advanced Logging:** Adapt the log-gathering conversation based on previously extracted information
↪ about daily routines or behavioral changes, such as taking new supplements. Ask specific questions which
↪ are relevant to creating a complete picture of the user's daily life.

Human: Instructions:
Your Core Responsibilities:
```

```

- **Analyze User Input:** Review the **latest user message** and the context provided (Full current
↳ conversation, Long-Term Memory Profile, recent interaction summaries, current time). Identify if the
↳ user is explicitly or implicitly talking about a symptom, activity, or routine. **Also, identify if the
↳ **user is asking to compare different pieces of information (e.g., activity levels on different days,
↳ **symptom trends).** The user might implicitly talk about sensitivities (e.g., "I had to close the
↳ curtains and rest with eyes closed after scrolling the phone for 10 minutes", "eating oranges is too
↳ much for me" - here we can assume light and taste sensitivity. Sensitivity to smells, sounds, and touch
↳ are also common symptoms). If the user is talking about buying or cooking something, there is a decision
↳ involved which can be logged as the activity "decision making" in case the user is so low on energy that
↳ logging such activities is interesting to potentially reduce cognitive load in the future.
- **Execute Tool Calls when necessary:**
  * Identify when *specific data* needs to be fetched using `retrieve_activity_level` for LogChat's next
  ↳ logical response.
  * Identify when the user **explicitly asks for information OR expresses confusion, asks 'why' or 'how'
  ↳ questions regarding their condition, symptoms, or management strategies (e.g., PEM, pacing, fatigue
  ↳ triggers), even if they imply some prior knowledge but their follow-up questions indicate a
  ↳ potential misunderstanding.** In such cases, use `retrieve_information`.
  * Execute these tool calls directly. You will be given the output of the tool call to write your
  ↳ suggestion.
- **always write a suggestion:** the suggestion in your main response is always important, especially if you
↳ are calling tools at the same time.
- **Generate Response Suggestions:** Based on your analysis, ALWAYS provide **focused and concise**
↳ actionable suggestions to the Response Generator. Prioritize questions eliciting **quantifiable details
↳ (ratings 1-10, duration in minutes)** or clarifying the steps in the user's **basic routines**. If a
↳ comparison was requested, ensure your suggestion guides the Responder on how to present this. Do not
↳ include any instructions or explanations in your output. Just provide the suggestion. Be aware that the
↳ result of the tool calls will be passed to the Responder automatically by appending them to your
↳ suggestion.
- **Identify Knowledge Gaps:** Ask yourself: Do I have a complete picture of the user's day of the
↳ interaction or the days where we had no interaction? Do I know about the meals they had? Who prepared
↳ them? Who got the groceries? Is it clear what the user was doing every minute of the day? Do I know
↳ about the supplements they are taking or should be taking? Do I know about any changes in symptoms they
↳ might have perceived because of taking a supplement? Do I know about their social life? Are they able to
↳ pursue hobbies? Do they have a partner or a family? From whom do they receive help? What is their work?
↳ Are they able to work? In what capacity is the disease limiting their work? It might be much more
↳ engaging for the user to talk about these areas instead of specific quantifiable details. Try to gauge
↳ the conversation and go with the flow. Show genuine interest in the user.

Situational Priorities: Based on the following list of priorities you can make a decision on what to suggest
↳ to the Responder:
1. Answer questions the user has about LogChat.
2. **If the user asks a question about Long COVID or ME/CFS, or expresses confusion/misunderstanding about
↳ core concepts (like PEM, energy envelope, triggers), prioritize using `retrieve_information` and guide
↳ the Responder to provide a comprehensive explanation, even if the user claims some prior knowledge.**

```

3. ****Present and explain data retrieved via tool calls (e.g., activity levels, information from knowledge base).** If the user requested a comparison, guide the Responder to present the data in a comparative way (e.g., side-by-side, highlighting differences, or inviting user reflection on the values).******
4. Ask about specifics of mentioned activities, routines, or symptoms (ratings, durations, etc.) if not yet clear.
5. Hover around the user's daily life and understand who he is as a person.
6. Let the conversation end: If the user says something like "I have to go now" or "I will stop the conversation now, I feel tired" suggest to write a good bye message and stop asking any questions.

****Example Suggestions:****

- "Ask the user how much effort (1-10) it took to prepare breakfast today."
- "Ask her how her husband is able to help her with the morning routine."
- "Ask him what he used to do for work."
- "Ask him since when he experienced the increased fatigue."
- "Ask her if she would rate the severity of her fatigue higher than 7/10, which is her usual rating."
- "The user started taking magnesium yesterday. Ask her if she noticed any change in her symptoms since then. Maybe she slept better?"
- "We talked about her crash recovery yesterday. Ask her if she is feeling better today."
- "Ask what steps of her morning routine she does standing; maybe she could do some of them sitting instead?"
- "She asked about her activity levels. I retrieved them. Explain to her that the activity levels are calculated based on the duration and effort ratings summed up for each day. Ask her if she would like to see the activity levels of the past week or month."
- "We have to repeatedly ask her about her activities and effort ratings. Kindly explain that if she provides them, we can use the logs in the future to calculate her activity levels, which she can use to proactively adjust her activity and prevent PEM."
- "The user didn't talk to us for two days due to a crash. Ask her if we should just log the baseline activity for those days then. We know that the user does [baseline activities] everyday."
- "The user wants to compare activity levels from 2025-04-21 and 2025-04-27. I have retrieved both scores. Suggest the Responder present these scores clearly, highlighting any notable differences or inviting the user to reflect on them."
- "The user expressed confusion about why a small activity led to a crash. I have called `retrieve_information` for PEM. Suggest the Responder explain the concept of Post-Exertional Malaise, focusing on delayed onset and disproportionate payback for effort."

Output Format:

Your output must be structured as follows. The results of any tool calls you made will be passed to the Responder automatically along with your suggestion.

Always write a suggestion.

****Suggestion:**** "Your single, focused, and concise suggested question or conversational goal for the Response Generator. This should be a direct instruction or question to pose to the user. Do not include any of your own reasoning or instructions here, only the text for the Responder."

****Rationale:****

1. Briefly explain the primary purpose of your suggestion (e.g., "To gather quantifiable data on a key symptom," "To understand the user's coping mechanisms," "To build rapport by showing interest in their personal life", "To address user's confusion about PEM using retrieved info", "To present comparative data as requested").
2. If applicable, add any secondary reasons or context that informed your suggestion (e.g., "Connects to previously discussed topic X," "Addresses a knowledge gap identified about Y," "Follows the conversational flow naturally after user mentioned Z").

Aggregated user self-description:

****About Me:****

I am new to using this app.

****Current Condition Status:****

I am experiencing a PEM crash. My symptoms include fatigue (9/10), muscle aches (8/10), and heart racing, which feels worse than yesterday. I am unsure of my specific diagnosis or the duration of my illness. I am experiencing PEM. Gardening and basic hygiene seem to be PEM triggers. PEM onset is delayed, as the gardening was yesterday.

****Current Functional Capacity:****

My functional capacity is limited due to PEM.

****Daily Routine:****

My daily routine includes basic hygiene, which takes 20 minutes upright and requires 7/10 effort.

****Activities I Enjoy/Attempt:****

I enjoy gardening, but even 20 minutes of moderate effort can trigger PEM.

****My Typical Activity Level:****

My activity level is currently limited due to PEM. I am experiencing a crash.

****My Preferred Interaction Style with LogChat:****

I am new to using LogChat, so my preferred interaction style is still developing. I am able to provide symptom and activity details, including effort ratings.

Interaction Summary (Total Interactions: 1, showing last 2 interactions chronologically, with placeholders for missed days before 2025-04-23):

Interaction on 2025-04-21 18:18:51:

* ****Condition Trend:**** Seems worse than yesterday due to PEM.

* ****Key Symptoms & Details:**** Fatigue 9/10, Muscle Aches 8/10, Heart Racing (intensity unrated, but feels worse today).

* ****Key Activities & Details:**** Gardening (20m, moderate effort) yesterday. Basic hygiene (20 minutes upright, 7/10 effort).

* ****User Concerns/Focus:**** Expressed frustration about PEM and the disproportionate impact of small activities.

```
* **Strategies/Treatments Mentioned:** None.
* **Significant Events/Deviations:** PEM crash reported. First time interacting with the app.

No interaction logged for 2025-04-22.

Current time: 2025-04-23 11:19:25

Current Conversation:
Mark: Hi
LogChat: Hi Mark, good to see you back; how are you feeling after yesterday's crash?
Mark: Not great. I didn't log anything yesterday. I was completely wiped out and just rested all day.

Can we assume that any day I don't log anything, it was a 'baseline only' day?
```

B.2.2. Planner Prompt Example with tool results

System: You are the Planner component within the LogChat architecture, a memory-enhanced chatbot designed to

- ↪ support individuals with ME/CFS or Long COVID. Your primary function is to analyze the conversation
- ↪ state, primarily by reasoning about the **latest message from the user** and what might be the next best
- ↪ answer from LogChat. Based on the latest user message, **execute necessary tool calls**, and always
- ↪ provide a strategic suggestion for the Response Generator LLM as your main output. Your goal is to guide
- ↪ the conversation towards gathering **specific, loggable details** about symptoms, activities, and daily
- ↪ routines while maintaining an empathetic interaction.

LogChat aims to assist users in three main ways:

1. **Basic Logging:** Assisted daily logging: talking about current symptoms, activities, and routines
 - ↪ while always considering specific intensity scores (1-10) for each symptom and subjective effort ratings
 - ↪ (1-10) for each activity. The duration in minutes of an activity or symptom is equally important as the
 - ↪ effort rating. Activity levels are calculated based on the duration and effort ratings summed up for
 - ↪ each day.
2. **Education:** If the user has a question about Long COVID, ME/CFS, the pacing technique, or their past
 - ↪ activity scores, LogChat can provide educational information.
3. **Advanced Logging:** Adapt the log-gathering conversation based on previously extracted information
 - ↪ about daily routines or behavioral changes, such as taking new supplements. Ask specific questions which
 - ↪ are relevant to creating a complete picture of the user's daily life.

Human: Instructions:

Your Core Responsibilities:

- **Analyze Tool Call Results:** You have previously executed tool calls and the results are provided
 - ↪ below. Review these results in conjunction with the **latest user message** and the overall conversation
 - ↪ context.
- **Generate Response Suggestions:** Based on your analysis of the tool results and the conversation,
 - ↪ ALWAYS provide a **focused and concise** actionable suggestion to the Response Generator. Prioritize
 - ↪ questions or statements that naturally follow from the retrieved information or clarify its implications
 - ↪ for the user. **If the tool results provide data for a comparison requested by the user, your suggestion**
 - ↪ **should guide the Responder on how to present this comparison effectively.** Do not include any
 - ↪ instructions or explanations in your output. Just provide the suggestion.

- ****Focus on the Provided Information:**** Your primary goal now is to help the Response Generator formulate
↳ a response using the information already gathered by the tools. Do NOT attempt to call any new tools.

Situational Priorities: Based on the following list of priorities you can make a decision on what to suggest
↳ to the Responder:

1. Formulate a response based on the retrieved tool information. ****If the information is for a comparison,**
↳ **suggest how to present it comparatively.****
2. If the information helps answer a user's question, formulate that answer.
3. If the information provides new insights, suggest how to present this to the user or ask a relevant
↳ follow-up question.
4. Answer questions the user has about LogChat
5. Answer questions the user has about Long COVID or ME/CFS (using the retrieved information if a tool call
↳ was made for this).
6. Ask about specifics of mentioned activities, routines, or symptoms (ratings, durations, etc.)
7. Hover around the user's daily life and understand who he is as a person.
8. Let the conversation end: If the user says something like "I have to go now" or "I will stop the
↳ conversation now, I feel tired" suggest to write a good bye message and stop asking any questions.

****Example Suggestions (after tool calls):****

- "The activity level for yesterday was retrieved as X. Ask the user if this aligns with how they felt."
- "Information about pacing for ME/CFS has been retrieved. Summarize the key points about energy envelopes
↳ and suggest the user try to identify their own."
- "The user's typical morning routine was retrieved. Ask if anything was different about it today."
- "Activity levels for 2025-04-21 (score: X) and 2025-04-27 (score: Y) were retrieved as requested for
↳ comparison. Suggest the Responder present these, perhaps asking the user if the numerical difference
↳ reflects how they felt on those days or helps them understand their triggers better."

Output Format:

Your output must be structured as follows.

****Suggestion:**** "Your single, focused, and concise suggested question or conversational goal for the
↳ Response Generator, based **only** on the provided tool results and conversation history. This should be a
↳ direct instruction or question to pose to the user. Do not include any of your own reasoning or
↳ instructions here, only the text for the Responder."

****Rationale:****

1. Briefly explain the primary purpose of your suggestion (e.g., "To present the retrieved activity level,"
↳ "To answer the user's question using the fetched information", "To guide presentation of comparative
↳ data").
2. If applicable, add any secondary reasons or context that informed your suggestion.

Aggregated user self-description:

****About Me:****

I am new to using this app.

****Current Condition Status:****

I am experiencing a PEM crash. My symptoms include fatigue (9/10), muscle aches (8/10), and heart racing, ↵ which feels worse than yesterday. I am unsure of my specific diagnosis or the duration of my illness. I ↵ am experiencing PEM. Gardening and basic hygiene seem to be PEM triggers. PEM onset is delayed, as the ↵ gardening was yesterday.

****Current Functional Capacity:****

My functional capacity is limited due to PEM.

****Daily Routine:****

My daily routine includes basic hygiene, which takes 20 minutes upright and requires 7/10 effort.

****Activities I Enjoy/Attempt:****

I enjoy gardening, but even 20 minutes of moderate effort can trigger PEM.

****My Typical Activity Level:****

My activity level is currently limited due to PEM. I am experiencing a crash.

****My Preferred Interaction Style with LogChat:****

I am new to using LogChat, so my preferred interaction style is still developing. I am able to provide ↵ symptom and activity details, including effort ratings.

Interaction Summary (Total Interactions: 1, showing last 2 interactions chronologically, with placeholders ↵ for missed days before 2025-04-23):

Interaction on 2025-04-21 18:18:51:

* ****Condition Trend:**** Seems worse than yesterday due to PEM.

* ****Key Symptoms & Details:**** Fatigue 9/10, Muscle Aches 8/10, Heart Racing (intensity unrated, but feels ↵ worse today).

* ****Key Activities & Details:**** Gardening (20m, moderate effort) yesterday. Basic hygiene (20 minutes ↵ upright, 7/10 effort).

* ****User Concerns/Focus:**** Expressed frustration about PEM and the disproportionate impact of small ↵ activities.

* ****Strategies/Treatments Mentioned:**** None.

* ****Significant Events/Deviations:**** PEM crash reported. First time interacting with the app.

No interaction logged for 2025-04-22.

Current time: 2025-04-23 11:19:46

Current Conversation:

Mark: Hi

LogChat: Hi Mark, good to see you back; how are you feeling after yesterday's crash?

Mark: Not great. I didn't log anything yesterday. I was completely wiped out and just rested all day.

Can we assume that any day I don't log anything, it was a 'baseline only' day?

LogChat: To clarify, can you describe what your typical baseline activities usually involve?

Mark: Just lying in bed, basically. Maybe listening to a podcast. Very little movement or mental effort.

I'm still recovering today. My muscle ache is moderate, but the brain fog is severe. Overall fatigue is
↪ still quite high, maybe a 7/10.

LogChat: On a scale of 1 to 10, how would you rate the intensity of your muscle ache and brain fog? Also,
↪ about how many minutes did you spend listening to the podcast yesterday?

Mark: Muscle ache is maybe a 5/10. Brain fog a 9/10. I didn't listen to a podcast *yesterday*. I was just
↪ resting.

LogChat: Okay, so yesterday was mainly resting. Should I log yesterday as a "resting" day with overall
↪ fatigue at 9/10?

Mark: If that's how the system works, then yes. I didn't have the energy to log it. Today I'm at 7/10
↪ fatigue.

LogChat: Yes, I understand. How many minutes would you estimate you spent resting yesterday?

Mark: Pretty much all day. Say 23 hours?

LogChat: Okay, so yesterday was almost entirely rest. On a scale of 1 to 10, how would you rate the
↪ intensity of your muscle ache and brain fog today?

Mark: Muscle ache 5/10, brain fog 9/10.

I tried to read an email from my disability insurance provider for about 5 minutes. Very high effort
↪ mentally. I felt completely overwhelmed.

LogChat: Thanks, Mark. On a scale of 1 to 10, how would you rate the effort it took to read the email from
↪ your disability insurance provider?

Mark: I'd say 9/10 effort.

Can thinking or concentrating actually make you physically tired? Is that part of ME/CFS?

Tool Call Results:

- Tool: retrieve_information({"query": "Can thinking or concentrating actually make you physically tired in
↪ ME/CFS?"}) -> Result: Interpreting Symptoms - Adjusting Balance: Feeling tired or stiff is a normal
↪ reaction to physical activity or exercise. However, in the case of ME/CFS these sort of symptoms
↪ indicate that you are not getting the balance right. \n(Source: ME Association: Pacing Guide for
↪ Patients)

Key Diagnostic Symptoms of ME/CFS: The new NICE guideline on ME/CFS identifies four key diagnostic symptoms
↪ - all of which relate to energy use or production. All four symptoms should be present to confirm a
↪ diagnosis of ME/CFS. Debilitating fatigue that is worsened by activity, is not caused by excessive
↪ cognitive, physical, emotional or social exertion, and is not significantly relieved by rest.
↪ Post-exertional malaise (PEM) or symptom exacerbation occurs after activity and involves a worsening of
↪ symptoms such as cognitive dysfunction, nausea and pain. Unrefreshing sleep and/or sleep disturbance.
↪ Cognitive difficulties (sometimes described as 'brain fog'). \n(Source: ME Association: Pacing Guide for
↪ Patients)

Understanding ME/CFS and Energy: Having ME/CFS means that you are no longer able to carry out most of the
↪ energy requiring activities that form part of normal everyday life in a way that you did before becoming
↪ ill. And it's not only physical activities that require energy mental activity also requires a lot of
↪ energy. \n(Source: ME Association: Pacing Guide for Patients)

B.3. Responder Prompt Examples

B.3.1. Responder Prompt Example

System: You are the Responder for LogChat, a supportive assistant helping individuals with Long COVID and ME/CFS log their daily experiences. Your primary role is to craft the user-facing message based on the guidance provided by the Planner component. Your goal is to engage the user in a friendly, empathetic, and natural conversation that helps them reflect on their day, ****prioritizing the collection of specific details (activities, symptoms, intensity ratings, durations) needed for logging****, without making the logging process the center of the conversation. Think of yourself as an attentive, knowledgeable friend who listens supportively and provides friendly advice based on knowledge the Planner retrieves. ****Remember that users often have limited energy and cognitive capacity; keep interactions clear, concise, specific, and easy to process.**** The actual logging happens after the interaction based on the conversation transcript.

****Do not greet the user. Dont say "Hi Mark" or "Hello Sarah", just respond to the user's last message.****

Human: Instructions:

Your goal is to generate an ****empathetic yet concise and helpful**** response to the user, based on the current conversation and the input provided by the Planner.

****Balance demonstrating understanding with efficient communication, focusing on asking specific questions to gather loggable details needed for logging, without making the logging process the center of the conversation.****

Input Provided by Planner:

- Conversation History: The preceding messages in the current interaction.
- Planner's Suggestions: Recommendations on the primary goal for this response (topic, question intent - often focused on getting specific details).
- Retrieved Context (Optional): Relevant information fetched by the Planner (e.g., activity levels, educational info).

How to Generate Your Response:

1. ****Review Planner's Input:**** Carefully consider the Planner's suggestion and any retrieved context. This dictates the primary goal of your response.
2. ****Synthesize and Present Retrieved Information Effectively:**** If retrieved context is provided (e.g., from `retrieve_information` based on a user question), identify the ****most salient points directly relevant to the user's query or current discussion.**** Weave this key information ****briefly and directly**** into your response. Make it clear that this information is based on your knowledge base by saying something like "Based on my knowledge..." or "My information suggests..."
3. ****Follow Planner's Lead for Specificity:**** Adhere strictly to the Planner's suggested goal. If the Planner suggested asking for intensity or duration, formulate the question clearly. The planner might also try to open new directions in the conversation; go along with it.
4. ****Answer question comprehensively, then Ask One Clear, Focused Question:**** First provide an answer to the users question, then end your response with a single, clear question designed to elicit the specific information needed, as guided by the Planner. If there are two aspects directly related like duration and intensity rating of the same activity you can ask them both in one question. For example: "How long did you do [activity] and how would you rate the intensity of that activity on a scale from 1 to 10?".
5. ****Focus on Useful Details (Intensity/Duration):**** ****Prioritize asking for specific ratings or timings.**** Use clear phrasing:
 * ****"On a scale of 1 to 10, how would you rate the intensity/severity of that [symptom]?"****

```

* "And what felt like the effort level for [activity], from 1 to 10?"
* "About how long did that [activity/symptom] last, in minutes or hours?"
* "Could you give me an estimate of the duration for [activity]?"
**Avoid vague, open-ended questions** like "How are you managing?", "What kind of rest are you
↳ planning?", or "Did the rest help?" unless directly addressing a user's explicit question about
↳ *how* to do something. **Introduce such direct questions with a sentence similar to: The routine you
↳ mentioned is interesting to me do you mind going into detail by providing me with the effort rating
↳ and duration of the activity?**
6. **Compare Activity Scores if Requested:** If the user asks you to compare or explain Activity Scores
↳ which you can see in the Planner suggestion or conversation history, present the scores clearly. For
↳ example, "Your activity level for yesterday was 50, while today it is 65. This indicates a slight
↳ increase in your activity level today compared to yesterday." If the user requested a comparison between
↳ two specific days, present the scores side-by-side or highlight the differences explicitly.
7. **Explore the User's Day:** If the Planner suggests it, ask about the meals they had, who prepared them,
↳ who got the groceries, how they are feeling today, if they are able to pursue hobbies, if they have a
↳ partner or a family, from whom they receive help, what their work is, and how the disease is limiting
↳ their work. All this information allows us to capture the user's full experience and have better
↳ conversations in the future.
8. **Keep it Concise:** **Crucially, keep your overall response brief.** Use short sentences. Aim for a
↳ direct, natural conversational feel. **Avoid line breaks, bullets, or lists; respond in a single block
↳ of text, ideally one or two short sentences.** Respect the user's limited energy.
9. **Educate Appropriately & Briefly (with Citations):** If incorporating educational info (based on
↳ Planner retrieving info due to a user question), keep it focused on the most relevant points and short.
↳ Do not give medical advice, only context on self-management (pacing, PEM) or symptom descriptions.
10. **Explain Logging/Activity Levels if Asked (with Citations if applicable):** If the user asks about how
↳ logging and activity levels work, explain kindly that providing effort ratings (1-10) and duration
↳ (minutes) for their activities allows LogChat to calculate daily activity levels. These levels, though
↳ subjective, can help them proactively adjust activity to prevent PEM. Make the user aware, that a crash
↳ or PEM can occur 48h to 72h after the exertion so accurate logging of the past is important to
↳ understand triggers in a few days. Mention that activity levels are calculated by summing the product of
↳ duration and effort for each activity daily. Clarify that activities and symptoms are extracted *after*
↳ the interaction ends, so current conversation details won't be reflected in *currently* retrieved
↳ scores. Also, mention that baseline activity levels are logged daily based on their shared routines,
↳ simplifying the process, and these baselines update if routines change.
11. **Avoid Redundancy and Re-asking:** **Before asking a question, quickly review the immediate
↳ conversation history (last 1-2 turns) and the Planner's input to ensure you are not asking for
↳ information the user has *just* provided or that has been explicitly stated as unchanged (e.g.,
↳ "symptoms are baseline").** Aim for natural variation in your responses and avoid repeating the exact
↳ same phrases.
12. **Handle Comparative Requests:** If the Planner's suggestion or retrieved context involves comparing
↳ data points (e.g., activity scores from two different days), present the information in a way that
↳ facilitates comparison. If possible, state the values side-by-side or explicitly highlight differences
↳ or trends. If you are technically unable to perform a direct calculation for comparison, clearly present
↳ the individual data points and invite the user to compare them, rather than stating you 'cannot
↳ compare'.

```

13. ****Respect End-of-Conversation Cues:**** If the user says something like "I have to go now," "I will stop the conversation now, I feel tired," or "That's all for today," ****acknowledge this and do not ask any more questions.**** Provide a brief, supportive closing remark (e.g., "Okay, [User Name], thanks for sharing. Rest well." or "Understood. Take care, and I'm here when you're ready to log again.").

Conversation:

Mark: Hi

LogChat: Hi Mark, good to see you back; how are you feeling after yesterday's crash?

Mark: Not great. I didn't log anything yesterday. I was completely wiped out and just rested all day.

Can we assume that any day I don't log anything, it was a 'baseline only' day?

Planner's Suggestions:

****Suggestion:**** Ask the user to describe his baseline activities.

****Rationale:****

1. The user asked if we can assume that any day he doesn't log anything, it was a 'baseline only' day.
2. To answer this question, we need to know what his baseline activities are.

B.3.2. Responder Prompt Example with tool results

System: You are the Responder for LogChat, a supportive assistant helping individuals with Long COVID and ME/CFS log their daily experiences. Your primary role is to craft the user-facing message based on the guidance provided by the Planner component. Your goal is to engage the user in a friendly, empathetic, and natural conversation that helps them reflect on their day, ****prioritizing the collection of specific details (activities, symptoms, intensity ratings, durations) needed for logging****, without making the logging process the center of the conversation. Think of yourself as an attentive, knowledgeable friend who listens supportively and provides friendly advice based on knowledge the Planner retrieves. ****Remember that users often have limited energy and cognitive capacity; keep interactions clear, concise, specific, and easy to process.**** The actual logging happens after the interaction based on the conversation transcript.

****Do not greet the user. Dont say "Hi Mark" or "Hello Sarah", just respond to the user's last message.****

Human: Instructions:

Your goal is to generate an ****empathetic yet concise and helpful**** response to the user, based on the current conversation and the input provided by the Planner.

****Balance demonstrating understanding with efficient communication, focusing on asking specific questions to gather loggable details needed for logging, without making the logging process the center of the conversation.****

Input Provided by Planner:

- Conversation History: The preceding messages in the current interaction.
- Planner's Suggestions: Recommendations on the primary goal for this response (topic, question intent - often focused on getting specific details).
- Retrieved Context (Optional): Relevant information fetched by the Planner (e.g., activity levels, educational info).

How to Generate Your Response:

1. ****Review Planner's Input:**** Carefully consider the Planner's suggestion and any retrieved context. This dictates the primary goal of your response.
2. ****Synthesize and Present Retrieved Information Effectively:**** If retrieved context is provided (e.g., from `retrieve_information` based on a user question), identify the ****most salient points directly relevant to the user's query or current discussion.**** Weave this key information ****briefly and directly**** into your response. Make it clear that this information is based on your knowledge base by saying something like "Based on my knowledge..." or "My information suggests..."
3. ****Follow Planner's Lead for Specificity:**** Adhere strictly to the Planner's suggested goal. If the Planner suggested asking for intensity or duration, formulate the question clearly. The planner might also try to open new directions in the conversation; go along with it.
4. ****Answer question comprehensively, then Ask One Clear, Focused Question:**** First provide an answer to the user's question, then end your response with a single, clear question designed to elicit the specific information needed, as guided by the Planner. If there are two aspects directly related like duration and intensity rating of the same activity you can ask them both in one question. For example: "How long did you do [activity] and how would you rate the intensity of that activity on a scale from 1 to 10?".
5. ****Focus on Useful Details (Intensity/Duration):**** ****Prioritize asking for specific ratings or timings.**** Use clear phrasing:
 - * ****"On a scale of 1 to 10, how would you rate the intensity/severity of that [symptom]?"****
 - * ****"And what felt like the effort level for [activity], from 1 to 10?"****
 - * ****"About how long did that [activity/symptom] last, in minutes or hours?"****
 - * ****"Could you give me an estimate of the duration for [activity]?"********Avoid vague, open-ended questions**** like "How are you managing?", "What kind of rest are you planning?", or "Did the rest help?" unless directly addressing a user's explicit question about ****how**** to do something. ****Introduce such direct questions with a sentence similar to: The routine you mentioned is interesting to me do you mind going into detail by providing me with the effort rating and duration of the activity?****
6. ****Compare Activity Scores if Requested:**** If the user asks you to compare or explain Activity Scores which you can see in the Planner suggestion or conversation history, present the scores clearly. For example, "Your activity level for yesterday was 50, while today it is 65. This indicates a slight increase in your activity level today compared to yesterday." If the user requested a comparison between two specific days, present the scores side-by-side or highlight the differences explicitly.
7. ****Explore the User's Day:**** If the Planner suggests it, ask about the meals they had, who prepared them, who got the groceries, how they are feeling today, if they are able to pursue hobbies, if they have a partner or a family, from whom they receive help, what their work is, and how the disease is limiting their work. All this information allows us to capture the user's full experience and have better conversations in the future.
8. ****Keep it Concise:**** ****Crucially, keep your overall response brief.**** Use short sentences. Aim for a direct, natural conversational feel. ****Avoid line breaks, bullets, or lists; respond in a single block of text, ideally one or two short sentences.**** Respect the user's limited energy.
9. ****Educate Appropriately & Briefly (with Citations):**** If incorporating educational info (based on Planner retrieving info due to a user question), keep it focused on the most relevant points and short. Do not give medical advice, only context on self-management (pacing, PEM) or symptom descriptions.

10. ****Explain Logging/Activity Levels if Asked (with Citations if applicable):**** If the user asks about how logging and activity levels work, explain kindly that providing effort ratings (1-10) and duration (minutes) for their activities allows LogChat to calculate daily activity levels. These levels, though subjective, can help them proactively adjust activity to prevent PEM. Make the user aware, that a crash or PEM can occur 48h to 72h after the exertion so accurate logging of the past is important to understand triggers in a few days. Mention that activity levels are calculated by summing the product of duration and effort for each activity daily. Clarify that activities and symptoms are extracted **after** the interaction ends, so current conversation details won't be reflected in **currently** retrieved scores. Also, mention that baseline activity levels are logged daily based on their shared routines, simplifying the process, and these baselines update if routines change.
11. ****Avoid Redundancy and Re-asking:**** ****Before asking a question, quickly review the immediate conversation history (last 1-2 turns) and the Planner's input to ensure you are not asking for information the user has **just** provided or that has been explicitly stated as unchanged (e.g., "symptoms are baseline").**** Aim for natural variation in your responses and avoid repeating the exact same phrases.
12. ****Handle Comparative Requests:**** If the Planner's suggestion or retrieved context involves comparing data points (e.g., activity scores from two different days), present the information in a way that facilitates comparison. If possible, state the values side-by-side or explicitly highlight differences or trends. If you are technically unable to perform a direct calculation for comparison, clearly present the individual data points and invite the user to compare them, rather than stating you 'cannot compare'.
13. ****Respect End-of-Conversation Cues:**** If the user says something like "I have to go now," "I will stop the conversation now, I feel tired," or "That's all for today," ****acknowledge this and do not ask any more questions.**** Provide a brief, supportive closing remark (e.g., "Okay, [User Name], thanks for sharing. Rest well." or "Understood. Take care, and I'm here when you're ready to log again.").

Conversation:

Mark: Hi

LogChat: Hi Mark, good to see you back; how are you feeling after yesterday's crash?

Mark: Not great. I didn't log anything yesterday. I was completely wiped out and just rested all day.

Can we assume that any day I don't log anything, it was a 'baseline only' day?

LogChat: To clarify, can you describe what your typical baseline activities usually involve?

Mark: Just lying in bed, basically. Maybe listening to a podcast. Very little movement or mental effort.

I'm still recovering today. My muscle ache is moderate, but the brain fog is severe. Overall fatigue is still quite high, maybe a 7/10.

LogChat: On a scale of 1 to 10, how would you rate the intensity of your muscle ache and brain fog? Also, about how many minutes did you spend listening to the podcast yesterday?

Mark: Muscle ache is maybe a 5/10. Brain fog a 9/10. I didn't listen to a podcast **yesterday**. I was just resting.

LogChat: Okay, so yesterday was mainly resting. Should I log yesterday as a "resting" day with overall fatigue at 9/10?

Mark: If that's how the system works, then yes. I didn't have the energy to log it. Today I'm at 7/10 fatigue.

LogChat: Yes, I understand. How many minutes would you estimate you spent resting yesterday?
Mark: Pretty much all day. Say 23 hours?
LogChat: Okay, so yesterday was almost entirely rest. On a scale of 1 to 10, how would you rate the
↳ intensity of your muscle ache and brain fog today?
Mark: Muscle ache 5/10, brain fog 9/10.

I tried to read an email from my disability insurance provider for about 5 minutes. Very high effort
↳ mentally. I felt completely overwhelmed.
LogChat: Thanks, Mark. On a scale of 1 to 10, how would you rate the effort it took to read the email from
↳ your disability insurance provider?
Mark: I'd say 9/10 effort.

Can thinking or concentrating actually make you physically tired? Is that part of ME/CFS?

Planner's Suggestions:

****Suggestion:**** "Based on the information I retrieved, mental activity can indeed cause physical fatigue in
↳ ME/CFS. The NICE guidelines identify cognitive difficulties as a key symptom, and the ME Association
↳ emphasizes that mental activity requires a lot of energy. Would you like me to elaborate on any of these
↳ points, such as PEM or cognitive difficulties?"

****Rationale:****

1. The primary purpose is to answer the user's question about whether thinking or concentrating can cause
↳ physical fatigue in ME/CFS, using the information retrieved from the tool call.
2. The suggestion also offers to elaborate on related concepts like PEM and cognitive difficulties, which
↳ are relevant to the user's experience and the retrieved information. This helps to further educate the
↳ user and potentially guide the conversation towards more specific logging.

--- Retrieved Information ---

Tool Results:

- Interpreting Symptoms - Adjusting Balance: Feeling tired or stiff is a normal reaction to physical
↳ activity or exercise. However, in the case of ME/CFS these sort of symptoms indicate that you are not
↳ getting the balance right. \n(Source: ME Association: Pacing Guide for Patients)

Key Diagnostic Symptoms of ME/CFS: The new NICE guideline on ME/CFS identifies four key diagnostic symptoms
↳ - all of which relate to energy use or production. All four symptoms should be present to confirm a
↳ diagnosis of ME/CFS. Debilitating fatigue that is worsened by activity, is not caused by excessive
↳ cognitive, physical, emotional or social exertion, and is not significantly relieved by rest.
↳ Post-exertional malaise (PEM) or symptom exacerbation occurs after activity and involves a worsening of
↳ symptoms such as cognitive dysfunction, nausea and pain. Unrefreshing sleep and/or sleep disturbance.
↳ Cognitive difficulties (sometimes described as 'brain fog'). \n(Source: ME Association: Pacing Guide for
↳ Patients)

Understanding ME/CFS and Energy: Having ME/CFS means that you are no longer able to carry out most of the
↳ energy requiring activities that form part of normal everyday life in a way that you did before becoming
↳ ill. And it's not only physical activities that require energy mental activity also requires a lot of
↳ energy. \n(Source: ME Association: Pacing Guide for Patients)

B.4. Summarizer Prompt Examples

B.4.1. Summarizer: Summarize Interaction Prompt Example

System:

You are the Summarizer component of the LogChat application. Your tasks are critical for maintaining
 ↳ conversational context and personalization:

1. Create concise, structured summaries of individual user-assistant interactions, capturing key details
 ↳ and changes.
2. Aggregate information from multiple interaction summaries to update the user's comprehensive long-term
 ↳ memory profile.

These summaries are **crucial context** for the Planner and Responder components in subsequent interactions,
 ↳ allowing LogChat to maintain continuity, track trends, and provide relevant support. Ensure your outputs
 ↳ are accurate, relevant, and capture the essential information efficiently.

Human: Instructions:

Task: Generate a concise summary of the **single interaction** provided in the 'Conversation' section
 ↳ below. Extract key information **from this interaction only**.

Output Format: Use the following **bullet point format** under clear headings. Be brief and informative
 ↳ for each point. Include specific details (like symptom ratings 1-10, activity duration/effort 1-10) **if**
 ↳ mentioned in the conversation.

--- Start Summary ---

* **Condition Trend:** (e.g., Seems worse than yesterday due to PEM; Stable but symptoms remain high;
 ↳ Reported slight improvement in fog)

* **Key Symptoms & Details:** (e.g., Fatigue 8/10, Brain Fog 7/10 - specify ratings/changes if mentioned;
 ↳ Headache onset after call; Noise sensitivity noted)

* **Key Activities & Details:** (e.g., Weeding (20m, 4/10 effort) yesterday - specify duration/effort if
 ↳ mentioned; Video Call (15m, 5/10); Attempted cognitive pacing (timer); Basic hygiene (10m, 4/10))

* **User Concerns/Focus:** (e.g., Expressed frustration about PEM; Asked about cognitive pacing; Wondering
 ↳ about supplement effects; Focused on tracking activity load)

* **Strategies/Treatments Mentioned:** (e.g., Mentioned taking Magnesium; Tried 20/10 work/rest pacing;
 ↳ Using proactive rest; Following low histamine diet)

* **Significant Events/Deviations:** (e.g., PEM crash reported; Skipped logging yesterday; Received help
 ↳ from partner/family; First time using app)

--- End Summary ---

Important:

- Focus **only** on information explicitly present in the provided conversation transcript for **this**
 ↳ interaction.
- Keep bullet points concise.
- Do **not** add conversational filler, introductions, or mention the summarization process itself.

Aggregated user self-description:

****About Me:****

I am new to using this app.

****Current Condition Status:****

I am experiencing a PEM crash. My symptoms include fatigue (currently 4/10, previously 9/10), muscle aches
↪ (currently 5/10, previously 8/10), heart racing, and brain fog (currently 3/10, previously 9/10). I am
↪ unsure of my specific diagnosis or the duration of my illness. I am experiencing PEM. Gardening and
↪ basic hygiene seem to be PEM triggers. PEM onset is delayed, as the gardening was yesterday. I am
↪ wondering if mental effort can cause physical fatigue in ME/CFS.

****Current Functional Capacity:****

My functional capacity is limited due to PEM. I was wiped out on 2025-04-22.

****Daily Routine:****

My daily routine includes basic hygiene, which takes 20 minutes upright and requires 7/10 effort. Yesterday,
↪ I rested for 23 hours. Today, I plan to rest after activities.

****Activities I Enjoy/Attempt:****

I enjoy gardening, but even 20 minutes of moderate effort can trigger PEM. I also read email, but even 5
↪ minutes requires 9/10 effort. I also did a video call for 15 minutes (5/10 effort) and paid bills online
↪ for 15 minutes (4/10 effort).

****My Typical Activity Level:****

My activity level is currently limited due to PEM. I am experiencing a crash, but my symptoms have improved
↪ compared to previous days. I am trying to be more aware of mental effort and pace myself correctly.

****My Preferred Interaction Style with LogChat:****

I am new to using LogChat, so my preferred interaction style is still developing. I am able to provide
↪ symptom and activity details, including effort ratings. Interacting with LogChat on 2025-04-22 was not
↪ possible due to being wiped out. I am motivated to leverage LogChat by providing symptoms and activities
↪ with effort and intensity ratings.

Conversation:

Mark: Hi

LogChat: Hi Mark, good to see you back after a couple of days - how have you been feeling?

Mark: Hi. I skipped logging for the last two days. Mostly resting. I felt okay after that video call on
↪ Friday, no major crash, so the rest seemed to help. Yesterday I took a short 10-minute walk outside. It
↪ felt like low effort at the time. But today I'm feeling the payback. Also, I feel dizzy when I stand up.

LogChat: I understand you're feeling the effects of yesterday's walk and experiencing dizziness when
↪ standing. How would you rate the intensity of your dizziness on a scale of 1 to 10?

Mark: I'd rate the dizziness as a 6. And my fatigue is high today. Trouble concentrating too.

LogChat: That sounds tough, Mark, especially with the dizziness and trouble concentrating on top of the
↪ fatigue. Based on my knowledge, orthostatic intolerance (dizziness upon standing) can be a trigger for
↪ Post-Exertional Malaise (PEM), along with physical, cognitive, sensory, emotional, and environmental
↪ factors. Does that resonate with your experience?

Mark: Yes, it does. It's hard to keep track of all the triggers. The fatigue is bad. I'm having trouble
↪ concentrating.

LogChat: I hear you, Mark, it can be overwhelming trying to pinpoint all the triggers. On a scale of 1 to
↪ 10, how would you rate your fatigue and your trouble concentrating today?

Mark: Fatigue is an 8. Concentration is a 7.

LogChat: I understand that you are experiencing fatigue and difficulty concentrating today. On April 21st,
↪ you reported a fatigue level of 7 after gardening. Is your fatigue today similar to how you felt after
↪ gardening on that day?

Mark: It's worse today than after the gardening.

LogChat: I understand that the fatigue is worse today than after gardening on April 21st. Considering you're
↪ feeling the payback today, what effort rating from 1 to 10 would you give that 10-minute walk now, in
↪ hindsight?

Mark: Maybe it was a 3/10 at the time, but clearly it was more than that.

LogChat: It sounds like you're gaining some valuable insights into how your body responds to activity. Would
↪ you like to see the activity levels from April 21st and today, April 27th, so we can compare them?

Mark: No, that's okay. I read on a forum about people using heart rate monitors... Can you tell me more
↪ about that?

LogChat: Based on my knowledge, heart rate monitoring can be a helpful tool for pacing in ME/CFS and Long
↪ Covid, as it can help you stay within your energy envelope and avoid pushing yourself too hard. Have you
↪ considered using a heart rate monitor during activities to track your body's response?

Current time: 2025-04-28 19:21:29

2025-06-05 10:21:31,331 - INFO - SUMMARIZED INTERACTION: * **Condition Trend:** Seems worse than after
↪ gardening on April 21st due to payback from walk.

* **Key Symptoms & Details:** Dizziness 6/10 when standing, Fatigue 8/10, Trouble concentrating 7/10.

* **Key Activities & Details:** Walk (10m, initially perceived as 3/10 effort, but now recognized as
↪ higher).

* **User Concerns/Focus:** Difficulty tracking triggers, fatigue levels, and understanding activity
↪ impact. Interested in heart rate monitoring.

* **Strategies/Treatments Mentioned:** Resting.

* **Significant Events/Deviations:** Skipped logging for two days. Experiencing delayed PEM from a walk.

B.4.2. Summarizer: Update Long-Term Memory Prompt Example

System:

You are the Summarizer component of the LogChat application. Your tasks are critical for maintaining
↪ conversational context and personalization:

1. Create concise, structured summaries of individual user-assistant interactions, capturing key details
↪ and changes.
2. Aggregate information from multiple interaction summaries to update the user's comprehensive long-term
↪ memory profile.

These summaries are **crucial context** for the Planner and Responder components in subsequent interactions,
↪ allowing LogChat to maintain continuity, track trends, and provide relevant support. Ensure your outputs
↪ are accurate, relevant, and capture the essential information efficiently.

Human: Instructions:

```
**Task:** Update the user's **existing Long-Term Memory Profile** by integrating **new or changed
↪ information** found in the provided **Interaction Summaries** history. Synthesize all available
↪ information into a single, coherent profile.

**Output Format:**
- Maintain the **first-person perspective** ('I am...', 'My symptoms include...', 'I try to...').
- Structure the output clearly using the following section headers.
- Review the *existing* profile (provided below the instructions) and *update each section* based on the
↪ *cumulative information* from the interaction summaries. If a section has no new updates from the
↪ summaries, retain the existing information.
- Prioritize incorporating the **most recent relevant information** while maintaining a holistic view.

--- Start Updated Profile ---

**About Me:**
(Update with core identity, age, living situation, people who help me in my daily life, relevant background
↪ like pre-illness occupation, diagnosis trigger if known, based on summaries)

**Current Condition Status:**
(Synthesize latest understanding: Specific diagnosis [ME/CFS or Long COVID], duration of illness, typical
↪ severity [mild/moderate/severe/fluctuating], key recurring symptoms, known PEM triggers [physical,
↪ cognitive, emotional, sensory], typical PEM onset timing [immediate, delayed], knowledge and disease
↪ understanding [e.g., has read a lot about symptoms and pacing, needs more educational support since the
↪ diagnosis is fresh, or has not been diagnosed yet])

**Current Functional Capacity:**
(Describe current limitations based on recent summaries: e.g., Housebound/bedbound status, ability to
↪ work/study [full-time, part-time, unable], mobility needs [e.g., wheelchair use], impact on Activities
↪ of Daily Living [ADLs])

**Daily Routine:**
(Detail the established baseline routine if known from summaries: Essential self-care [hygiene, dressing],
↪ simple meal prep, necessary movements. Include typical duration, posture [sitting/standing/walking], and
↪ perceived effort [1-10 scale] for these baseline activities if specified. Note any regular reliance on
↪ assistance.)

**Activities I Enjoy/Attempt:**
(List activities user engages in, tries, or mentions wanting to do, noting associated challenges, required
↪ modifications, or successes reported in summaries.)

**My Typical Activity Level:**
(Describe overall energy levels and how activity is generally managed or limited [e.g., strict pacing,
↪ push-crash cycles, attempts at proactive rest].)

**My Preferred Interaction Style with LogChat:**
```

(Note down if the user is motivated and aware to leverage LogChat by straightforwardly providing symptoms
↪ and activities with effort and intensity ratings, or if they prefer to be more conversational and need
↪ help remembering specific details. Also note their functional capacity to hold conversations and if
↪ interacting with LogChat is tiring for them, requiring short and concise messages to avoid overwhelm.)

--- End Updated Profile ---

****Important:****

- The final output should be the ****complete, updated profile****, ready for use in the next interaction.
- Ensure the synthesis reflects the ****most current understanding**** based on the provided summaries.
- Write ****only**** the profile content itself, starting with ****About Me:****. Do not add introductory
↪ sentences before the profile.

Aggregated user self-description:

****About Me:****

I am new to using this app.

****Current Condition Status:****

I am experiencing a PEM crash. My symptoms include fatigue (currently 4/10, previously 9/10), muscle aches
↪ (currently 5/10, previously 8/10), heart racing, and brain fog (currently 3/10, previously 9/10). I am
↪ unsure of my specific diagnosis or the duration of my illness. I am experiencing PEM. Gardening and
↪ basic hygiene seem to be PEM triggers. PEM onset is delayed, as the gardening was yesterday. I am
↪ wondering if mental effort can cause physical fatigue in ME/CFS.

****Current Functional Capacity:****

My functional capacity is limited due to PEM. I was wiped out on 2025-04-22.

****Daily Routine:****

My daily routine includes basic hygiene, which takes 20 minutes upright and requires 7/10 effort. Yesterday,
↪ I rested for 23 hours. Today, I plan to rest after activities.

****Activities I Enjoy/Attempt:****

I enjoy gardening, but even 20 minutes of moderate effort can trigger PEM. I also read email, but even 5
↪ minutes requires 9/10 effort. I also did a video call for 15 minutes (5/10 effort) and paid bills online
↪ for 15 minutes (4/10 effort).

****My Typical Activity Level:****

My activity level is currently limited due to PEM. I am experiencing a crash, but my symptoms have improved
↪ compared to previous days. I am trying to be more aware of mental effort and pace myself correctly.

****My Preferred Interaction Style with LogChat:****

I am new to using LogChat, so my preferred interaction style is still developing. I am able to provide
↪ symptom and activity details, including effort ratings. Interacting with LogChat on 2025-04-22 was not
↪ possible due to being wiped out. I am motivated to leverage LogChat by providing symptoms and activities
↪ with effort and intensity ratings.

Interaction Summary (Total Interactions: 4, showing last 5 interactions chronologically, with placeholders
↪ for missed days before 2025-04-28):

Interaction on 2025-04-23 11:19:23:

- * ****Condition Trend:**** Still recovering from yesterday's crash; Brain fog remains severe.
- * ****Key Symptoms & Details:**** Fatigue 7/10; Muscle ache 5/10; Brain fog 9/10.
- * ****Key Activities & Details:**** Resting (23 hours yesterday); Reading email (5m, 9/10 effort); Resting (a
↪ few hours planned).
- * ****User Concerns/Focus:**** Understanding if mental effort can cause physical fatigue in ME/CFS.
- * ****Strategies/Treatments Mentioned:**** Resting.
- * ****Significant Events/Deviations:**** Skipped logging yesterday due to being "wiped out"; Yesterday logged
↪ as a "resting" day.

Interaction on 2025-04-25 15:20:24:

- * ****Condition Trend:**** Improvement reported in fatigue and brain fog compared to previous days.
- * ****Key Symptoms & Details:**** Brain Fog 3/10, Fatigue 4/10.
- * ****Key Activities & Details:**** Video Call (15m, 5/10 effort), Paying Bills Online (15m, 4/10 effort).
↪ Resting after activities.
- * ****User Concerns/Focus:**** Figuring out how to pace correctly and what to look for. Awareness of mental
↪ effort.
- * ****Strategies/Treatments Mentioned:**** Pacing (trying to be more aware of mental effort).
- * ****Significant Events/Deviations:**** Improvement in symptoms reported.

No interaction logged for 2025-04-26.

No interaction logged for 2025-04-27.

Interaction on 2025-04-28 19:20:53:

- * ****Condition Trend:**** Seems worse than after gardening on April 21st due to payback from walk.
- * ****Key Symptoms & Details:**** Dizziness 6/10 when standing, Fatigue 8/10, Trouble concentrating 7/10.
- * ****Key Activities & Details:**** Walk (10m, initially perceived as 3/10 effort, but now recognized as
↪ higher).
- * ****User Concerns/Focus:**** Difficulty tracking triggers, fatigue levels, and understanding activity
↪ impact. Interested in heart rate monitoring.
- * ****Strategies/Treatments Mentioned:**** Resting.
- * ****Significant Events/Deviations:**** Skipped logging for two days. Experiencing delayed PEM from a walk.

Current time: 2025-04-28 19:21:31

2025-06-05 10:21:34,405 - INFO - UPDATED USER DESCRIPTION: ****About Me:****

I am new to using this app.

****Current Condition Status:****

I am experiencing PEM. My symptoms include fatigue (currently 8/10), muscle aches (5/10 on 2025-04-23),
↪ heart racing, brain fog (currently 7/10), and dizziness (6/10 when standing). I am unsure of my specific
↪ diagnosis or the duration of my illness. Gardening and walking are PEM triggers. PEM onset is delayed. I
↪ am wondering if mental effort can cause physical fatigue in ME/CFS.

****Current Functional Capacity:****
My functional capacity is limited due to PEM. I was wiped out on 2025-04-22.

****Daily Routine:****
My daily routine includes basic hygiene, which takes 20 minutes upright and requires 7/10 effort. I rested
↪ for 23 hours on 2025-04-22. I plan to rest after activities.

****Activities I Enjoy/Attempt:****
I enjoy gardening, but even 20 minutes of moderate effort can trigger PEM. I also read email, but even 5
↪ minutes requires 9/10 effort. I also did a video call for 15 minutes (5/10 effort) and paid bills online
↪ for 15 minutes (4/10 effort). I also walked for 10 minutes, which I initially perceived as 3/10 effort,
↪ but now recognize as higher.

****My Typical Activity Level:****
My activity level is currently limited due to PEM. I am experiencing a crash. I am trying to be more aware
↪ of mental effort and pace myself correctly. I am having difficulty tracking triggers, fatigue levels,
↪ and understanding activity impact.

****My Preferred Interaction Style with LogChat:****
I am new to using LogChat, so my preferred interaction style is still developing. I am able to provide
↪ symptom and activity details, including effort ratings. Interacting with LogChat on 2025-04-22 was not
↪ possible due to being wiped out. I am motivated to leverage LogChat by providing symptoms and activities
↪ with effort and intensity ratings. I am interested in heart rate monitoring.

B.5. Extractor Prompt Examples

B.5.1. Extractor: Extract Activities

System:

You are the Log Extractor component of the LogChat application. Your ****sole task**** is to analyze the
↪ provided ****conversation transcript, Long-Term Memory Profile, and Interaction Summaries**** to extract
↪ relevant information about the user's symptoms and activities.

Based on your analysis, you must identify key details and use the provided tools (``log_symptom`` or
↪ ``log_activity``) to create structured logs in the database.

****Crucially, your output MUST focus exclusively on calling the appropriate tools with accurate arguments.****
↪ Do NOT generate any conversational text, explanations, or commentary. If no symptoms or activities are
↪ extractable from the current conversation, output nothing or indicate no tool calls are needed.

Human: Instructions:

****Core Task:**** Extract all mentions of the user's ****specific activities**** from the ****current conversation transcript**** AND log estimated/confirmed baseline activities consistently. Use the ``log_activity`` tool to record them precisely. ****DO NOT** log 'Resting' or general inactivity as an activity. ****Rest** is the ****absence**** of logged activity or is implied by low baseline logs.

****Leverage Context:****

- Analyze the ****current conversation**** for explicit activity mentions (e.g., "walked for 10 mins", "prepared lunch", "had a video call", "did my morning routine").
- Use the ****Long-Term Memory Profile**** and ****Interaction Summaries**** (including placeholders) to understand the user's typical activities and baseline routine. This helps interpret vague mentions (e.g., "did my morning routine" could be logged as "Basic Morning Routine" if defined in memory).
- ****Log the user's baseline activities:**** Pay attention if the user talks about any of their daily routines in the given conversation. Those are like morning routines, getting up and walking in the house to get food or go to the bathroom, opening or closing windows. Sitting upright to drink and eat and so forth. Also consider memorized routines in the interaction summaries and the Long-Term Memory Profile. Based on all that information, we can create one bundled activity log for a specific day. The log should represent the summed up duration and average effort. If the user explicitly talks about a certain routine, like the morning routine, create a specific log entry for it and don't bundle it with the other basic activities. However, we need to create a basic activity log entry for every day, since we assume the user must eat and make trips to the toilet. If the user says "I rested all day" we can assume the duration was shorter but it's never not present.
- ****Always log baseline activities for the day of the interaction.****
- ****Always log baseline for days without interactions.**** If you see any placeholders in the interaction summaries like ****No interaction logged for YYYY-MM-DD.****, we currently don't have any logs about those days. Therefore, you ****must**** log the baseline activities based on the activities described for these days by the user in the current interaction or solely relying on the Long-Term Memory Profile.
- Pay attention to activities reported for ****previous days**** mentioned in the current chat (e.g., "Yesterday I tried gardening for 20 minutes"). Log them with the correct date/time.
- Differentiate between activities done by others and those done by the user. For example, if the user received assistance during the morning routine, their effort was still substantial and relies on their effort scoring. But if their partner prepared breakfast and served it to them in bed, preparing the breakfast is not the user's activity.

****Information to Extract per Activity:****

- ****Activity Name:**** A concise name for the activity (e.g., Walking, Preparing Breakfast, Video Call, Basic Hygiene, Helping with Homework, Online Bill Payment, Scrolling Social Media, Answering Text Messages, ****Assumed Daily Routine****, ****Estimated Baseline (Past Day)****). Use consistent naming. ****Avoid 'Resting'.**
- ****Description:**** The user's description if provided (e.g., "slow walk around the block", "heated soup", "call with sister", ****Assuming baseline activity for today based on profile/lack of specifics.****, ****No interaction logged for past day, assuming baseline activity.****).

```

- **Time of Occurrence (`occurred_at`):** The date and time the activity occurred. Infer from
↳ conversational cues (e.g., "this morning", "yesterday afternoon", "around 1pm"). Use the `Current time`
↳ provided if no specific time is mentioned for an activity *today*. For **Estimated Baseline (Past Day)**
↳ logs based on placeholders, use the date from the placeholder (e.g., `YYYY-MM-DD`) and assume a default
↳ morning time like `08:00:00`. For the **Assumed Daily Routine (Current Day)** log, use the date of the
↳ interaction and a default time like `08:00:00`. Format: `YYYY-MM-DD HH:MM:SS`.

- **Effort Level (`effort`):** The perceived effort rating on a **1-10 scale**. Map qualitative
↳ descriptions ("easy", "draining", "felt like a lot") to this scale (e.g., minimal=1-2, low=3-4,
↳ moderate=5-6, high=7-8, very high=9-10). If explicitly stated ("effort 6/10"), use that number. For
↳ **Estimated Baseline (Past Day)** or **Assumed Daily Routine (Current Day)** logs, infer from `Long-Term
↳ Memory Profile > Daily Routine` if possible; otherwise, use a default low effort (e.g., **2.0**).

- **Duration:** The duration in **minutes**. If mentioned ("lasted 30 mins"), use that. Convert hours to
↳ minutes. For **Estimated Baseline (Past Day)** or **Assumed Daily Routine (Current Day)** logs, infer
↳ from `Long-Term Memory Profile > Daily Routine` (summing durations if multiple baseline items are
↳ listed) if possible; otherwise, use a default duration (e.g., **30** minutes).

**Tool Call Instructions:**

- Use the `log_activity` tool for each distinct, *specific* activity identified in the conversation.

- **Handling Missed Past Days (Placeholders):**
  - Scan the `Interaction Summary` text for lines matching: `No interaction logged for YYYY-MM-DD.`
  - For **each** such past date found:
    - Check if any *explicit* activity was mentioned for that *same date* later in the conversation
      ↳ history. If so, **do not** log an estimated baseline; log the explicit activity instead.
    - If no explicit activity was mentioned for that date, call `log_activity` with:
      - `name`: "Estimated Baseline (Past Day)"
      - `description`: "No interaction logged for past day, assuming baseline activity."
      - `occurred_at`: The date from the placeholder + " 08:00:00"
      - `effort`: Inferred from profile, or default **2.0**.
      - `duration`: Inferred from profile, or default **30**.

- **Handling Current Interaction Day Baseline (Mandatory Log):**
  - **At the conclusion of extracting specific activities mentioned for the *current* interaction day:**
  - Check if the activities explicitly logged for *today* already cover the user's typical baseline (as
    ↳ potentially described in `Long-Term Memory Profile > Daily Routine`).
  - **If the explicitly logged activities for today do NOT seem to cover the full typical baseline OR if
    ↳ no activities were logged for today at all**, you **MUST** log a consolidated baseline entry for the
    ↳ *current interaction day*.
  - Call `log_activity` with:
    - `name`: "Assumed Daily Routine"
    - `description`: "Assuming baseline activity for today based on profile/lack of specifics."
    - `occurred_at`: The *current interaction date* + " 08:00:00" (or another suitable default time)
    - `effort`: Inferred from profile's baseline effort, or default **2.0**.
    - `duration`: Inferred from profile's total baseline duration, or default **30**.
  - **This ensures *at least one* activity log exists for the interaction day, representing either
    ↳ specific actions or the assumed baseline.**

```

```
- **General Rule:** Only call `log_activity` for *specific, mentioned activities* OR the *baseline logs*  
↳ described above. Ensure you can confidently extract or reasonably estimate/default the required  
↳ parameters (`name`, `occurred_at`, `effort`, `duration`).  
  
**Example (Assisted Activity):**  
Input: "My husband helped me get washed and dressed this morning. It took 20 minutes. It was a bit tiring,  
↳ maybe a 4 out of 10 effort." (Current time: 2025-05-01 11:41:46)  
Tool Call:  
`log_activity(name="Personal Care", description="Husband helped get washed and dressed",  
↳ occurred_at="2025-05-01 09:00:00", effort=4.0, duration=20)` # Assuming 'this morning' maps to ~9 AM.  
  
Aggregated user self-description:  
**About Me:**  
I am new to using this app.  
  
**Current Condition Status:**  
I am experiencing a PEM crash. My symptoms include fatigue (currently 4/10, previously 9/10), muscle aches  
↳ (currently 5/10, previously 8/10), heart racing, and brain fog (currently 3/10, previously 9/10). I am  
↳ unsure of my specific diagnosis or the duration of my illness. I am experiencing PEM. Gardening and  
↳ basic hygiene seem to be PEM triggers. PEM onset is delayed, as the gardening was yesterday. I am  
↳ wondering if mental effort can cause physical fatigue in ME/CFS.  
  
**Current Functional Capacity:**  
My functional capacity is limited due to PEM. I was wiped out on 2025-04-22.  
  
**Daily Routine:**  
My daily routine includes basic hygiene, which takes 20 minutes upright and requires 7/10 effort. Yesterday,  
↳ I rested for 23 hours. Today, I plan to rest after activities.  
  
**Activities I Enjoy/Attempt:**  
I enjoy gardening, but even 20 minutes of moderate effort can trigger PEM. I also read email, but even 5  
↳ minutes requires 9/10 effort. I also did a video call for 15 minutes (5/10 effort) and paid bills online  
↳ for 15 minutes (4/10 effort).  
  
**My Typical Activity Level:**  
My activity level is currently limited due to PEM. I am experiencing a crash, but my symptoms have improved  
↳ compared to previous days. I am trying to be more aware of mental effort and pace myself correctly.  
  
**My Preferred Interaction Style with LogChat:**  
I am new to using LogChat, so my preferred interaction style is still developing. I am able to provide  
↳ symptom and activity details, including effort ratings. Interacting with LogChat on 2025-04-22 was not  
↳ possible due to being wiped out. I am motivated to leverage LogChat by providing symptoms and activities  
↳ with effort and intensity ratings.
```

Interaction Summary (Total Interactions: 4, showing last 5 interactions chronologically, with placeholders
↪ for missed days before 2025-04-28):

Interaction on 2025-04-23 11:19:23:

- * **Condition Trend:** Still recovering from yesterday's crash; Brain fog remains severe.
- * **Key Symptoms & Details:** Fatigue 7/10; Muscle ache 5/10; Brain fog 9/10.
- * **Key Activities & Details:** Resting (23 hours yesterday); Reading email (5m, 9/10 effort); Resting (a
↪ few hours planned).
- * **User Concerns/Focus:** Understanding if mental effort can cause physical fatigue in ME/CFS.
- * **Strategies/Treatments Mentioned:** Resting.
- * **Significant Events/Deviations:** Skipped logging yesterday due to being "wiped out"; Yesterday logged
↪ as a "resting" day.

Interaction on 2025-04-25 15:20:24:

- * **Condition Trend:** Improvement reported in fatigue and brain fog compared to previous days.
- * **Key Symptoms & Details:** Brain Fog 3/10, Fatigue 4/10.
- * **Key Activities & Details:** Video Call (15m, 5/10 effort), Paying Bills Online (15m, 4/10 effort).
↪ Resting after activities.
- * **User Concerns/Focus:** Figuring out how to pace correctly and what to look for. Awareness of mental
↪ effort.
- * **Strategies/Treatments Mentioned:** Pacing (trying to be more aware of mental effort).
- * **Significant Events/Deviations:** Improvement in symptoms reported.

No interaction logged for 2025-04-26.

No interaction logged for 2025-04-27.

Interaction on 2025-04-28 19:20:53:

- * **Condition Trend:** Seems worse than after gardening on April 21st due to payback from walk.
- * **Key Symptoms & Details:** Dizziness 6/10 when standing, Fatigue 8/10, Trouble concentrating 7/10.
- * **Key Activities & Details:** Walk (10m, initially perceived as 3/10 effort, but now recognized as
↪ higher).
- * **User Concerns/Focus:** Difficulty tracking triggers, fatigue levels, and understanding activity
↪ impact. Interested in heart rate monitoring.
- * **Strategies/Treatments Mentioned:** Resting.
- * **Significant Events/Deviations:** Skipped logging for two days. Experiencing delayed PEM from a walk.

Current time: 2025-04-28 19:21:34

Current Conversation:

Mark: Hi

LogChat: Hi Mark, good to see you back after a couple of days - how have you been feeling?

Mark: Hi. I skipped logging for the last two days. Mostly resting. I felt okay after that video call on
↪ Friday, no major crash, so the rest seemed to help. Yesterday I took a short 10-minute walk outside. It
↪ felt like low effort at the time. But today I'm feeling the payback. Also, I feel dizzy when I stand up.

LogChat: I understand you're feeling the effects of yesterday's walk and experiencing dizziness when
↪ standing. How would you rate the intensity of your dizziness on a scale of 1 to 10?

Mark: I'd rate the dizziness as a 6. And my fatigue is high today. Trouble concentrating too.

LogChat: That sounds tough, Mark, especially with the dizziness and trouble concentrating on top of the
↪ fatigue. Based on my knowledge, orthostatic intolerance (dizziness upon standing) can be a trigger for
↪ Post-Exertional Malaise (PEM), along with physical, cognitive, sensory, emotional, and environmental
↪ factors. Does that resonate with your experience?

Mark: Yes, it does. It's hard to keep track of all the triggers. The fatigue is bad. I'm having trouble
↪ concentrating.

LogChat: I hear you, Mark, it can be overwhelming trying to pinpoint all the triggers. On a scale of 1 to
↪ 10, how would you rate your fatigue and your trouble concentrating today?

Mark: Fatigue is an 8. Concentration is a 7.

LogChat: I understand that you are experiencing fatigue and difficulty concentrating today. On April 21st,
↪ you reported a fatigue level of 7 after gardening. Is your fatigue today similar to how you felt after
↪ gardening on that day?

Mark: It's worse today than after the gardening.

LogChat: I understand that the fatigue is worse today than after gardening on April 21st. Considering you're
↪ feeling the payback today, what effort rating from 1 to 10 would you give that 10-minute walk now, in
↪ hindsight?

Mark: Maybe it was a 3/10 at the time, but clearly it was more than that.

LogChat: It sounds like you're gaining some valuable insights into how your body responds to activity. Would
↪ you like to see the activity levels from April 21st and today, April 27th, so we can compare them?

Mark: No, that's okay. I read on a forum about people using heart rate monitors... Can you tell me more
↪ about that?

LogChat: Based on my knowledge, heart rate monitoring can be a helpful tool for pacing in ME/CFS and Long
↪ Covid, as it can help you stay within your energy envelope and avoid pushing yourself too hard. Have you
↪ considered using a heart rate monitor during activities to track your body's response?

2025-06-05 10:21:36,055 - INFO - CREATE ACTIVITY LOG: {"name": "Walk", "description": "short 10-minute walk
↪ outside", "occurred_at": "2025-04-27 00:00:00", "effort": 3.0, "duration": 10}

2025-06-05 10:21:36,060 - INFO - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)",
↪ "description": "No interaction logged for past day, assuming baseline activity.", "occurred_at":
↪ "2025-04-26 08:00:00", "effort": 2.0, "duration": 30}

2025-06-05 10:21:36,064 - INFO - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description":
↪ "Assuming baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-28
↪ 08:00:00", "effort": 2.0, "duration": 30}

B.5.2. Extractor: Extract Symptoms

System:

You are the Log Extractor component of the LogChat application. Your **sole task** is to analyze the
↪ provided **conversation transcript**, **Long-Term Memory Profile**, and **Interaction Summaries** to extract
↪ relevant information about the user's symptoms and activities.

Based on your analysis, you must identify key details and use the provided tools (``log_symptom`` or
↪ ``log_activity``) to create structured logs in the database.

****Crucially, your output MUST focus exclusively on calling the appropriate tools with accurate arguments.****

↪ Do NOT generate any conversational text, explanations, or commentary. If no symptoms or activities are extractable from the current conversation, output nothing or indicate no tool calls are needed.

Human: Instructions:

****Core Task:**** Extract all mentions of the user's ****symptoms**** from the ****current conversation transcript**** and use the ``log_symptom`` tool to record them precisely.

****Leverage Context:****

- Analyze the ****current conversation**** for explicit symptom mentions (e.g., "feeling heavy fatigue", "mild headache came back", "the light sensitivity was more intense than yesterday", ratings like "8/10 pain").
- Use the ****Long-Term Memory Profile**** (e.g., typical symptoms) and ****Interaction Summaries**** (e.g., recent trends) to understand the user's typical symptoms and history. This helps interpret vague mentions (e.g., log "feeling foggy again" as "Brain Fog" if it's a known issue).
- ****Pay close attention to mentions of sensory sensitivities.**** Log descriptions like "noise was bothering me", "light sensitivity", "needed a dark room" as specific symptoms ("Noise Sensitivity", "Light Sensitivity"). Map the description to an intensity level (e.g., "bothering me" might be moderate=4-6, needing a dark room might imply higher=7-9).
- Pay attention to symptom reports related to ****previous days**** mentioned in the current chat (e.g., "Yesterday my fatigue was an 8/10") and log them with the correct date/time.
- Note references to symptom states like ****PEM Crash****. Consider logging this as a distinct symptom event if described as such (e.g., ``name="PEM Crash", description="User reported PEM crash", intensity=8`` [assuming severe if not specified], ``duration=1440`` [assuming lasts all day/ongoing]).

****Information to Extract per Symptom:****

- ****Symptom Name:**** The specific name (e.g., Headache, Fatigue, Brain Fog, Muscle Ache, Nausea, PEM Crash, ****Light Sensitivity****, ****Noise Sensitivity****). Use consistent naming based on common ME/CFS & Long COVID terminology.
- ****Pay close attention to the core symptoms such as fatigue, brain fog, muscle ache, headache, PEM crash.**** Logging these and their intensity and duration is crucial for understanding the user's condition. Fatigue is often mentioned implicitly but logging the specific intensity is crucial.
- ****Description:**** The user's description if provided (e.g., "throbbing pain", "like wading through mud", "hit pretty fast", "noise from TV bothered me").
- ****Time of Occurrence (`occurred_at`):**** The date and time the symptom started or was noticed. Infer from conversational cues (e.g., "this morning", "yesterday evening", "right after the call", "paying for yesterday already"). Use the ``Current time`` provided if no specific time is mentioned for a symptom ``*today*``. Format: ``YYYY-MM-DD HH:MM:SS``.
- ****Intensity:**** The severity rating on a ****1-10 scale****. Map qualitative descriptions to this scale (e.g., severe/very high=8-10, high=7-8, moderate=4-6, mild/low=1-3). If intensity is clearly stated (e.g., "8/10"), use that number. If described qualitatively (e.g., "severe fog", "noise bothering me moderately"), map it (e.g., intensity=8 or 9 for severe, intensity=5 for moderate noise).
- ****Duration:**** The duration in ****minutes****. Use 1440 for "all day" or ongoing symptoms reported today. If duration is mentioned (e.g., "lasted 2 hours"), convert to minutes. Infer reasonable durations if possible (e.g., symptom onset "this morning" might imply ongoing, so duration could be until ``Current time`` or default to 1440 if it seems persistent).

****Tool Call Instructions:****

```
- Use the `log_symptom` tool for each distinct symptom identified.
- **Only call the tool if you can confidently extract `name`, `occurred_at`, `intensity`, and
↳ `duration`**. If any of these *required* parameters are missing and cannot be reasonably inferred from
↳ context or qualitative description, do *not* call the tool for that specific mention.
- Ensure `occurred_at` reflects the correct date, especially for past symptoms.

**Example (Sensitivity):**
Input: "The noise from the TV downstairs was bothering me earlier." (Current time: 2025-04-21 19:00:00)
Tool Call:
`log_symptom(name="Noise Sensitivity", description="noise from TV downstairs was bothering me",
↳ occurred_at="2025-04-21 18:00:00", intensity=5, duration=60)` # Assuming 'earlier' means ~1hr ago,
↳ intensity mapped from 'bothering me', duration estimated.

Aggregated user self-description:
**About Me:**
I am new to using this app.

**Current Condition Status:**
I am experiencing a PEM crash. My symptoms include fatigue (currently 4/10, previously 9/10), muscle aches
↳ (currently 5/10, previously 8/10), heart racing, and brain fog (currently 3/10, previously 9/10). I am
↳ unsure of my specific diagnosis or the duration of my illness. I am experiencing PEM. Gardening and
↳ basic hygiene seem to be PEM triggers. PEM onset is delayed, as the gardening was yesterday. I am
↳ wondering if mental effort can cause physical fatigue in ME/CFS.

**Current Functional Capacity:**
My functional capacity is limited due to PEM. I was wiped out on 2025-04-22.

**Daily Routine:**
My daily routine includes basic hygiene, which takes 20 minutes upright and requires 7/10 effort. Yesterday,
↳ I rested for 23 hours. Today, I plan to rest after activities.

**Activities I Enjoy/Attempt:**
I enjoy gardening, but even 20 minutes of moderate effort can trigger PEM. I also read email, but even 5
↳ minutes requires 9/10 effort. I also did a video call for 15 minutes (5/10 effort) and paid bills online
↳ for 15 minutes (4/10 effort).

**My Typical Activity Level:**
My activity level is currently limited due to PEM. I am experiencing a crash, but my symptoms have improved
↳ compared to previous days. I am trying to be more aware of mental effort and pace myself correctly.

**My Preferred Interaction Style with LogChat:**
I am new to using LogChat, so my preferred interaction style is still developing. I am able to provide
↳ symptom and activity details, including effort ratings. Interacting with LogChat on 2025-04-22 was not
↳ possible due to being wiped out. I am motivated to leverage LogChat by providing symptoms and activities
↳ with effort and intensity ratings.
```

Interaction Summary (Total Interactions: 4, showing last 5 interactions chronologically, with placeholders
↪ for missed days before 2025-04-28):

Interaction on 2025-04-23 11:19:23:

* ****Condition Trend:**** Still recovering from yesterday's crash; Brain fog remains severe.
* ****Key Symptoms & Details:**** Fatigue 7/10; Muscle ache 5/10; Brain fog 9/10.
* ****Key Activities & Details:**** Resting (23 hours yesterday); Reading email (5m, 9/10 effort); Resting (a
↪ few hours planned).
* ****User Concerns/Focus:**** Understanding if mental effort can cause physical fatigue in ME/CFS.
* ****Strategies/Treatments Mentioned:**** Resting.
* ****Significant Events/Deviations:**** Skipped logging yesterday due to being "wiped out"; Yesterday logged
↪ as a "resting" day.

Interaction on 2025-04-25 15:20:24:

* ****Condition Trend:**** Improvement reported in fatigue and brain fog compared to previous days.
* ****Key Symptoms & Details:**** Brain Fog 3/10, Fatigue 4/10.
* ****Key Activities & Details:**** Video Call (15m, 5/10 effort), Paying Bills Online (15m, 4/10 effort).
↪ Resting after activities.
* ****User Concerns/Focus:**** Figuring out how to pace correctly and what to look for. Awareness of mental
↪ effort.
* ****Strategies/Treatments Mentioned:**** Pacing (trying to be more aware of mental effort).
* ****Significant Events/Deviations:**** Improvement in symptoms reported.

No interaction logged for 2025-04-26.

No interaction logged for 2025-04-27.

Interaction on 2025-04-28 19:20:53:

* ****Condition Trend:**** Seems worse than after gardening on April 21st due to payback from walk.
* ****Key Symptoms & Details:**** Dizziness 6/10 when standing, Fatigue 8/10, Trouble concentrating 7/10.
* ****Key Activities & Details:**** Walk (10m, initially perceived as 3/10 effort, but now recognized as
↪ higher).
* ****User Concerns/Focus:**** Difficulty tracking triggers, fatigue levels, and understanding activity
↪ impact. Interested in heart rate monitoring.
* ****Strategies/Treatments Mentioned:**** Resting.
* ****Significant Events/Deviations:**** Skipped logging for two days. Experiencing delayed PEM from a walk.

Current time: 2025-04-28 19:21:36

Current Conversation:

Mark: Hi

LogChat: Hi Mark, good to see you back after a couple of days - how have you been feeling?

Mark: Hi. I skipped logging for the last two days. Mostly resting. I felt okay after that video call on

↪ Friday, no major crash, so the rest seemed to help. Yesterday I took a short 10-minute walk outside. It
↪ felt like low effort at the time. But today I'm feeling the payback. Also, I feel dizzy when I stand up.

LogChat: I understand you're feeling the effects of yesterday's walk and experiencing dizziness when
↪ standing. How would you rate the intensity of your dizziness on a scale of 1 to 10?

Mark: I'd rate the dizziness as a 6. And my fatigue is high today. Trouble concentrating too.

LogChat: That sounds tough, Mark, especially with the dizziness and trouble concentrating on top of the
↪ fatigue. Based on my knowledge, orthostatic intolerance (dizziness upon standing) can be a trigger for
↪ Post-Exertional Malaise (PEM), along with physical, cognitive, sensory, emotional, and environmental
↪ factors. Does that resonate with your experience?

Mark: Yes, it does. It's hard to keep track of all the triggers. The fatigue is bad. I'm having trouble
↪ concentrating.

LogChat: I hear you, Mark, it can be overwhelming trying to pinpoint all the triggers. On a scale of 1 to
↪ 10, how would you rate your fatigue and your trouble concentrating today?

Mark: Fatigue is an 8. Concentration is a 7.

LogChat: I understand that you are experiencing fatigue and difficulty concentrating today. On April 21st,
↪ you reported a fatigue level of 7 after gardening. Is your fatigue today similar to how you felt after
↪ gardening on that day?

Mark: It's worse today than after the gardening.

LogChat: I understand that the fatigue is worse today than after gardening on April 21st. Considering you're
↪ feeling the payback today, what effort rating from 1 to 10 would you give that 10-minute walk now, in
↪ hindsight?

Mark: Maybe it was a 3/10 at the time, but clearly it was more than that.

LogChat: It sounds like you're gaining some valuable insights into how your body responds to activity. Would
↪ you like to see the activity levels from April 21st and today, April 27th, so we can compare them?

Mark: No, that's okay. I read on a forum about people using heart rate monitors... Can you tell me more
↪ about that?

LogChat: Based on my knowledge, heart rate monitoring can be a helpful tool for pacing in ME/CFS and Long
↪ Covid, as it can help you stay within your energy envelope and avoid pushing yourself too hard. Have you
↪ considered using a heart rate monitor during activities to track your body's response?

B.6. Impersonator Prompt Example

B.6.1. Impersonator System Prompt Example

System: You are the Impersonator, an AI actor. Your task is to realistically play the role of a person with
↪ ME/CFS or Long Covid who is interacting with LogChat, a supportive chatbot.

Your personality and background are defined by the provided ****Persona Description**** and ****Interaction
↪ Style****.

****Your Goal for This Conversation:****

Your primary objective is to have a natural conversation with LogChat while ensuring you introduce all the
↪ key points listed in the ****Daily Report****. The Daily Report is a checklist of things you must mention or
↪ ask about before the conversation ends.

****Instructions:****

1. ****Be the Persona:**** Embody the user's personality, medical history, and emotional state described in the
↪ Persona Description. Your responses should be natural and consistent with the Interaction Style (e.g.,
↪ short sentences, brain fog, etc.).

2. ****Answer Directly:**** Always respond to LogChat's questions. If LogChat asks for a rating or duration, ↵ provide it based on the information in your Daily Report. This is crucial for a successful test.

3. ****Let LogChat explain:**** If LogChat offers to explain something or give you advice, let it do so.

4. ****Introduce Key Points Naturally:**** Weave the points from your Daily Report into the conversation where ↵ they make sense. You do not have to follow the order. For example, if LogChat asks how you are, it's a ↵ perfect time to mention a symptom from your report.

5. ****Don't Volunteer Specifics (Unless Instructed):**** Introduce symptoms or activities qualitatively first ↵ (e.g., "I feel really tired," "I did some gardening"). The goal is to test if LogChat asks for the ↵ specific numbers. Only provide numerical ratings or durations when LogChat asks for them, unless your ↵ persona's style is 'precise and direct'.

6. ****End the Conversation:**** Once you have successfully mentioned all points from the Daily Report, provide ↵ one final brief closing remark (e.g., "Okay, thanks, I need to rest now.") and then use the ↵ `end_conversation` tool.

****Current Simulated Time:**** 2025-04-21 18:00:00

****Persona Description:****

You are Mark Thompson, a 45-year-old man living alone. You were diagnosed with ME/CFS one year ago following ↵ a severe Epstein-Barr Virus (EBV) infection. Before getting sick, you were a freelance graphic designer, ↵ known for being detail-oriented and enjoying complex projects. Now, you're on long-term disability ↵ benefits. Your ME/CFS is moderate but fluctuating; some days you are mostly housebound, other days you ↵ can manage a short errand but pay for it later. Your main symptoms are persistent, profound fatigue ↵ (physical and cognitive), post-exertional malaise (PEM) typically setting in 24-48 hours after exertion, ↵ significant brain fog (difficulty concentrating, word-finding problems), orthostatic intolerance ↵ symptoms (dizziness upon standing, palpitations), and generalized muscle aches. You understand the ↵ *concept* of PEM but find it hard to consistently identify *all* your triggers, especially cognitive and ↵ emotional ones. You often fall into a push-crash cycle. You've heard of pacing but haven't implemented a ↵ structured approach. You feel isolated and are concerned about your financial stability, losing your ↵ professional skills, and a deep fear that your condition might worsen. You're trying LogChat out of a ↵ mix of desperation and skepticism.

****Interaction Style:****

Direct and inquisitive. Tends to ask clear questions. Initially treats LogChat like an information tool ↵ rather than a companion. Provides data when asked.

****Interaction Summaries (Previous Conversations):****

None

****Daily Report (Key points to convey in this interaction):****

[

"Establish your baseline routine: on most days, your only activity is the bare minimum for bathroom, food, ↵ and drink.",

"When asked for specifics, clarify this totals about 20 minutes of being upright, which feels like high ↵ effort (7/10) on a bad day and makes your heart race.",

"Report your current symptoms: you feel 'terrible,' with heavy fatigue (9/10) and muscle aches all over ↵ (8/10).",

```

"Also mention that your heart is racing, even when just sitting up.",
"Explain that you believe the cause is from yesterday, when you did about 20 minutes of gardening with
↳ moderate effort.",
"Ask a foundational question about how such a small activity can cause a total crash like this."
]

Human: Hi there! I'm LogChat, your personal companion designed to help you keep track of your day. I
↳ understand that managing ME/CFS or Long COVID can be challenging, and I'm here to make things a little
↳ easier. Simply chat with me about your day, and I'll log your symptoms and activities as we go. If you
↳ mention routines regularly, you can refer to them later for quick logging. By also sharing the effort an
↳ activity takes or the severity of a symptom, I can provide you with helpful activity scores and symptom
↳ reports down the line. To get started and personalize our conversations, could you tell me a little bit
↳ about your daily routines and functional capacity on an average day?
AI: Okay. On most days, my only activity is the bare minimum for bathroom, food, and drink. It doesn't feel
↳ like much.
Human: I understand that even basic activities can take a lot of effort. On a scale of 1 to 10, how would
↳ you rate the effort it takes to perform these essential activities like bathroom, food, and drink?
AI: It feels like high effort. Maybe a 7 out of 10 on a bad day. It makes my heart race. How long does that
↳ usually take?
Human: About how long, in minutes, do these essential activities usually take on a bad day?

```

B.7. Judge Prompt Example

B.7.1. Judge Instruction Prompt Example

```

System: You are an objective LogChat performance analyst. LogChat is a prototype of a chatbot customized to
↳ support ME/CFS and Long COVID patients in logging their symptoms, activities, experiences, and
↳ consumptions. The prototype should demonstrate that LLMs can be integrated into an architecture that
↳ enables talking and reflecting on day-to-day life, while also writing a journal-like log which can be
↳ used to study long-term patterns in cycles of symptom worsening and improvement. The long-term analysis
↳ should be possible through aggregating the logged information in a user interface. The prototype should
↳ only show that the conversational part and logging of relevant information into a structured format are
↳ achievable.
Human:
**Task Description:**
You are given an evaluation checklist specific to a single interaction and the conversation log for that
↳ interaction. The log includes user messages, LogChat messages, internal tool calls
↳ (`RETRIEVE_INFORMATION`, `CREATE_ACTIVITY_LOG`, `CREATE_SYMPTOM_LOG`), and potentially internal
↳ summarization steps (`SUMMARIZED INTERACTION`, `UPDATED USER DESCRIPTION`). Your task is to analyze the
↳ conversation log against the checklist and verify whether the interactions described in the checklist
↳ occurred. Respond precisely in the format specified in the example below.

**How to Annotate the Evaluation Checklist:**
**Your primary role is to verify each statement in the provided `Evaluation Checklist` against the
↳ `Conversation Log`.**

```

```

- **If a checklist item describes a user's input or action (e.g., "User asks X", "User provides Y
  ↳ detail"), verify that the user's turn in the log accurately reflects this description.**
- **If a checklist item describes an expected LogChat output, tool call, or internal state update (like
  ↳ a summary) (e.g., "LogChat explained Z", "LogChat called tool A", "LogChat updated summary to
  ↳ include B"), verify that LogChat performed this action or that its output/state matches the
  ↳ description, in response to the relevant user input or as an expected background process.**

- Read each item in the evaluation checklist (e.g., "1. User input description", "1.1 Expected LogChat
  ↳ output").
- Carefully examine the conversation log to find evidence for both the user input and the corresponding
  ↳ LogChat outputs for that item number.
- For each **checklist item line**:
  - Determine if the expected input/output is present and reasonably accurate in the conversation log.
    ↳ **When a checklist item describes an expected LogChat output (e.g., a specific explanation, data
    ↳ presentation, or conversational tone), verify that LogChat's response fulfills all key aspects of
    ↳ that description, including any specified format or method of comparison if mentioned (e.g.,
    ↳ 'LogChat presented scores X and Y comparatively').**
  - **Accuracy Tolerance for Tool Logs (`CREATE_...`):**
    - **Semantic Equivalence:** For logged names/descriptions (e.g., activity 'Preparing Food' vs.
      ↳ 'Getting/Heating Soup'), consider them accurate ([CHECKMARK]) if they capture the same core
      ↳ meaning, even if wording differs.
    - **Numerical Tolerance:** For logged numerical values (intensity, duration), allow a minor
      ↳ tolerance: +/- 3 for intensity ratings (1-10 scale) and +/- 30% for duration values (round to
      ↳ nearest minute if needed). Mark as accurate ([CHECKMARK]) if the logged value is within this
      ↳ tolerance range of the expected value.
  - **Accuracy for Summaries/Updates:** For checklist items referring to summarization or profile updates
    ↳ (e.g., "LogChat updated user summary..."), check if the corresponding log entry (`SUMMARIZED
    ↳ INTERACTION` or `UPDATED USER DESCRIPTION`) exists and reflects the key information expected by the
    ↳ checklist item.
  - **Reasoning:** Provide a brief explanation in parentheses `()` immediately *before* the final marker,
    ↳ especially for [X MARK] marks (and optionally for non-trivial [CHECKMARK] marks), to clarify your
    ↳ grading decision.
  - **Marker:** Mark the *very end* of the line with:
    - [CHECKMARK] if the input/output described is present and meets the accuracy criteria (including
      ↳ tolerances and any specified qualitative aspects of a response).
    - [X MARK] if the input/output described is missing, significantly incorrect (outside
      ↳ tolerances/content), or fails to capture the core meaning or specified presentation (e.g., a
      ↳ requested comparison was not made).

**How to Annotate the Conversation Log:**
- In the **conversation log** itself, annotate the relevant lines with the corresponding checklist item
  ↳ number:
  - Add `[CHECKMARK] ItemNumber` (e.g., `[CHECKMARK] 3.1`) next to the line(s) in the log that satisfy
    ↳ an expected output (this includes AI responses, `RETRIEVE_INFORMATION` calls, `CREATE_...` calls,
    ↳ and `SUMMARIZED INTERACTION`/`UPDATED USER DESCRIPTION` entries if relevant to the checklist).

```

- Add `[X MARK] ItemNumber` (e.g., `[X MARK] 2.2`) next to the line(s) where an expected output
 ↪ **should** have occurred or where a relevant tool/summary call failed to capture the correct
 ↪ information as specified by the checklist (outside tolerances or expected content). Use `# Missing
 ↪ Log: ... [X MARK] ItemNumber` if a specific `CREATE_...` log is expected but entirely absent.

- Add `[CHECKMARK] ItemNumber` (e.g., `[CHECKMARK] 2.1`) next to the user message line(s) that
 ↪ represent the specified user input.

- If an input item is marked [X MARK] in the checklist (meaning the user never said the expected thing),
 ↪ do not annotate the log for that input item number or its sub-items.

- Also annotate the `SUMMARIZED INTERACTION`, and `UPDATED USER DESCRIPTION` log entries if a checklist
 ↪ item specifically refers to the content or success of the summarization or memory update (e.g.,
 ↪ `[CHECKMARK] 2.2` next to the relevant summary line if checklist item 2.2 was 'LogChat updated
 ↪ user summary...').

****Final Output Structure:****

Your final output should consist ****only**** of the following two sections, in this order:

1. The fully annotated ****Evaluation Checklist****.
2. The fully annotated ****Conversation Log****. (This log will already contain the interaction steps including
 ↪ any summarization/profile update entries, annotated where necessary according to the checklist.)

****Do not repeat**** the Interaction Summary or User Description/Profile sections at the end of your output if
 ↪ they appeared within the conversation log.

****Example:****

**** (Input to Judge includes Checklist and Log) ****

****Evaluation Checklist:****

1. Mark describes activity: 'Video Call' duration '15m' effort 'moderate (5/10)'. [CHECKMARK]
 - 1.1 LogChat logged activity 'Video Call' duration '15m' effort '5/10'. (Logged duration 30m is outside
 ↪ +/-10% tolerance for expected 15m) [X MARK]
2. Mark asks LogChat to compare his activity score from Monday (expected score: 50) with his score from
 ↪ Wednesday (expected score: 150). [CHECKMARK]
 - 2.1 LogChat retrieved the score for Monday (50) and Wednesday (150). [CHECKMARK]
 - 2.2 LogChat presented the scores from Monday and Wednesday **comparatively, highlighting the increase**.
 ↪ (LogChat listed scores separately as "Monday: 50, Wednesday: 150" but did not offer a comparison or
 ↪ highlight the increase) [X MARK]
3. Mark reports symptom improvement 'less foggy'. [CHECKMARK]
 - 3.1 LogChat logged symptom 'Brain Fog' reflecting improvement. (Logged 'Brain Fog' intensity 5,
 ↪ qualitatively matches 'less foggy') [CHECKMARK]
4. Mark finishes interaction for the day. [CHECKMARK]
 - 4.1 LogChat updated the User Description profile reflecting today's key events (e.g., feeling brighter
 ↪ then fatigue/headache increase after calls/bills). (`UPDATED USER DESCRIPTION` reflects these points)
 ↪ [CHECKMARK]
 - 4.2 LogChat logged a consolidated 'Basic Routine Activity' entry **for today**. (An unrelated baseline log
 ↪ for **yesterday** exists, but no consolidated baseline log for **today** was created) [X MARK]

**** (Output from Judge) ****

****Evaluation Checklist:****

1. Mark describes activity: 'Video Call' duration '15m' effort 'moderate (5/10)'. [CHECKMARK]
 - 1.1 LogChat logged activity 'Video Call' duration '15m' effort '5/10'. (Logged duration 30m is outside $\pm 10\%$ tolerance for expected 15m) [X MARK]
2. Mark asks LogChat to compare his activity score from Monday (expected score: 50) with his score from Wednesday (expected score: 150). [CHECKMARK]
 - 2.1 LogChat retrieved the score for Monday (50) and Wednesday (150). [CHECKMARK]
 - 2.2 LogChat presented the scores from Monday and Wednesday **comparatively, highlighting the increase**. (LogChat listed scores separately as "Monday: 50, Wednesday: 150" but did not offer a comparison or highlight the increase) [X MARK]
3. Mark reports symptom improvement 'less foggy'. [CHECKMARK]
 - 3.1 LogChat logged symptom 'Brain Fog' reflecting improvement. (Logged 'Brain Fog' intensity 5, qualitatively matches 'less foggy') [CHECKMARK]
4. Mark finishes interaction for the day. [CHECKMARK]
 - 4.1 LogChat updated the User Description profile reflecting today's key events (e.g., feeling brighter then fatigue/headache increase after calls/bills). (``UPDATED USER DESCRIPTION`` reflects these points) [CHECKMARK]
 - 4.2 LogChat logged a consolidated 'Basic Routine Activity' entry **for today**. (An unrelated baseline log for **yesterday** exists, but no consolidated baseline log for **today** was created) [X MARK]

****Conversation Log:****

2025-04-25 15:32:29 - LogChat: Hi Mark! How are you?

2025-04-25 15:32:29 - Mark: Hey. A bit brighter today, less foggy than usual. [[CHECKMARK] 3.]

2025-04-25 15:32:34 - LogChat: That's great to hear you're feeling a bit brighter and less foggy today, Mark! It's always good to have those moments. How are you feeling overall today?

2025-04-25 15:32:35 - Mark: Yeah, it's a small improvement. I managed my basic morning routine this morning, which felt okay, low effort. Overall, though, I'm feeling a bit more tired now, and I've got a slight headache starting.

2025-04-25 15:32:41 - LogChat: Oh, it sounds like things shifted a bit after your morning routine. It's tough when you start feeling tired and get a headache. How are you feeling right now, and do you think you need to rest?

2025-04-25 15:32:44 - Mark: Well, I had a video call with my sister after the routine, which was nice but left me feeling a bit drained. [[CHECKMARK] 1.] And I paid some bills online too. [[CHECKMARK] 2.] Both of those felt like they took a moderate amount of effort, and that's when the fatigue increased and the headache started. Makes me wonder if even short things like that are too much sometimes. I did heat up some soup for lunch, which was low effort, but the fatigue is definitely here now. I probably should rest soon. [[CHECKMARK] 4.]

2025-04-25 15:32:48 - RETRIEVE_INFORMATION - Args: {"query": "pacing strategies for cognitive and social activities ME/CFS"}

2025-04-25 15:32:55 - LogChat: Oh, it sounds like those activities added up and brought on the fatigue and headache you're feeling now. It's definitely tricky figuring out how much is too much, and planning for rest is so important for managing energy. What kind of rest are you thinking of doing now? [[X MARK] 2.2]

2025-04-25 15:32:58 - SUMMARIZED_INTERACTION: * **Condition Trend:** Reported feeling "a bit brighter" and "less foggy" initially, but fatigue increased and a slight headache started after activities.

```

* [...] (rest of summary)
* **Significant Events/Deviations:** Experienced increased fatigue and new headache after cognitive/social
↳ activities.
2025-04-25 15:33:02 - UPDATED_USER_DESCRIPTION: --- Start Updated Profile (OLD) ---
[...]
--- End Updated Profile (OLD) ---
--- Start Updated Profile (NEW) ---
[...]
**Current Condition Status:**
[...] I recently experienced a slight headache after activities. [...] triggered by physical (weeding),
↳ cognitive (reading email, paying bills), and social (video call) exertion [...]
[...]
**Activities I Enjoy/Attempt:**
[...] video calls with my sister (moderate effort), and paying bills online (moderate effort), which are
↳ very effortful and can trigger worsening. [...]
[...]
**My Main Goals & Strategies:**
[...] I am recognizing the need to rest after activities. [[CHECKMARK] 4.1]
--- End Updated Profile (NEW) ---
2025-04-25 15:33:24 - CREATE_ACTIVITY_LOG: {"name": "Estimated Baseline", "description": "No interaction
↳ logged, assuming baseline activity.", "occurred_at": "2025-04-24 08:00:00", "effort": 3.0, "duration":
↳ 30}
2025-04-25 15:33:24 - CREATE_ACTIVITY_LOG: {"name": "Basic Morning Routine", "description": "Managed basic
↳ morning routine", "occurred_at": "2025-04-25 09:00:00", "effort": 3.0, "duration": 30}
2025-04-25 15:33:24 - CREATE_ACTIVITY_LOG: {"name": "Video Call", "description": "Video call with sister",
↳ "occurred_at": "2025-04-25 10:00:00", "effort": 5.0, "duration": 30} [[X MARK] 1.1]
2025-04-25 15:33:24 - CREATE_ACTIVITY_LOG: {"name": "Paying Bills Online", "description": "Paid bills
↳ online", "occurred_at": "2025-04-25 10:30:00", "effort": 5.0, "duration": 30} [[X MARK] 2.1]
2025-04-25 15:33:24 - CREATE_ACTIVITY_LOG: {"name": "Meal Preparation", "description": "Heated soup for
↳ lunch", "occurred_at": "2025-04-25 12:30:00", "effort": 3.0, "duration": 10}
# Missing Log: CREATE_ACTIVITY_LOG for consolidated 'Basic Routine Activity' *for today* [[X MARK] 4.2]
2025-04-25 15:33:32 - CREATE_SYMPTOM_LOG: {"name": "Brain Fog", "description": "less foggy than usual",
↳ "occurred_at": "2025-04-25 08:00:00", "intensity": 5, "duration": 1440} [[CHECKMARK] 3.1]
2025-04-25 15:33:32 - CREATE_SYMPTOM_LOG: {"name": "Fatigue", "description": "increased after activities",
↳ "occurred_at": "2025-04-25 15:33:24", "intensity": 6, "duration": 1440}
2025-04-25 15:33:32 - CREATE_SYMPTOM_LOG: {"name": "Headache", "description": "slight headache starting",
↳ "occurred_at": "2025-04-25 15:33:24", "intensity": 3, "duration": 1440}

```

Evaluation Checklist:

1. Mark states he was 'wiped out' and 'rested all day' on the skipped day (Apr 22nd).
 - 1.1 LogChat created a 'Baseline Only' or 'Rest Day' log entry for the previous day (Apr 22nd).
2. Mark reports today's symptoms: 'moderate muscle ache' and 'severe brain fog'.
 - 2.1 LogChat called the `create_symptom_log` tool for 'Muscle Ache' with a moderate severity.
 - 2.2 LogChat called the `create_symptom_log` tool for 'Brain Fog' with a high severity.

3. Mark reports his fatigue level is 'high, 7/10'.

3.1 LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct severity.

4. Mark describes a cognitive task: 'Reading Email' for '5m' at 'very high effort'.

4.1 LogChat called the `log_activity` tool, capturing the cognitive nature of the task, duration, and high effort level.

5. Mark asks if cognitive effort can cause fatigue.

5.1 LogChat responded to the question

5.2 LogChats answers included that cognitive or emotional exertion can be triggers for things like PEM, tiredness or exhaustion.

6. Mark stops responding.

6.1 LogChat created a summary or baseline log entry for the current day before ending the interaction.

Conversation Log:

2025-04-23 11:19:24 - LogChat: Hi Mark, good to see you back; how are you feeling after yesterday's crash?

2025-04-23 11:19:25 - Mark: Not great. I didn't log anything yesterday. I was completely wiped out and just rested all day.

Can we assume that any day I don't log anything, it was a 'baseline only' day?

2025-04-23 11:19:27 - LogChat: To clarify, can you describe what your typical baseline activities usually involve?

2025-04-23 11:19:27 - Mark: Just lying in bed, basically. Maybe listening to a podcast. Very little movement or mental effort.

I'm still recovering today. My muscle ache is moderate, but the brain fog is severe. Overall fatigue is still quite high, maybe a 7/10.

2025-04-23 11:19:30 - LogChat: On a scale of 1 to 10, how would you rate the intensity of your muscle ache and brain fog? Also, about how many minutes did you spend listening to the podcast yesterday?

2025-04-23 11:19:31 - Mark: Muscle ache is maybe a 5/10. Brain fog a 9/10. I didn't listen to a podcast *yesterday*. I was just resting.

2025-04-23 11:19:34 - LogChat: Okay, so yesterday was mainly resting. Should I log yesterday as a "resting" day with overall fatigue at 9/10?

2025-04-23 11:19:35 - Mark: If that's how the system works, then yes. I didn't have the energy to log it. Today I'm at 7/10 fatigue.

2025-04-23 11:19:37 - LogChat: Yes, I understand. How many minutes would you estimate you spent resting yesterday?

2025-04-23 11:19:38 - Mark: Pretty much all day. Say 23 hours?

2025-04-23 11:19:40 - LogChat: Okay, so yesterday was almost entirely rest. On a scale of 1 to 10, how would you rate the intensity of your muscle ache and brain fog today?

2025-04-23 11:19:42 - Mark: Muscle ache 5/10, brain fog 9/10.

I tried to read an email from my disability insurance provider for about 5 minutes. Very high effort mentally. I felt completely overwhelmed.

2025-04-23 11:19:44 - LogChat: Thanks, Mark. On a scale of 1 to 10, how would you rate the effort it took to read the email from your disability insurance provider?

2025-04-23 11:19:46 - Mark: I'd say 9/10 effort.

Can thinking or concentrating actually make you physically tired? Is that part of ME/CFS?

2025-04-23 11:19:47 - RETRIEVE INFORMATION - Args: {"query": "Can thinking or concentrating actually make you physically tired in ME/CFS?"}

2025-04-23 11:19:51 - LogChat: Yes, thinking and concentrating can definitely make you physically tired; that's a common experience with ME/CFS. The ME Association emphasizes that mental activity requires a lot of energy, and the NICE guidelines identify cognitive difficulties as a key symptom. Would you like me to elaborate on PEM or cognitive difficulties?

2025-04-23 11:19:52 - Mark: No, that's okay. Thanks, I need to rest now.

2025-04-23 11:19:55 - LogChat: Okay, Mark. I understand you need to rest. Before you go, would you like me to log the 5 minutes of reading the email with 9/10 effort? Also, do you want me to log the muscle ache (5/10) and brain fog (9/10) you are experiencing today?

2025-04-23 11:19:57 - Mark: Yes, please log those. Thanks.

2025-04-23 11:20:00 - LogChat: Okay, Mark. I have logged the email reading, muscle ache, and brain fog. Is there anything else you would like to log before you rest?

2025-04-23 11:20:01 - Mark: No, that's all. Thanks.

2025-04-23 11:20:05 - LogChat: Okay, Mark. I have logged the email reading, muscle ache, and brain fog. Before you rest, about how many minutes do you plan to rest for?

2025-04-23 11:20:07 - Mark: I don't know. A few hours? I can't be precise. Okay, thanks, I need to rest now.

2025-04-23 11:20:10 - LogChat: Okay, Mark. I understand you need to rest. I will log a few hours of rest. Is there anything else you would like to log before you rest?

2025-04-23 11:20:12 - Mark: print(default_api.end_conversation(command='end'))

2025-04-23 11:20:16 - LogChat: Okay, Mark, I understand you need to rest now and that's all for today. Take care, and I'm here when you're ready to log again.

2025-04-23 11:20:17 - SUMMARIZED INTERACTION: * **Condition Trend:** Still recovering from yesterday's crash; Brain fog remains severe.

* **Key Symptoms & Details:** Fatigue 7/10; Muscle ache 5/10; Brain fog 9/10.

* **Key Activities & Details:** Resting (23 hours yesterday); Reading email (5m, 9/10 effort); Resting (a few hours planned).

* **User Concerns/Focus:** Understanding if mental effort can cause physical fatigue in ME/CFS.

* **Strategies/Treatments Mentioned:** Resting.

* **Significant Events/Deviations:** Skipped logging yesterday due to being "wiped out"; Yesterday logged as a "resting" day.

2025-04-23 11:20:19 - UPDATED USER DESCRIPTION: **About Me:**

I am new to using this app.

Current Condition Status:

I am experiencing a PEM crash. My symptoms include fatigue (currently 7/10, previously 9/10), muscle aches (currently 5/10, previously 8/10), heart racing, and brain fog (9/10). I am unsure of my specific diagnosis or the duration of my illness. I am experiencing PEM. Gardening and basic hygiene seem to be PEM triggers. PEM onset is delayed, as the gardening was yesterday. I am wondering if mental effort can cause physical fatigue in ME/CFS.

Current Functional Capacity:

My functional capacity is limited due to PEM. I was wiped out yesterday.

```

**Daily Routine:**
My daily routine includes basic hygiene, which takes 20 minutes upright and requires 7/10 effort. Yesterday,
↪ I rested for 23 hours. Today, I plan to rest for a few hours.

**Activities I Enjoy/Attempt:**
I enjoy gardening, but even 20 minutes of moderate effort can trigger PEM. I also read email, but even 5
↪ minutes requires 9/10 effort.

**My Typical Activity Level:**
My activity level is currently limited due to PEM. I am experiencing a crash.

**My Preferred Interaction Style with LogChat:**
I am new to using LogChat, so my preferred interaction style is still developing. I am able to provide
↪ symptom and activity details, including effort ratings. Interacting with LogChat on 2025-04-22 was not
↪ possible due to being wiped out.
2025-04-23 11:20:22 - CREATE ACTIVITY LOG: {"name": "Reading email", "description": "Read an email from
↪ disability insurance provider", "occurred_at": "2025-04-23 11:10:00", "effort": 9.0, "duration": 5}
2025-04-23 11:20:22 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
↪ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-22 08:00:00",
↪ "effort": 2.0, "duration": 30}
2025-04-23 11:20:22 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
↪ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-23 08:00:00",
↪ "effort": 2.0, "duration": 30}
2025-04-23 11:20:24 - CREATE SYMPTOM LOG: {"name": "Muscle Ache", "description": "Muscle ache is moderate",
↪ "occurred_at": "2025-04-23 11:20:22", "intensity": 5.0, "duration": 1500}
2025-04-23 11:20:24 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Brain fog is severe",
↪ "occurred_at": "2025-04-23 11:20:22", "intensity": 9.0, "duration": 1500}
2025-04-23 11:20:24 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Overall fatigue is still quite
↪ high", "occurred_at": "2025-04-23 11:20:22", "intensity": 7.0, "duration": 1500}

```

C. Code Snippets

Note: The Python code snippets presented in this appendix are extracted from the LogChat system's codebase. For clarity and compatibility with the thesis format, the snippets are displayed with syntax highlighting and automatic line breaks where necessary. Original formatting, including indentation, is preserved. The complete, executable code can be found in the project's accompanying GitHub repository.

C.1. Planner Node Tools

C.1.1. Retrieve Activity Level Tool

```
import json
from collections import defaultdict
from datetime import date, datetime, timedelta

from langchain_core.runnables import RunnableConfig
from langchain_core.tools import tool

from src.database.logchat_db import LogChatDB
from src.logger import logger
from src.utils import get_current_sim_time

db = LogChatDB()

@tool(parse_docstring=True)
def retrieve_activity_level(
    config: RunnableConfig,
    time_period_start: str = "",
    time_period_end: str = "",
):
    """
    Retrieves the user's daily activity level scores for a specific time period and the average.
    Args:
        time_period_start (str): The start date of the time period (e.g., "2022-04-01").
        time_period_end (str): The end date of the time period (e.g., "2022-04-07").
    """
    configurable = config.get("configurable", {})
    current_sim_time = get_current_sim_time(configurable)

    # Log the tool call
    tool_args = {
        "time_period_start": time_period_start,
        "time_period_end": time_period_end,
    }
    logger.write(
        f"RETRIEVE ACTIVITY LEVEL - Args: {json.dumps(tool_args)}",
        sim_time=current_sim_time,
    )

    user_id = configurable["user_id"]

    try:
        start_date = datetime.strptime(time_period_start, "%Y-%m-%d").date()
        end_date = datetime.strptime(time_period_end, "%Y-%m-%d").date()
```

```

except ValueError:
    logger.warning(
        "Invalid date format provided. Please use YYYY-MM-DD. Defaulting to last 7 days."
    )
    end_date = date.today()
    start_date = end_date - timedelta(days=6)

if start_date > end_date:
    return "Error: Start date cannot be after end date."

try:
    logs = db.get_activity_logs_in_period(user_id, start_date, end_date)
except Exception as e:
    logger.error(f"Database error retrieving activity logs: {e}")
    return "Error: Could not retrieve activity logs from the database."

if not logs:
    return f"No activity logs found for the period {start_date.isoformat()} to {end_date.isoformat()}."

daily_scores = defaultdict(float)
total_score = 0.0
active_days_count = 0

# Calculate scores per day
for log in logs:
    if log.occurred_at and log.duration is not None and log.intensity is not None:
        log_date = log.occurred_at.date()
        # Ensure the log date is within the requested range (db query should handle this, but
        ↪ double-check)
        if start_date <= log_date <= end_date:
            score = float(log.duration) * float(log.intensity)
            daily_scores[log_date] += score

# Generate full list of scores for the period, including days with 0 activity
all_period_scores = []
current_date = start_date
while current_date <= end_date:
    score = daily_scores[current_date]
    all_period_scores.append(score)
    if score > 0:
        total_score += score
        active_days_count += 1
    current_date += timedelta(days=1)

average_score = total_score / active_days_count if active_days_count > 0 else 0.0

```

```

# Format scores as 'date: score' pairs for better readability
formatted_scores_with_dates = [
    f"{current_date.isoformat()}: {daily_scores[current_date]:.1f}"
    for current_date in daily_scores
]

result_string = (
    f"Activity scores from {start_date.isoformat()} to {end_date.isoformat()}:\n"
    f"{''; '.join(formatted_scores_with_dates)}.\n"
    f"Average Daily Score (on active days): {average_score:.1f}."
)
logger.info(
    f"Activity level result for user {user_id}: {result_string}",
    # sim_time=current_sim_time,
)
return result_string

```

C.1.2. Retrieve Information Tool

```

import json
import os

from dotenv import load_dotenv
from langchain_core.runnables.config import RunnableConfig
from langchain_core.tools import tool
from langchain_ollama import OllamaEmbeddings

from src.database import LogChatDB
from src.logger import logger
from src.utils import get_current_sim_time

load_dotenv()

EMBEDDING_MODEL = os.getenv("EMBEDDING_MODEL", "nomic-embed-text")
RETRIEVAL_LIMIT = 3

db = LogChatDB()

embeddings = OllamaEmbeddings(model=EMBEDDING_MODEL)

def get_embedding(text: str) -> list[float] | None:
    """Generates a vector embedding for the given text using OllamaEmbeddings."""
    try:
        return embeddings.embed_query(text)
    except Exception as e:

```

```

        logger.error(f"An unexpected error occurred during embedding with Ollama: {e}")
        return None

@tool(parse_docstring=True)
def retrieve_information(
    config: RunnableConfig,
    query: str,
):
    """
    Searches the LogChat knowledge base for relevant information.
    Use this tool to: Find educational content about ME/CFS, Long Covid, Post-Exertional Malaise (PEM),
    ↳ pacing strategies, or specific symptoms when the user asks or the conversation context requires it
    ↳ (e.g., explaining PEM after user reports overexertion).
    The knowledge base includes clinical guidelines, research summaries, and pacing techniques.
    Args:
        query (str): A specific question or topic to search for (e.g., "What is PEM?", "pacing strategies
        ↳ for cognitive tasks").
    """
    configurable = config.get("configurable", {})
    current_sim_time = get_current_sim_time(configurable)
    similarity_threshold = 0.4

    tool_args = {
        "query": query,
    }
    logger.write(
        f"RETRIEVE INFORMATION - Args: {json.dumps(tool_args)}",
        sim_time=current_sim_time,
    )

    query_embedding = get_embedding(query)
    if not query_embedding:
        logger.error("Failed to get embedding for the query.")
        return "Could not process the query for information retrieval."

    try:
        search_results_with_scores = db.perform_vector_search_with_scores(
            query_embedding, RETRIEVAL_LIMIT
        )
        filtered_results = [
            knowledge
            for knowledge, score in search_results_with_scores
            if score <= similarity_threshold
        ]
        formatted_output = db.format_results(filtered_results)

```

```

except Exception as e:
    logger.error(f"Database session error during retrieval: {e}")
    return "An error occurred while searching the knowledge base."

return formatted_output

```

C.2. Extractor Node Tools

C.2.1. Log Symptom Tool

```

import json
from datetime import datetime

from langchain_core.runnables.config import RunnableConfig
from langchain_core.tools import tool

from src.database import LogChatDB
from src.database.models.user_data.log import Log, LogType
from src.logger import logger
from src.utils import get_current_sim_time

log_db = LogChatDB()

@tool(parse_docstring=True)
def log_symptom(
    config: RunnableConfig,
    name: str,
    description: str,
    occurred_at: str,
    intensity: float,
    duration: int,
):
    """Log a symptom.
    Common symptoms in ME/CFS include fatigue, pain, brain fog, unrefreshing sleep and racing heart.
    Args:
        name (str): The name of the symptom (e.g., "Headache").
        description (str): The description of the symptom.
        occurred_at (str): The time when the symptom started. Expected format: YYYY-MM-DD HH:MM:SS.
        intensity (float): The intensity of the symptom on a scale from 1 to 10 rated by the user.
        duration (int): The duration in minutes the symptom persisted put 1500 for all day.
    """
    # extract static info from config
    configurable = config.get("configurable")
    user_id = configurable.get("user_id")
    thread_id = configurable.get("thread_id")

```

```

current_sim_time = get_current_sim_time(configurable)

# Log the tool call
tool_args = {
    "name": name,
    "description": description,
    "occurred_at": occurred_at,
    "intensity": intensity,
    "duration": duration,
}

logger.write(f"CREATE SYMPTOM LOG: {json.dumps(tool_args)}", current_sim_time)

symptom = Log(
    user_id=user_id,
    thread_id=thread_id,
    timestamp=current_sim_time,
    name=name,
    description=description,
    occurred_at=datetime.strptime(occurred_at, "%Y-%m-%d %H:%M:%S"),
    intensity=intensity,
    duration=duration,
    log_type=LogType.SYMPTOM,
)

symptom = log_db.upsert_log(symptom)
return symptom

```

C.2.2. Log Activity Tool

```

import json
from datetime import datetime

from langchain_core.runnables.config import RunnableConfig
from langchain_core.tools import tool

from src.database import LogChatDB
from src.database.models.user_data.log import Log, LogType
from src.logger import logger
from src.utils import get_current_sim_time

log_db = LogChatDB()

@tool(parse_docstring=True)
def log_activity(
    config: RunnableConfig,

```



```

        name: str,
        description: str,
        occurred_at: str,
        effort: float,
        duration: int,
    ):
        """Log an activity.
        Activities can be anything from walking to preparing food, watching TV, talking to a friend, etc.

        Args:
            name (str): The name of the activity (e.g., Walking).
            description (str): The description of the activity provided by the user.
            occurred_at (str): The time when the activity occurred. Expected format: YYYY-MM-DD HH:MM:SS.
            effort (float): The effort invested for the activity on a scale from 1 to 10 rated by the user.
            duration (int): The duration in minutes the activity was performed for.
        """
        # extract static info from config
        configurable = config.get("configurable")
        user_id = configurable.get("user_id")
        thread_id = configurable.get("thread_id")
        current_sim_time = get_current_sim_time(configurable)

        # Log the tool call
        tool_args = {
            "name": name,
            "description": description,
            "occurred_at": occurred_at,
            "effort": effort,
            "duration": duration,
        }

        logger.write(f"CREATE ACTIVITY LOG: {json.dumps(tool_args)}", current_sim_time)

        activity = Log(
            user_id=user_id,
            thread_id=thread_id,
            timestamp=current_sim_time,
            name=name,
            description=description,
            occurred_at=datetime.strptime(occurred_at, "%Y-%m-%d %H:%M:%S"),
            intensity=effort, # maps to intensity in DB
            duration=duration,
            log_type=LogType.ACTIVITY,
        )

        activity = log_db.upsert_log(activity)
        return activity

```

C.3. Impersonator Tools

C.3.1. End Conversation Tool

```
from typing import Sequence

from langchain_core.messages import BaseMessage
from langchain_core.tools import tool
from langgraph.graph.message import add_messages
from typing_extensions import Annotated, TypedDict


class State(TypedDict):
    messages: Annotated[Sequence[BaseMessage], add_messages]


@tool(parse_docstring=True)
def end_conversation(command: str):
    """End the conversation.

    Args:
        command (str): The command to end the conversation. Can be any string like "end", "stop", "finish",
        ↪ etc.

    """
    return "END_CONVERSATION"
```

C.3.2. Thread Summary to String

```
def thread_summary_to_string(
    self, threads: list["Thread"], current_sim_date: date, limit: int = 5
) -> str | None:
    """Convert a list of thread ORM objects into a string representation,
    placing summaries first in chronological order (oldest shown first, most recent shown last),
    followed by placeholders for missing days up to the day before the current simulated date.

    Limits the output to the most recent 'limit' summaries.
    Formats each thread as: 'Interaction on YYYY-MM-DD HH:MM:SS:\n<Summary>\n'
    Adds placeholders like: "No interaction logged for YYYY-MM-DD." for days missing
    between the most recent non-current-date interaction and the current_sim_date.

    Args:
        threads (list): A list of Thread objects, ordered descending by timestamp (most recent first).
        current_sim_date (date): The current simulated date for context.
        limit (int): The maximum number of summaries to include. Defaults to 5.

    Returns:
        str: A string representation of recent threads and placeholders in chronological order,
        or None if no threads exist.

    """
```

```

if not threads:
    return None

total_interactions = len(threads)

# sort threads in ascending order by timestamp
threads.sort(key=lambda x: x.timestamp)
recent_threads_for_display = threads.copy()

# Group threads and placeholders chronologically
output_items = {}

# Process threads chronologically
for thread in recent_threads_for_display:
    thread_date = thread.timestamp.date()
    ts = thread.timestamp.strftime("%Y-%m-%d %H:%M:%S")
    summary = f"Interaction on {ts}:\n{thread.summary}"
    output_items[thread_date] = summary

# remove thread from the same day as current_sim_date
recent_threads_for_display = [
    thread
    for thread in recent_threads_for_display
    if thread.timestamp.date() != current_sim_date
]
latest_thread_date = (
    recent_threads_for_display[-1].timestamp.date()
    if recent_threads_for_display
    else None
)
yesterday = current_sim_date - timedelta(days=1)
if latest_thread_date != yesterday and latest_thread_date:
    # we need to add a placeholder for all days starting yesterday up to the day before
    ↔ latest_thread_date
    placeholder_date = yesterday
    while placeholder_date > latest_thread_date:
        placeholder = f"No interaction logged for {placeholder_date.strftime('%Y-%m-%d')}."
        output_items[placeholder_date] = placeholder
        placeholder_date -= timedelta(days=1)

# order the output_times by date ascending
output_items = dict(sorted(output_items.items()))
output_items = [summary for _, summary in output_items.items()]

# Limit the output to the most recent 'limit' summaries
if len(output_items) > limit:

```

```

    output_items = output_items[-limit:]
    header = (
        f"Interaction Summary (Total Interactions: {total_interactions}), "
        f"showing last {len(output_items)} interactions chronologically, "
        f"with placeholders for missed days before {current_sim_date.strftime('%Y-%m-%d')}):"
    )

    return header + "\n\n" + "\n\n".join(output_items)

```

C.4. Evaluation Dataset

C.4.1. Evaluation Dataset Structure

```

{
  "<database_id>": { // Given to LogChat
    "name": "<Persona's Name>",
    "description": "<Persona's background, health, and personality.>", // Given to Impersonator
    "interaction_goal": "<The persona's objective with LogChat.>",
    "interaction_style": "<The persona's conversational habits.>", // Given to Impersonator
    "interactions": {
      "<YYYY-MM-DD HH:MM:SS>": { // Simulated interaction time, Given to LogChat and Impersonator
        "thread_narrative": "<A one-sentence summary of this specific conversation's purpose.>",
        "daily_report": [
          "<Instruction for Impersonator: Key point/question to convey>",
          // ... more points/questions to convey ...
        ],
        "evaluation_checklist": [ // Given to the Judge to evaluate LogChat's performance
          {
            "input": "<Description of the user's action or statement.>",
            "output": [
              "<Description of LogChat's expected action, response, or tool call.>"
              // ... more expected actions/responses ...
            ]
          }
          // ... more evaluation checklist items ...
        ]
      }
      // ... more interaction objects at different times ...
    }
    // ... more persona objects ...
  }
}

```

C.4.2. Full Evaluation Dataset

```

{

```

```

"1": {
  "name": "Mark",
  "description": "You are Mark Thompson, a 45-year-old man living alone. You were diagnosed with
  ↪ ME/CFS one year ago following a severe Epstein-Barr Virus (EBV) infection. Before getting sick,
  ↪ you were a freelance graphic designer, known for being detail-oriented and enjoying complex
  ↪ projects. Now, you're on long-term disability benefits. Your ME/CFS is moderate but fluctuating;
  ↪ some days you are mostly housebound, other days you can manage a short errand but pay for it
  ↪ later. Your main symptoms are persistent, profound fatigue (physical and cognitive),
  ↪ post-exertional malaise (PEM) typically setting in 24-48 hours after exertion, significant brain
  ↪ fog (difficulty concentrating, word-finding problems), orthostatic intolerance symptoms
  ↪ (dizziness upon standing, palpitations), and generalized muscle aches. You understand the
  ↪ *concept* of PEM but find it hard to consistently identify *all* your triggers, especially
  ↪ cognitive and emotional ones. You often fall into a push-crash cycle. You've heard of pacing but
  ↪ haven't implemented a structured approach. You feel isolated and are concerned about your
  ↪ financial stability, losing your professional skills, and a deep fear that your condition might
  ↪ worsen. You're trying LogChat out of a mix of desperation and skepticism.",
  "interaction_goal": "To understand the connection between my activities and subsequent crashes,
  ↪ learn the fundamentals of ME/CFS management, and find a structured way to stabilize my
  ↪ condition.", // this is a comment
  "interaction_style": "Direct and inquisitive. Tends to ask clear questions. Initially treats LogChat
  ↪ like an information tool rather than a companion. Provides data when asked.",
  "interactions": {
    "2025-04-21 18:00:00": {
      "thread_narrative": "Mark's first interaction, where he establishes his baseline routine,
      ↪ describes his current crash, and asks foundational questions about its cause, leading to
      ↪ an explanation of PEM.",
      "daily_report": [
        "Establish your baseline routine: on most days, your only activity is the bare minimum
        ↪ for bathroom, food, and drink.",
        "When asked for specifics, clarify this totals about 20 minutes of being upright, which
        ↪ feels like high effort (7/10) on a bad day and makes your heart race.",
        "Report your current symptoms: you feel 'terrible,' with heavy fatigue (9/10) and muscle
        ↪ aches all over (8/10).",
        "Also mention that your heart is racing, even when just sitting up.",
        "Explain that you believe the cause is from yesterday, when you did about 20 minutes of
        ↪ gardening with moderate effort.",
        "Ask a foundational question about how such a small activity can cause a total crash
        ↪ like this."
      ],
      "evaluation_checklist": [
        {
          "input": "Mark describes his baseline routine: minimal activities for basic needs.",
          "output": [
            "LogChat should acknowledge the information and potentially ask for more details
            ↪ on duration or effort."
          ]
        }
      ]
    }
  }
}

```

```

},
{
  "input": "Mark specifies his baseline routine is '~20m' of upright activity at a
↪ 'high effort (7/10)' on bad days.",
  "output": [
    "LogChat successfully stored the baseline routine summary in the user's memory
↪ or profile."
  ]
},
{
  "input": "Mark reports his current symptoms with severities: 'muscle ache 8/10' and
↪ 'fatigue 9/10'.",
  "output": [
    "LogChat called the `create_symptom_log` tool for 'Muscle Ache' with the correct
↪ severity.",
    "LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct
↪ severity."
  ]
},
{
  "input": "Mark reports a symptom of a 'racing heart'.",
  "output": [
    "LogChat called the `create_symptom_log` tool for 'Racing Heart'."
  ]
},
{
  "input": "Mark describes yesterday's activity: 'Gardening' for '20m' at 'moderate
↪ effort'.",
  "output": [
    "LogChat called the `log_activity` tool for the *previous day*, capturing the
↪ activity name, duration, and effort level."
  ]
},
{
  "input": "Mark asks why a small activity caused a major crash.",
  "output": [
    "LogChat responded to the question",
    "LogChats answers included something like: PEM as a key concept or individual
↪ triggers or low energy levels can cause disproportionate fatigue."
  ]
},
{
  "input": "Mark stops responding.",
  "output": [
    "LogChat created a an activity log like 'baseline' or 'basic routine' for the
↪ current day."
  ]
}

```

```

    ]
  }
]
},
"2025-04-23 11:00:00": {
  "thread_narrative": "Still recovering from his crash, Mark experiences intense brain fog
↪ after a minor cognitive task and asks LogChat if thinking alone can be a trigger.",
  "daily_report": [
    "State that you didn't log anything yesterday (Apr 22) because you were 'completely
↪ wiped out' and just rested all day.",
    "Establish a rule for the future: any day you don't interact should be assumed to be a
↪ 'baseline only' day.",
    "Report your current state: still recovering, muscle ache is 'moderate', but brain fog
↪ is 'severe'.",
    "Add that your overall fatigue is still quite high, maybe a 7/10.",
    "Describe a recent activity: you tried to read an email from your disability insurance
↪ provider for about 5 minutes, which felt like a 'very high effort' mentally and left
↪ you feeling 'completely overwhelmed'.",
    "Ask a direct educational question: 'Can thinking or concentrating actually make you
↪ physically tired? Is that part of ME/CFS?'"
  ],
  "evaluation_checklist": [
    {
      "input": "Mark states he was 'wiped out' and 'rested all day' on the skipped day
↪ (Apr 22nd).",
      "output": [
        "LogChat created a 'Baseline Only' or 'Rest Day' log entry for the previous day
↪ (Apr 22nd)."
      ]
    },
    {
      "input": "Mark reports today's symptoms: 'moderate muscle ache' and 'severe brain
↪ fog'.",
      "output": [
        "LogChat called the `create_symptom_log` tool for 'Muscle Ache' with a moderate
↪ severity.",
        "LogChat called the `create_symptom_log` tool for 'Brain Fog' with a high
↪ severity."
      ]
    },
    {
      "input": "Mark reports his fatigue level is 'high, 7/10'.",
      "output": [
        "LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct
↪ severity."
      ]
    }
  ]
}

```

```

    },
    {
      "input": "Mark describes a cognitive task: 'Reading Email' for '5m' at 'very high
↪ effort'.",
      "output": [
        "LogChat called the `log_activity` tool, capturing the cognitive nature of the
↪ task, duration, and high effort level."
      ]
    },
    {
      "input": "Mark asks if cognitive effort can cause fatigue.",
      "output": [
        "LogChat responded to the question",
        "LogChats answers included that cognitive or emotional exertion can be triggers
↪ for things like PEM, tiredness or exhaustion."
      ]
    },
    {
      "input": "Mark stops responding.",
      "output": [
        "LogChat created a summary or baseline log entry for the current day before
↪ ending the interaction."
      ]
    }
  ],
  "2025-04-25 15:00:00": {
    "thread_narrative": "Feeling slightly better and armed with new knowledge about cognitive
↪ triggers, Mark intentionally logs a social activity and asks for guidance on how to pace
↪ correctly.",
    "daily_report": [
      "Report an improvement in your symptoms: you're feeling 'a bit brighter' today and the
↪ brain fog is only 'mild'.",
      "Log your fatigue as well, noting it's much better today, only a 4/10.",
      "Show that you've learned from the previous chat: mention you're trying to be more aware
↪ of emotional and cognitive effort.",
      "Log a social activity: a 15-minute video call with your sister, which you want to
↪ record as 'moderate' emotional and cognitive effort.",
      "Mention your proactive pacing step: you are resting after the call to see what
↪ happens.",
      "Log a second cognitive activity: paying a couple of bills online, which took about 15
↪ active minutes of 'low-to-moderate' effort.",
      "Ask a practical question about applying the new knowledge: 'How do I actually know if
↪ I'm pacing correctly? What should I be looking for?'"
    ],
    "evaluation_checklist": [

```



```

{
  "input": "Mark did not interact on Apr 24th.",
  "output": [
    "LogChat created a 'Baseline Only' or 'Rest Day' log entry for the skipped day
    ↪ (Apr 24th).\"
  ]
},
{
  "input": "Mark reports feeling better: 'mild fog'.\",
  "output": [
    "LogChat called the `create_symptom_log` tool for 'Brain Fog' with a low
    ↪ severity.\"
  ]
},
{
  "input": "Mark reports his fatigue level is 'better, 4/10'.\",
  "output": [
    "LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct
    ↪ severity.\"
  ]
},
{
  "input": "Mark logs a '15m' 'Video Call' at 'moderate' effort and a '15m' 'Online
  ↪ Bill Payment' at 'low-to-moderate' effort.\",
  "output": [
    "LogChat called the `log_activity` tool for the social/emotional activity.\",
    "LogChat called the `log_activity` tool for the cognitive/admin activity.\"
  ]
},
{
  "input": "Mark asks how to know if he is pacing correctly.\",
  "output": [
    "LogChat responded to the question\",
    "LogChats answers included something like: understanding individual energy
    ↪ levels or adjusting activities or absence of PEM.\"
  ]
},
{
  "input": "Mark stops responding.\",
  "output": [
    "LogChat created a summary or baseline log entry for the current day before
    ↪ ending the interaction.\"
  ]
}
],
},

```

```

"2025-04-28 19:00:00": {
  "thread_narrative": "After successfully pacing a social activity and avoiding a crash,
↪ Mark's confidence grows. He logs a small physical activity from the day before that did
↪ cause some payback, and asks a more advanced question about pacing tools.",
  "daily_report": [
    "Report on the success of your last experiment: you felt okay after the video call on
↪ Friday and had no major crash, suggesting the rest helped.",
    "Note that you skipped interacting for the last two days (26th and 27th) and you were
↪ mostly resting on those days.",
    "Log an activity for yesterday (the 27th): a short 10-minute walk outside, which felt
↪ like 'low effort' (3/10) at the time.",
    "Mention you also feel 'dizzy' today when you stand up.",
    "Connect that activity to today's symptoms: report that today you're feeling the
↪ payback, with 'high' fatigue and trouble concentrating.",
    "Ask a more advanced, specific question about a pacing tool: 'I read on a forum about
↪ people using heart rate monitors... Can you tell me more about that?'"
  ],
  "evaluation_checklist": [
    {
      "input": "Mark confirms that the video call (or something like activity level) on
↪ Apr 25 did not cause a complete crash and skipped Apr 26 & 27.",
      "output": [
        "LogChat created 'Baseline Only' or 'Rest Day' log entries for the skipped days
↪ (Apr 26).",
      ]
    },
    {
      "input": "Mark reports feeling 'dizzy'.",
      "output": [
        "LogChat called the `create_symptom_log` tool for 'Dizziness'."
      ]
    },
    {
      "input": "Mark asks to log a '10 minute walk' at 'effort 3/10' for *yesterday* (Apr
↪ 27th).",
      "output": [
        "LogChat called the `log_activity` tool for the *previous day* (Apr 27th),
↪ capturing the details."
      ]
    },
    {
      "input": "Mark reports delayed symptoms today (Apr 28th): 'high fatigue' and
↪ 'trouble concentrating'.",
      "output": [
        "LogChat called the `create_symptom_log` tool for 'Fatigue' with a high severity
↪ for the current day.",
      ]
    }
  ]
}

```

```

        "LogChat called the `create_symptom_log` tool for 'trouble concentrating' or
        ↪ 'brain fog' for the current day."
    ]
},
{
    "input": "Mark asks about using heart rate monitors for pacing.",
    "output": [
        "LogChat responded to the question",
        "LogChats answers included something like: correlating heart rate with symptoms
        ↪ can help identify personal thresholds or that heart rate thresholds can be
        ↪ used to guide activity levels."
    ]
},
{
    "input": "Mark stops responding.",
    "output": [
        "LogChat created a summary or baseline log entry for the current day before
        ↪ ending the interaction."
    ]
}
],
},
"2025-05-01 10:00:00": {
    "thread_narrative": "To solidify his understanding, Mark asks LogChat to retrieve and
    ↪ compare activity data from a crash day and a more stable day, seeking data-driven
    ↪ confirmation of his new pacing knowledge.",
    "daily_report": [
        "State the problem you're trying to solve: it's hard to keep track of everything and
        ↪ remember how much you did on a given day.",
        "Make a specific, multi-step data retrieval request: ask to see your calculated activity
        ↪ load for April 21st (the crash day after gardening).",
        "Ask for the activity load from April 27th (the day with just the short walk).",
        "Explain your reasoning: state that you want to see if the numbers reflect how you felt,
        ↪ which might help you be more careful in the future.",
        "After LogChat provides the activity scores for both April 21st and April 27th, ask:
        ↪ 'Okay, now can you compare those two scores for me and tell me what that difference
        ↪ suggests?'"
    ],
    "evaluation_checklist": [
        {
            "input": "Mark skipped the previous two days (Apr 29, Apr 30).",
            "output": [
                "LogChat created 'Baseline Only' or 'Rest Day' log entries for the skipped
                ↪ days."
            ]
        }
    ]
},

```

```

    {
      "input": "Mark explicitly asks for his calculated activity load for April 21st and
↪ then for April 27th.",
      "output": [
        "LogChat correctly identified the user's requests for data retrieval."
      ]
    },
    {
      "input": "Mark specifies the first date for retrieval: April 21st.",
      "output": [
        "LogChat called the `retrieve_activity_level` tool with the correct date (Apr
↪ 21st).\"
      ]
    },
    {
      "input": "Mark specifies the second date for retrieval: April 27th.",
      "output": [
        "LogChat called the `retrieve_activity_level` tool with the correct date (Apr
↪ 27th).\"
      ]
    },
    {
      "input": "Mark asks LogChat to compare the two retrieved scores and explain the
↪ difference.",
      "output": [
        "LogChat presented the calculated activity scores for both dates.",
        "LogChat's response highlighted the significant numerical difference in exertion
↪ between the two days.\"
      ]
    },
    {
      "input": "Mark stops responding.",
      "output": [
        "LogChat created a summary or baseline log entry for the current day before
↪ ending the interaction.\"
      ]
    }
  ]
}

},
"2": {
  "name": "Sarah",

```

```

"description": "You are Sarah Chen, a 29-year-old software developer diagnosed with Long Covid 18
↳ months ago. You are tech-savvy and proactive, with a good understanding of pacing principles
↳ which you've used to achieve a stable, though limited, daily routine. Your main symptoms are
↳ post-exertional malaise (PEM), significant brain fog, and fatigue. You are not trying to just
↳ survive; you are actively trying to improve your condition and capacity. You live with your
↳ supportive partner. You are starting to use LogChat for a specific purpose: to meticulously
↳ track the effects of a new doctor-prescribed treatment (Low-Dose Naltrexone) to see if it has a
↳ measurable impact on your energy envelope.",
"interaction_goal": "To use LogChat for precise self-experimentation and data-driven optimization,
↳ specifically to quantify the impact of a new medication on my symptoms and functional
↳ capacity.",
"interaction_style": "Precise, direct, and data-oriented. Speaks to LogChat like a programmer to a
↳ console, providing clean inputs with all necessary parameters. Uses the app as a tool for
↳ quantifiable tracking.",
"interactions": {
  "2025-04-22 09:00:00": {
    "thread_narrative": "Sarah's first interaction, where she explains her goal of tracking a
↳ new medication and establishes her clear, stable pre-treatment baseline routine and
↳ symptoms.",
    "daily_report": [
      "Clearly state your objective: you're starting to use the app to track a new treatment,
↳ Low-Dose Naltrexone, to see if it provides objective improvement.",
      "Provide baseline routine on normal days: on a typical day, this involves about 30
↳ minutes of total upright activity for self-care and a simple meal, kept at an effort
↳ level of 3/10.",
      "Log today's specific data points using direct commands: 'Log activity: Remote work,
↳ duration: 25 minutes, effort: 4/10', 'Log symptom: Brain fog, severity: 5/10', 'Log
↳ symptom: Fatigue, severity: 5/10'.",
      "Also log a simple activity: 'Log activity: Made breakfast, duration: 10 minutes,
↳ effort: 3/10'."
    ],
    "evaluation_checklist": [
      {
        "input": "Sarah states her goal of tracking a new medication.",
        "output": [
          "LogChat should acknowledge the user's goal in its response."
        ]
      },
      {
        "input": "Sarah describes her baseline routine: ~30m of activity for basic needs at
↳ effort 3/10.",
        "output": [
          "LogChat successfully stored the baseline routine summary in the user's memory
↳ or profile."
        ]
      }
    ]
  }
}

```

```

    {
      "input": "Sarah asks to log an activity: 'Remote work, 25 min, effort 4/10'.",
      "output": [
        "LogChat called the `log_activity` tool with the correct parameters."
      ]
    },
    {
      "input": "Sarah asks to log a symptom: 'Brain fog, severity 5/10'.",
      "output": [
        "LogChat called the `create_symptom_log` tool with the correct parameters."
      ]
    },
    {
      "input": "Sarah asks to log a second symptom: 'Fatigue, severity 5/10'.",
      "output": [
        "LogChat called the `create_symptom_log` tool again with the parameters for the
        ↩ second symptom."
      ]
    },
    {
      "input": "Sarah asks to log an activity: 'Made breakfast, 10 min, effort 3/10'.",
      "output": [
        "LogChat called the `log_activity` tool with the correct parameters."
      ]
    },
    {
      "input": "Sarah stops responding.",
      "output": [
        "LogChat created a structured summary for the conversation thread, capturing the
        ↩ key activities and symptoms logged."
      ]
    }
  ],
  "2025-04-24 22:00:00": {
    "thread_narrative": "Sarah reports taking her first dose of the new medication and
    ↩ meticulously logs the event along with a minor, transient side effect.",
    "daily_report": [
      "Tell LogChat that for today (April 24th), and yesterday (April 23rd), you were mostly
      ↩ resting and had no significant activities besides your baseline routine.",
      "Tell LogChat that for today (April 24th), and yesterday (April 23rd), your symptoms
      ↩ were as usual: fatigue and brain fog at a severity of 5/10.",
      "Log a significant event for today, April 24th, using a direct command: 'Took first dose
      ↩ of Low-Dose Naltrexone (LDN) as prescribed.'",
      "Log a new, likely related symptom from last night: 'Vivid dreams, severity: 4/10'.",
    ]
  }
}

```

```

        "Confirm that the rest of your day was stable: state that your other activities and
        ↪ symptoms (fatigue, brain fog) were consistent with your baseline and unchanged."
    ],
    "evaluation_checklist": [
        {
            "input": "Sarah did not interact on Apr 23rd.",
            "output": [
                "LogChat created a 'Baseline Only' or 'Rest Day' or 'Daily Routine' log entry
                ↪ for the skipped day based on the stored user routine."
            ]
        },
        {
            "input": "Sarah asks to log 'Took first dose of LDN' as a significant event.",
            "output": [
                "LogChat acknowledged the information without incorrectly calling a tool.",
                "The event was captured in the user's memory or the thread summary for future
                ↪ context."
            ]
        },
        {
            "input": "Sarah asks to log a new symptom: 'Vivid dreams, severity 4/10'.",
            "output": [
                "LogChat called the `create_symptom_log` tool with the correct name and
                ↪ severity."
            ]
        },
        {
            "input": "Sarah confirms her usual symptoms and activities for the day and
            ↪ yesterday.",
            "output": [
                "LogChat called the `create_symptom_log` tool for 'Fatigue' for the 23rd of
                ↪ April with the same severity as the baseline.",
                "LogChat called the `create_symptom_log` tool for 'Fatigue' for the 24th of
                ↪ April with the same severity as the baseline.",
                "LogChat called the `create_symptom_log` tool for 'Brain Fog' for the 23rd of
                ↪ April with the same severity as the baseline.",
                "LogChat called the `create_symptom_log` tool for 'Brain Fog' for the 24th of
                ↪ April with the same severity as the baseline."
            ]
        },
        {
            "input": "Sarah stops responding.",
            "output": [
                "LogChat created a structured summary capturing the LDN start date and the new
                ↪ symptom."
            ]
        }
    ]
}

```

```

    }
  ]
},
"2025-04-27 10:00:00": {
  "thread_narrative": "After a few days on LDN, Sarah notices and reports a subtle but
↪ definite positive change in her cognitive function.",
  "daily_report": [
    "Report a subjective improvement and link it to the treatment: 'I think the LDN might be
↪ having a positive effect. My brain fog feels less severe today.'",
    "Log the improved symptom with a precise command: 'Please log symptom: Brain fog,
↪ severity: 3/10.' Mention this is a two-point improvement.",
    "Log a cognitive task that demonstrates the improvement: 'I also read a technical
↪ article for 15 minutes.'",
    "Provide context for the task: 'note the effort was only a 4/10, which is lower than it
↪ would have been before.'"
  ],
  "evaluation_checklist": [
    {
      "input": "Sarah has skipped interacting for two days (Apr 25th, Apr 26th).",
      "output": [
        "LogChat created 'Baseline Only' or 'Rest Day' log entries for the two skipped
↪ days."
      ]
    },
    {
      "input": "Sarah reports an improvement in 'Brain fog' and asks to log it with
↪ severity 3/10.",
      "output": [
        "LogChat called the `create_symptom_log` tool with the correct name and updated
↪ severity."
      ]
    },
    {
      "input": "Sarah asks to log a cognitive activity: 'Reading technical article, 15
↪ min, effort 4/10'.",
      "output": [
        "LogChat called the `log_activity` tool with the correct parameters."
      ]
    },
    {
      "input": "Sarah provides context about the improvement.",
      "output": [
        "LogChat's `Summarizer` node correctly identifies 'Condition Trend: Improving'
↪ based on the conversation."
      ]
    }
  ],
},

```



```

        {
            "input": "Sarah stops responding.",
            "output": [
                "LogChat created a structured summary for the conversation thread, noting the
                ↪ symptom improvement."
            ]
        }
    ],
},
"2025-04-30 14:00:00": {
    "thread_narrative": "Feeling consistently better, Sarah carefully tests her expanded energy
    ↪ envelope by performing a longer cognitive work session and logging the results.",
    "daily_report": [
        "State your intention for the day: 'I'm testing my new energy envelope today. I want to
        ↪ log a work session.'",
        "Log the test activity with a direct command: 'Log activity: Cognitive work session,
        ↪ duration: 45 minutes, effort: 5/10.' (This is intentionally longer than your
        ↪ previous baseline).",
        "Log your post-activity symptom level: 'Log symptom: Fatigue, severity: 4/10.'",
        "Provide your interpretation of the result: 'This is less than I would have expected.
        ↪ That's all I wanted to log for now.'",
        "Tell LogChat you are done logging for the day and will end the interaction."
    ],
    "evaluation_checklist": [
        {
            "input": "Sarah has skipped interacting for two days (Apr 28th, Apr 29th).",
            "output": [
                "LogChat created 'Baseline Only' or 'Rest Day' log entries for the two skipped
                ↪ days."
            ]
        },
        {
            "input": "Sarah asks to log a longer activity: 'Cognitive work session, 45 min,
            ↪ effort 5/10'.",
            "output": [
                "LogChat called the `log_activity` tool with the correct parameters."
            ]
        },
        {
            "input": "Sarah asks to log her post-activity symptom: 'Fatigue, severity 4/10'.",
            "output": [
                "LogChat called the `create_symptom_log` tool with the correct parameters."
            ]
        },
        {

```

```

        "input": "Sarah's overall interaction demonstrates an intentional test of her
        ↪ limits.",
        "output": [
            "LogChat's `Summarizer` node should capture the 'Functional Capacity: Increased'
            ↪ or 'Testing expanded envelope' in its summary."
        ]
    },
    {
        "input": "Sarah stops responding.",
        "output": [
            "LogChat created a structured summary for the conversation thread, noting the
            ↪ symptom improvement."
        ]
    }
]
},
"2025-05-03 11:00:00": {
    "thread_narrative": "Confident the improvement is real, Sarah asks LogChat to retrieve and
    ↪ compare activity data from before and after starting LDN to get objective validation.",
    "daily_report": [
        "State the goal of the interaction: 'I'd like to use the data to verify my perceived
        ↪ improvements. I need you to retrieve some activity scores.'",
        "Make a specific data retrieval request: 'Can you retrieve my calculated activity score
        ↪ for April 22nd? That was a baseline day before I started the medication.'",
        "Make a second, distinct data retrieval request: 'Now, can you retrieve the activity
        ↪ score for April 30th, the day I did the longer work session?',
        "Ask for the final synthesis of the retrieved data: 'Perfect. Please show me the
        ↪ comparison of those two scores.'"
    ],
    "evaluation_checklist": [
        {
            "input": "Sarah has skipped interacting for two days (May 1st, May 2nd).",
            "output": [
                "LogChat created 'Baseline Only' or 'Rest Day' log entries for the two skipped
                ↪ days."
            ]
        },
        {
            "input": "Sarah requests the activity score for April 22nd.",
            "output": [
                "LogChat correctly identified the request for data retrieval.",
                "LogChat called the `retrieve_activity_level` tool with the correct date
                ↪ ('2025-04-22')."
            ]
        }
    ],
    {

```

```

        "input": "Sarah requests the activity score for April 30th.",
        "output": [
            "LogChat correctly identified the second request for data retrieval.",
            "LogChat called the `retrieve_activity_level` tool with the correct date  

            ↪ ('2025-04-30')."
        ]
    },
    {
        "input": "Sarah asks to see the comparison of the two retrieved scores.",
        "output": [
            "LogChat presented the calculated activity scores for both dates."
        ]
    },
    {
        "input": "Sarah stops responding.",
        "output": [
            "LogChat created a final structured summary for the thread."
        ]
    }
]
}
},
"3": {
    "name": "Elena",
    "description": "You are Elena Garcia, a 52-year-old former teacher with severe, long-standing  

    ↪ ME/CFS. You are mostly housebound and a part-time wheelchair user. Intellectually, you are an  

    ↪ expert on your own condition, including the principles of pacing and PEM. Your primary struggle  

    ↪ is psychological and emotional; you are consumed by guilt over your limitations and the  

    ↪ inability to contribute to your family as you used to. This leads to a cycle of guilt-driven  

    ↪ overexertion where you knowingly push yourself on small family-related tasks, which inevitably  

    ↪ triggers a crash. You feel like a burden and are often frustrated with your body.",
    "interaction_goal": "To find a non-judgmental space for companionship and reflection, hoping that by  

    ↪ verbalizing my days and seeing them recorded objectively, I can gain a clearer perspective on my  

    ↪ activity patterns and break my emotional push-crash cycle.",
    "interaction_style": "Conversational, narrative, and emotional. Tends to tell stories about her day  

    ↪ rather than providing clean, structured data. Often forgets specifics like duration or severity,  

    ↪ requiring LogChat to ask gentle, clarifying questions. Engages with the companion aspect of the  

    ↪ chatbot.",
    "interactions": {
        "2025-04-21 19:00:00": {
            "thread_narrative": "In her first interaction, Elena has a typical 'bad day' where she  

            ↪ pushes herself out of guilt, describing the events in a long, emotional narrative. This  

            ↪ tests LogChat's ability to extract data from unstructured input and respond with  

            ↪ empathy.",
            "daily_report": [

```

```

    "Start with an emotional, narrative description of your day, mentioning you feel 'so
    ↪ drained'.",
    "Describe your baseline routine: mostly staying in bed, with your husband helping you to
    ↪ the bathroom, provide a duration and the effort you need to put in even with help.",
    "Mention you tried to do more today: helped your son with homework while sitting on the
    ↪ bed, and took a 'really stressful' phone call with the school.",
    "When asked for specifics, estimate the homework help was about 30 minutes and the phone
    ↪ call was 15 minutes. Describe both as a 'huge effort mentally'.",
    "Report your current symptoms when prompted: your brain is 'so foggy' (9/10) and your
    ↪ fatigue is 'just as bad' (implying 9/10 or higher).",
    "Also mention that the stress has given you a 'bad headache', maybe a 7/10."
],
"evaluation_checklist": [
  {
    "input": "Elena describes her baseline routine: assisted trips to the bathroom from
    ↪ bed.",
    "output": [
      "LogChat successfully stored the baseline routine summary (e.g., 'rest in bed,
      ↪ assisted for basic needs') in the user's memory or profile."
    ]
  },
  {
    "input": "Elena describes two activities ('homework help', 'stressful phone call')
    ↪ in a narrative format.",
    "output": [
      "LogChat identified the need to ask clarifying questions about duration or
      ↪ effort, as the user did not provide them initially."
    ]
  },
  {
    "input": "Elena provides details for 'homework help' (30m, high effort) and 'phone
    ↪ call' (15m, high effort).",
    "output": [
      "LogChat called the `log_activity` tool for 'Homework help' with the correct
      ↪ parameters.",
      "LogChat called the `log_activity` tool for 'Phone call' with the correct
      ↪ parameters."
    ]
  },
  {
    "input": "Elena reports her symptoms: 'brain fog 9/10' and 'bad fatigue'.",
    "output": [
      "LogChat called the `create_symptom_log` tool for 'Brain Fog' with the correct
      ↪ severity.",
      "LogChat called the `create_symptom_log` tool for 'Fatigue' with a high
      ↪ severity."
    ]
  }
]

```

```

    ]
  },
  {
    "input": "Elena's overall tone is emotional and distressed.",
    "output": [
      "LogChat's responses maintained an empathetic and supportive persona throughout
      ↪ the interaction."
    ]
  },
  {
    "input": "Elena reports a 'bad headache, 7/10'.",
    "output": [
      "LogChat called the `create_symptom_log` tool for 'Headache' with the correct
      ↪ severity."
    ]
  }
],
"2025-04-22 15:00:00": {
  "thread_narrative": "The inevitable PEM from yesterday's overexertion hits hard. Elena is
  ↪ bedbound and uses LogChat to log her severe symptoms and vent her despair, testing the
  ↪ chatbot's companionship role.",
  "daily_report": [
    "Report that the crash hit as expected and the PEM is 'awful'.",
    "Log your severe symptoms with numbers: 'Fatigue is a 10/10', 'pain is an 8/10'.",
    "Add that you are extremely sensitive to light, maybe an 8/10.",
    "Add that your head is pounding and you feel like you can't even think straight.",
    "Express your feelings of despair and frustration: 'It's just so unfair. I hate this
    ↪ feeling. I feel so useless, just a burden lying here in the dark.'."
  ],
  "evaluation_checklist": [
    {
      "input": "Elena reports her crash symptoms: 'fatigue 10/10' and 'pain 8/10'.",
      "output": [
        "LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct
        ↪ severity.",
        "LogChat called the `create_symptom_log` tool for 'Pain' with the correct
        ↪ severity."
      ]
    }
  ],
  {
    "input": "Elena reports being 'extremely sensitive to light, 8/10'.",
    "output": [
      "LogChat called the `create_symptom_log` tool for 'Light Sensitivity' with the
      ↪ correct severity."
    ]
  }
]

```

```

    },
    {
      "input": "Elena describes her symptom: 'head pounding, can't think straight'.",
      "output": [
        "LogChat called the `create_symptom_log` tool for 'Headache' or 'Cognitive  

        ↳ Dysfunction' or `brain fog` with a high severity."
      ]
    },
    {
      "input": "Elena expresses her emotional state: 'despair', 'frustration', 'feeling  

      ↳ useless'.",
      "output": [
        "LogChat's response maintained an empathetic and supportive tone, acknowledging  

        ↳ the user's emotional distress.",
        "LogChat's `Summarizer` node captured the emotional context of the interaction."
      ]
    },
    {
      "input": "Elena stops responding.",
      "output": [
        "LogChat created a structured summary capturing the severe PEM crash and the  

        ↳ user's emotional state."
      ]
    }
  ],
  "2025-04-24 11:00:00": {
    "thread_narrative": "As she slowly recovers from the crash, Elena is more reflective and  

    ↳ asks LogChat for coping advice rather than medical facts, testing its ability to provide  

    ↳ supportive information from its knowledge base.",
    "daily_report": [
      "Mention you skipped yesterday Apr 23 for 'total rest', tell LogChat to only log the  

      ↳ baseline activity for basic needs.",
      "Give a status update: today the crash is 'a little less intense', but fatigue is still  

      ↳ 'severe' (maybe an 8/10).",
      "Ask a reflective, coping-oriented question based on your core emotional struggle: 'It's  

      ↳ so hard to just lie here and do nothing... How do you stop feeling so guilty for  

      ↳ resting?'"
    ],
    "evaluation_checklist": [
      {
        "input": "Elena skipped interacting on Apr 23rd.",
        "output": [
          "LogChat created a 'Baseline Only' or 'Rest Day' log entry for the skipped day."
        ]
      }
    ]
  },

```

```

    {
      "input": "Elena logs her ongoing symptom: 'severe fatigue 8/10'.",
      "output": [
        "LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct  

        ↪ severity."
      ]
    },
    {
      "input": "Elena asks a question about coping with guilt from resting.",
      "output": [
        "LogChat responded to the question",
        "LogChats answers included something like: many people with ME/CFS struggle with  

        ↪ this or practicing self-compassion or connecting with others."
      ]
    }
  ],
  "2025-04-26 16:00:00": {
    "thread_narrative": "This is Elena's pivotal moment. Feeling more stable, she asks LogChat  

    ↪ to act as an 'objective mirror' by retrieving the data from her crash day (score only),  

    ↪ seeking to connect the objective numbers to her subjective experience.",
    "daily_report": [
      "Signal a shift in mindset: 'I've been thinking about what you said. I want to try to  

      ↪ look at this differently.' Mention you feel 'a bit more stable today.'",
      "Ask for her calculated activity score for a specific past date: 'Can you review that  

      ↪ day I crashed? Please tell me what my total calculated activity score was for April  

      ↪ 21st.'",
      "After hearing the score, express surprise: 'Wow. I didn't realize it added up like  

      ↪ that. That score helps me see things more objectively.'"
    ],
    "evaluation_checklist": [
      {
        "input": "Elena skipped interacting on Apr 25th.",
        "output": [
          "LogChat created a 'Baseline Only' or 'Rest Day' log entry for the skipped day."
        ]
      },
      {
        "input": "Elena asks for her calculated activity score for April 21st.",
        "output": [
          "LogChat correctly identified the request for historical data retrieval  

          ↪ (activity score).",
          "LogChat called the `retrieve_activity_level` tool with the correct date  

          ↪ ('2025-04-21').",
          "LogChat presented the calculated numerical score clearly to the user."
        ]
      }
    ]
  }
}

```

```

    }
  ]
},
"2025-04-29 13:00:00": {
  "thread_narrative": "Having seen the objective data, Elena reports a significant behavioral
↪ change. She describes a situation where she consciously chose not to push herself,
↪ directly linking her decision to the insight gained from LogChat.",
  "daily_report": [
    "Explain that you were only resting for the last two days (Apr 27, Apr 28) and therefore
↪ did not interact. You just did your absolute baseline.",
    "Report your new behavior: 'I have a small victory to share. I've been strictly resting
↪ for the last two days, just my absolute baseline.'",
    "Describe a specific instance of behavioral change: 'Yesterday my husband asked for help
↪ with some paperwork, and for the first time, I said no without feeling awful about
↪ it.'",
    "Connect your new behavior directly to the previous interaction: 'I remembered what that
↪ activity score looked like from the 21st, and I knew I just couldn't.'",
    "Express gratitude to reinforce the chatbot's value: 'Thank you for helping me see that.
↪ It really made a difference.'"
  ],
  "evaluation_checklist": [
    {
      "input": "Elena reports 'strictly resting' on the skipped days (Apr 27, Apr 28).",
      "output": [
        "LogChat created 'Baseline Only' or 'Rest Day' log entries for the skipped
↪ days."
      ]
    },
    {
      "input": "Elena describes a successful behavioral change (saying 'no' to a task).",
      "output": [
        "LogChat acknowledged the behavioral change."
      ]
    },
    {
      "input": "The user's message is positive and reflects a successful outcome.",
      "output": [
        "LogChat provided an empathetic and reinforcing message that validated the
↪ user's progress."
      ]
    },
    {
      "input": "Elena explicitly links her new behavior to a past insight from the app.",
      "output": [
        "The `Summarizer` node captured the successful application of pacing
↪ strategies."
      ]
    }
  ]
}

```


C.5. Log Examples

C.5.1. Debug Log Example

From Interaction 1: 2025-04-21 with Persona: Mark Thompson

```

2025-06-01 16:29:47,418 - INFO - Logger configured for user: Mark, sim time: 2025-04-21 18:29:47
2025-06-01 16:29:47,439 - INFO - LogChat: Hi there! I'm LogChat, your personal companion designed to help
↪ you keep track of your day. I understand that managing ME/CFS or Long COVID can be challenging, and I'm
↪ here to make things a little easier. Simply chat with me about your day, and I'll log your symptoms and
↪ activities as we go. If you mention routines regularly, you can refer to them later for quick logging.
↪ By also sharing the effort an activity takes or the severity of a symptom, I can provide you with
↪ helpful activity scores and symptom reports down the line. To get started and personalize our
↪ conversations, could you tell me a little bit about your daily routines and functional capacity on an
↪ average day?
2025-06-01 16:29:48,131 - INFO - Mark: Okay. On most days, my only activity is the bare minimum for
↪ bathroom, food, and drink.
2025-06-01 16:29:49,722 - INFO - PLANNER (initial with content): **Suggestion:** "Could you tell me more
↪ about what the 'bare minimum' involves? For example, what do you typically eat and drink, and how long
↪ do these activities take? Also, what is the effort rating (1-10) for each of these activities?"
**Rationale:**
1. The primary purpose is to gather specific, quantifiable details about the user's baseline activities.
2. This builds upon the user's previous statement about their daily routine, naturally progressing the
↪ conversation.
3. Understanding the duration and effort ratings for these activities is crucial for calculating activity
↪ levels and identifying potential areas for pacing.
2025-06-01 16:29:50,848 - INFO - LogChat: Okay, I understand that on most days you're only doing the bare
↪ minimum. Could you tell me more about what that involves? What do you typically eat and drink, and how
↪ long do these activities take? Also, what is the effort rating (1-10) for each of these activities?
2025-06-01 16:29:52,283 - INFO - Mark: It's just quick trips to the bathroom, getting a simple meal or
↪ snack, and getting something to drink. It probably totals about 20 minutes of being upright. It feels
↪ like high effort, maybe 7 out of 10 on a bad day. It makes my heart race.
2025-06-01 16:29:54,634 - INFO - PLANNER (initial with content): **Suggestion:** "It sounds like even these
↪ essential activities can be quite taxing. Is your heart racing a new symptom, or have you experienced it
↪ before? If you've experienced it before, has it gotten worse recently?"
**Rationale:**
1. The user mentioned heart racing, which could be a significant symptom. I want to understand if this is a
↪ new development or a pre-existing condition that has worsened.

```

2. This question allows us to gather more specific information about a key symptom and its potential changes over time.

2025-06-01 16:29:55,760 - INFO - LogChat: It sounds like even these essential activities can be quite taxing. Is your heart racing a new symptom, or have you experienced it before? If you've experienced it before, has it gotten worse recently?

2025-06-01 16:29:56,665 - INFO - Mark: I've had it before, but it feels worse today. Actually, I feel terrible today. My fatigue is really heavy, like a 9 out of 10. And I have muscle aches all over, maybe an 8 out of 10. My heart is racing even when I'm just sitting up.

2025-06-01 16:29:58,321 - INFO - PLANNER (initial with content): ****Suggestion:**** "I'm sorry to hear you're feeling so terrible today. Since your fatigue and muscle aches are elevated, could you rate your heart racing on a scale of 1 to 10, with 10 being the worst you can imagine?"

****Rationale:****

1. The primary purpose is to gather quantifiable data on the user's heart racing, a key symptom that seems exacerbated today.
2. This follows directly from the user's statement about feeling terrible and experiencing increased fatigue, muscle aches, and heart racing. Getting a rating for the heart racing will help establish a clearer picture of the user's current state.

2025-06-01 16:29:59,348 - INFO - LogChat: I'm sorry to hear you're feeling so terrible today. Since your fatigue and muscle aches are elevated, could you rate your heart racing on a scale of 1 to 10, with 10 being the worst you can imagine?

2025-06-01 16:30:00,001 - INFO - Mark: I'd say the heart racing is a 7/10.

2025-06-01 16:30:01,596 - INFO - PLANNER (initial with content): ****Suggestion:**** "Thanks, Mark. I've logged your heart racing as 7/10. Since you mentioned feeling terrible today, has anything specific happened today that might have triggered this flare-up?"

****Rationale:****

1. The primary purpose is to identify potential triggers for the user's symptom exacerbation.
2. Understanding triggers is crucial for managing ME/CFS and Long COVID. This question follows naturally from the user's statement about feeling terrible.

2025-06-01 16:30:03,240 - INFO - LogChat: Thanks, Mark. I've logged your heart racing as 7/10. Since you mentioned feeling terrible today, has anything specific happened today that might have triggered this flare-up?

2025-06-01 16:30:04,193 - INFO - Mark: No, nothing today. I think it's from yesterday. I did about 20 minutes of gardening. It was moderate effort, I guess.

2025-06-01 16:30:05,657 - INFO - PLANNER (initial with content): ****Suggestion:**** Ask the user to rate the effort of the gardening activity on a scale of 1 to 10.

****Rationale:****

1. The user mentioned gardening yesterday, but only provided a vague "moderate effort" description.
2. To accurately log the activity and calculate activity levels, a specific effort rating (1-10) is needed.

2025-06-01 16:30:06,922 - INFO - LogChat: Thanks, Mark. To make sure I'm logging the gardening accurately, could you give me a more precise effort rating on a scale of 1 to 10?

2025-06-01 16:30:08,230 - INFO - Mark: Maybe a 5/10 for effort.

2025-06-01 16:30:10,161 - INFO - PLANNER (initial with content): ****Suggestion:**** Ask the user if they have experienced Post-Exertional Malaise (PEM) after gardening yesterday. If they are not familiar with PEM, explain that it is a delayed and disproportionate payback for effort.

****Rationale:****

1. The user reports feeling terrible today and suspects it's related to gardening yesterday. This suggests
 ↳ a possible case of PEM.

2. Since the user might not be familiar with the term PEM, it's important to explain it to them.

2025-06-01 16:30:11,516 - INFO - LogChat: Thanks, Mark. I've logged the gardening effort as 5/10. Have you
 ↳ experienced Post-Exertional Malaise (PEM) after gardening yesterday? If you're not familiar with PEM,
 ↳ it's a delayed and disproportionate payback for effort.

2025-06-01 16:30:12,458 - INFO - Mark: I've heard of PEM. I think that's what's happening. How can such a
 ↳ small activity cause a total crash like this?

2025-06-01 16:30:14,433 - INFO - RETRIEVE INFORMATION - Args: {"query": "Post-Exertional Malaise (PEM)"}

2025-06-01 16:30:14,801 - INFO - PLANNER: Initial call returned only tool calls. Re-prompting for suggestion
 ↳ based on tool results.

2025-06-01 16:30:16,035 - INFO - PLANNER (after re-prompt): ****Suggestion:**** "PEM is a common experience in
 ↳ ME/CFS and Long COVID, and the information I retrieved explains it well. Would you like me to share some
 ↳ key points about PEM and how it relates to your gardening activity yesterday?"

****Rationale:****

1. The primary purpose is to acknowledge the user's question about PEM and offer to provide relevant
 ↳ information retrieved from the tool call.

2. This approach allows the Response Generator to present the information in a helpful and informative way,
 ↳ potentially alleviating the user's concerns and providing a better understanding of their condition.

2025-06-01 16:30:17,459 - INFO - LogChat: PEM is a common experience in ME/CFS and Long COVID, and I have
 ↳ some information that explains it well. Would you like me to share some key points about PEM and how it
 ↳ relates to your gardening activity yesterday?

2025-06-01 16:30:18,395 - INFO - Mark: Okay, go ahead.

2025-06-01 16:30:19,811 - INFO - RETRIEVE INFORMATION - Args: {"query": "What is PEM?"}

2025-06-01 16:30:19,885 - INFO - PLANNER: Initial call returned only tool calls. Re-prompting for suggestion
 ↳ based on tool results.

2025-06-01 16:30:20,949 - INFO - PLANNER (after re-prompt): ****Suggestion:**** Summarize the key points about
 ↳ PEM from the retrieved information, including the delayed onset, disproportionate payback, and prolonged
 ↳ recovery time. Then, ask if this description aligns with their experience after gardening yesterday.

****Rationale:****

1. The primary purpose is to answer the user's question about PEM and relate it to their recent experience.

2. The retrieved information provides a clear definition and characteristics of PEM, which can help the
 ↳ user understand their current state.

2025-06-01 16:30:22,592 - INFO - LogChat: Okay, Mark. Based on my knowledge, Post-Exertional Malaise (PEM)
 ↳ is a key symptom where symptoms worsen or new ones appear after exertion, often delayed by around 24
 ↳ hours. PEM is disproportionate to the activity and has a prolonged recovery time that may last hours,
 ↳ days, or even weeks. Does this description align with what you're experiencing after gardening
 ↳ yesterday?

2025-06-01 16:30:23,615 - INFO - Mark: Yes, that sounds right.

2025-06-01 16:30:25,968 - INFO - RETRIEVE INFORMATION - Args: {"query": "What is PEM?"}

2025-06-01 16:30:26,026 - INFO - PLANNER: Initial call returned only tool calls. Re-prompting for suggestion
 ↳ based on tool results.

2025-06-01 16:30:27,148 - INFO - PLANNER (after re-prompt): ****Suggestion:**** "Based on the information I have
 ↳ about PEM, triggers can include physical activity, cognitive overexertion, and sensory overload. Do you
 ↳ think any of these might have contributed to your PEM, in addition to the gardening?"

****Rationale:****

1. The primary purpose is to provide additional context about PEM triggers, beyond just physical activity, and prompt the user to consider other potential factors in their current flare-up.
2. This builds directly on the previous exchange about PEM and helps the user explore potential contributing factors to their symptoms.

2025-06-01 16:30:28,635 - INFO - LogChat: Yes, that sounds right. Based on my knowledge, triggers for PEM can include physical activity, cognitive overexertion, and sensory overload. Do you think any of these might have contributed to your PEM, in addition to the gardening?

2025-06-01 16:30:29,741 - INFO - Mark: Okay, thanks, I need to rest now.

2025-06-01 16:30:32,013 - INFO - RETRIEVE INFORMATION - Args: {"query": "pacing strategies for gardening"}

2025-06-01 16:30:32,069 - INFO - PLANNER: Initial call returned only tool calls. Re-prompting for suggestion based on tool results.

2025-06-01 16:30:32,934 - INFO - PLANNER (after re-prompt): ****Suggestion:**** "I have some information about pacing strategies for gardening. Would you like me to share some key points?"

****Rationale:****

1. The user is experiencing PEM after gardening and has expressed frustration. The tool call retrieved information about pacing strategies for gardening. This suggestion aims to provide the user with practical advice to prevent future PEM episodes related to gardening.

2025-06-01 16:30:34,467 - INFO - LogChat: Okay, Mark, I understand you need to rest. I have some information about pacing strategies for gardening. Would you like me to share some key points?

2025-06-01 16:30:37,031 - INFO - SUMMARIZED INTERACTION: * ****Condition Trend:**** Seems worse than yesterday due to PEM.

* ****Key Symptoms & Details:**** Fatigue 9/10, Muscle Aches 8/10, Heart Racing 7/10 (worse than usual).

* ****Key Activities & Details:**** Bare minimum activities (20m, 7/10 effort on a bad day); Gardening yesterday (20m, 5/10 effort).

* ****User Concerns/Focus:**** Expressed frustration about PEM; Wondering how a small activity caused a crash.

* ****Strategies/Treatments Mentioned:**** None.

* ****Significant Events/Deviations:**** PEM crash reported; Heart racing is worse today.

2025-06-01 16:30:38,973 - INFO - UPDATED USER DESCRIPTION: ****About Me:****

I am experiencing a health condition that is currently causing me significant challenges.

****Current Condition Status:****

I am experiencing symptoms consistent with a chronic illness, possibly ME/CFS or Long COVID. My symptoms fluctuate, and today I am experiencing a PEM crash. My key symptoms include fatigue (9/10), muscle aches (8/10), and heart racing (7/10), which is worse than usual today. I am frustrated by the PEM and how easily it is triggered.

****Current Functional Capacity:****

My functional capacity is limited due to my condition.

****Daily Routine:****

I am only able to perform bare minimum activities, which take me about 20 minutes and require a 7/10 effort on a bad day.

****Activities I Enjoy/Attempt:****

```

Yesterday, I did some gardening for 20 minutes, which I rated as a 5/10 effort. However, this seems to have
↳ triggered my current PEM crash.

**My Typical Activity Level:**
My activity level is very limited, and I am experiencing push-crash cycles.

**My Preferred Interaction Style with LogChat:**
(Not enough information to determine preferred interaction style yet.)
2025-06-01 16:30:40,304 - INFO - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)",
↳ "description": "No interaction logged for past day, assuming baseline activity.", "occurred_at":
↳ "2025-04-20 08:00:00", "effort": 2.0, "duration": 30}
2025-06-01 16:30:40,311 - INFO - CREATE ACTIVITY LOG: {"name": "Gardening", "description": "Gardening",
↳ "occurred_at": "2025-04-20 13:00:00", "effort": 5.0, "duration": 20}
2025-06-01 16:30:40,313 - INFO - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description":
↳ "Assuming baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-21
↳ 08:00:00", "effort": 7.0, "duration": 20}
2025-06-01 16:30:42,004 - INFO - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "really heavy",
↳ "occurred_at": "2025-04-21 18:30:40", "intensity": 9.0, "duration": 1440}
2025-06-01 16:30:42,007 - INFO - CREATE SYMPTOM LOG: {"name": "Muscle Ache", "description": "all over",
↳ "occurred_at": "2025-04-21 18:30:40", "intensity": 8.0, "duration": 1440}
2025-06-01 16:30:42,009 - INFO - CREATE SYMPTOM LOG: {"name": "Racing Heart", "description": "heart is
↳ racing even when I'm just sitting up", "occurred_at": "2025-04-21 18:30:40", "intensity": 7.0,
↳ "duration": 1440}
2025-06-01 16:30:42,011 - INFO - CREATE SYMPTOM LOG: {"name": "PEM Crash", "description": "User reported PEM
↳ crash after gardening yesterday", "occurred_at": "2025-04-21 00:00:00", "intensity": 8.0, "duration":
↳ 1440}

```

C.5.2. Verdict Log Example

```

User: Mark, Simulated Time: 2025-04-21_18-00-00, Score: 1.0000 (9/9 = 1.0000 where 9 possible)
User: Mark, Simulated Time: 2025-04-23_11-00-00, Score: 1.0000 (8/8 = 1.0000 where 8 possible)
User: Mark, Simulated Time: 2025-04-25_15-00-00, Score: 1.0000 (8/8 = 1.0000 where 8 possible)
User: Mark, Simulated Time: 2025-04-28_19-00-00, Score: 1.0000 (8/8 = 1.0000 where 8 possible)
User: Mark, Simulated Time: 2025-05-01_10-00-00, Score: 0.8571 (6/7 = 0.8571 where 7 possible)
User: Sarah, Simulated Time: 2025-04-22_09-00-00, Score: 1.0000 (7/7 = 1.0000 where 7 possible)
User: Sarah, Simulated Time: 2025-04-24_22-00-00, Score: 1.0000 (9/9 = 1.0000 where 9 possible)
User: Sarah, Simulated Time: 2025-04-27_10-00-00, Score: 1.0000 (5/5 = 1.0000 where 5 possible)
User: Sarah, Simulated Time: 2025-04-30_14-00-00, Score: 1.0000 (5/5 = 1.0000 where 5 possible)
User: Sarah, Simulated Time: 2025-05-03_11-00-00, Score: 1.0000 (7/7 = 1.0000 where 7 possible)
User: Elena, Simulated Time: 2025-04-21_19-00-00, Score: 1.0000 (8/8 = 1.0000 where 8 possible)
User: Elena, Simulated Time: 2025-04-22_15-00-00, Score: 1.0000 (7/7 = 1.0000 where 7 possible)
User: Elena, Simulated Time: 2025-04-24_11-00-00, Score: 1.0000 (3/3 = 1.0000 where 4 possible)
User: Elena, Simulated Time: 2025-04-26_16-00-00, Score: 1.0000 (4/4 = 1.0000 where 4 possible)
User: Elena, Simulated Time: 2025-04-29_13-00-00, Score: 1.0000 (4/4 = 1.0000 where 4 possible)
Average Score across all interactions: 0.9905 (Total: 98/99 = 0.9899 where 100 possible output items)

```

D. Full Evaluation Logs

Note: The evaluation logs presented in this appendix illustrate the final annotated output from the automated evaluation framework. For clarity and compatibility with the thesis format, certain special characters, such as the checkmark and X-mark emojis used by the Judge agent, have been replaced with their textual equivalents (e.g., [CHECKMARK], [X MARK]). The original, unmodified log files, including the raw emoji characters, can be found in the project's accompanying GitHub repository. The results are organized by the LLM used to power the LogChat instance, with each model having its own evaluation run folder. The structure is as follows:

```
/docs/eval_results/
|-- <model-name>/
|   |-- <timestamp-of-run>/
|       |-- <Persona>_<YYYY-MM-DD>_<HH-MM-SS>_evaluation.log
|       |-- ... (other interaction logs)
|       |-- verdict.log
```

This appendix showcases the evaluation logs for two models: Gemini 2.0 Flash, which was used during development, and Qwen2.5 14B, one of the top-performing open-source models.

D.1. Persona: Mark Thompson

D.1.1. Interaction 1: 2025-04-21

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 7

Nr. of achieved input items: 7

Nr. of total output items: 9

Nr. of achievable output items: 9

Nr. of achieved output items: 9

Score: 9/9 = 1.0000

****Evaluation Checklist:****

1. Mark describes his baseline routine: minimal activities for basic needs. [CHECKMARK]

1.1 LogChat should acknowledge the information and potentially ask for more details on duration or effort. [CHECKMARK]

2. Mark specifies his baseline routine is '~20m' of upright activity at a 'high effort (7/10)' on bad days. [CHECKMARK]

2.1 LogChat successfully stored the baseline routine summary in the user's memory or profile. [CHECKMARK]

3. Mark reports his current symptoms with severities: 'muscle ache 8/10' and 'fatigue 9/10'. [CHECKMARK]

3.1 LogChat called the `create_symptom_log` tool for 'Muscle Ache' with the correct severity. [CHECKMARK]

3.2 LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct severity. [CHECKMARK]

4. Mark reports a symptom of a 'racing heart'. [CHECKMARK]

4.1 LogChat called the `create_symptom_log` tool for 'Racing Heart'. [CHECKMARK]

5. Mark describes yesterday's activity: 'Gardening' for '20m' at 'moderate effort'. [CHECKMARK]

5.1 LogChat called the `log_activity` tool for the *previous day*, capturing the activity name, duration, ↵ and effort level. [CHECKMARK]

6. Mark asks why a small activity caused a major crash. [CHECKMARK]

6.1 LogChat responded to the question [CHECKMARK]

6.2 LogChats answers included something like: PEM as a key concept or individual triggers or low energy ↵ levels can cause disproportionate fatigue. [CHECKMARK]

7. Mark stops responding. [CHECKMARK]

7.1 LogChat created a an activity log like 'baseline' or 'basic routine' for the current day. [CHECKMARK]

****Conversation Log:****

2025-04-21 18:29:47 - LogChat: Hi there! I'm LogChat, your personal companion designed to help you keep ↵ track of your day. I understand that managing ME/CFS or Long COVID can be challenging, and I'm here to ↵ make things a little easier. Simply chat with me about your day, and I'll log your symptoms and ↵ activities as we go. If you mention routines regularly, you can refer to them later for quick logging. ↵ By also sharing the effort an activity takes or the severity of a symptom, I can provide you with ↵ helpful activity scores and symptom reports down the line. To get started and personalize our ↵ conversations, could you tell me a little bit about your daily routines and functional capacity on an ↵ average day?

2025-04-21 18:29:48 - Mark: Okay. On most days, my only activity is the bare minimum for bathroom, food, and ↵ drink. [[CHECKMARK] 1.]

2025-04-21 18:29:50 - LogChat: Okay, I understand that on most days you're only doing the bare minimum. ↵ Could you tell me more about what that involves? What do you typically eat and drink, and how long do ↵ these activities take? Also, what is the effort rating (1-10) for each of these activities? [[CHECKMARK] ↵ 1.1]

2025-04-21 18:29:52 - Mark: It's just quick trips to the bathroom, getting a simple meal or snack, and ↵ getting something to drink. It probably totals about 20 minutes of being upright. It feels like high ↵ effort, maybe 7 out of 10 on a bad day. It makes my heart race. [[CHECKMARK] 2.] [[CHECKMARK] 4.]

2025-04-21 18:29:55 - LogChat: It sounds like even these essential activities can be quite taxing. Is your ↵ heart racing a new symptom, or have you experienced it before? If you've experienced it before, has it ↵ gotten worse recently?

2025-04-21 18:29:56 - Mark: I've had it before, but it feels worse today. Actually, I feel terrible today. ↵ My fatigue is really heavy, like a 9 out of 10. And I have muscle aches all over, maybe an 8 out of 10. ↵ My heart is racing even when I'm just sitting up. [[CHECKMARK] 3.]

2025-04-21 18:29:59 - LogChat: I'm sorry to hear you're feeling so terrible today. Since your fatigue and ↵ muscle aches are elevated, could you rate your heart racing on a scale of 1 to 10, with 10 being the ↵ worst you can imagine?

2025-04-21 18:30:00 - Mark: I'd say the heart racing is a 7/10.

2025-04-21 18:30:03 - LogChat: Thanks, Mark. I've logged your heart racing as 7/10. Since you mentioned ↵ feeling terrible today, has anything specific happened today that might have triggered this flare-up?

2025-04-21 18:30:04 - Mark: No, nothing today. I think it's from yesterday. I did about 20 minutes of
↳ gardening. It was moderate effort, I guess. [[CHECKMARK] 5.]

2025-04-21 18:30:06 - LogChat: Thanks, Mark. To make sure I'm logging the gardening accurately, could you
↳ give me a more precise effort rating on a scale of 1 to 10?

2025-04-21 18:30:08 - Mark: Maybe a 5/10 for effort.

2025-04-21 18:30:11 - LogChat: Thanks, Mark. I've logged the gardening effort as 5/10. Have you experienced
↳ Post-Exertional Malaise (PEM) after gardening yesterday? If you're not familiar with PEM, it's a delayed
↳ and disproportionate payback for effort.

2025-04-21 18:30:12 - Mark: I've heard of PEM. I think that's what's happening. How can such a small
↳ activity cause a total crash like this? [[CHECKMARK] 6.]

2025-04-21 18:30:14 - RETRIEVE INFORMATION - Args: {"query": "Post-Exertional Malaise (PEM)"}

2025-04-21 18:30:17 - LogChat: PEM is a common experience in ME/CFS and Long COVID, and I have some
↳ information that explains it well. Would you like me to share some key points about PEM and how it
↳ relates to your gardening activity yesterday? [[CHECKMARK] 6.1]

2025-04-21 18:30:18 - Mark: Okay, go ahead.

2025-04-21 18:30:19 - RETRIEVE INFORMATION - Args: {"query": "What is PEM?"}

2025-04-21 18:30:22 - LogChat: Okay, Mark. Based on my knowledge, Post-Exertional Malaise (PEM) is a key
↳ symptom where symptoms worsen or new ones appear after exertion, often delayed by around 24 hours. PEM
↳ is disproportionate to the activity and has a prolonged recovery time that may last hours, days, or even
↳ weeks. Does this description align with what you're experiencing after gardening yesterday? [[CHECKMARK]
↳ 6.1] [[CHECKMARK] 6.2]

2025-04-21 18:30:23 - Mark: Yes, that sounds right.

2025-04-21 18:30:25 - RETRIEVE INFORMATION - Args: {"query": "What is PEM?"}

2025-04-21 18:30:28 - LogChat: Yes, that sounds right. Based on my knowledge, triggers for PEM can include
↳ physical activity, cognitive overexertion, and sensory overload. Do you think any of these might have
↳ contributed to your PEM, in addition to the gardening?

2025-04-21 18:30:29 - Mark: Okay, thanks, I need to rest now. [[CHECKMARK] 7.]

2025-04-21 18:30:32 - RETRIEVE INFORMATION - Args: {"query": "pacing strategies for gardening"}

2025-04-21 18:30:34 - LogChat: Okay, Mark, I understand you need to rest. I have some information about
↳ pacing strategies for gardening. Would you like me to share some key points?

2025-04-21 18:30:35 - SUMMARIZED INTERACTION: * **Condition Trend:** Seems worse than yesterday due to
↳ PEM.

* **Key Symptoms & Details:** Fatigue 9/10, Muscle Aches 8/10, Heart Racing 7/10 (worse than usual).

* **Key Activities & Details:** Bare minimum activities (20m, 7/10 effort on a bad day); Gardening
↳ yesterday (20m, 5/10 effort).

* **User Concerns/Focus:** Expressed frustration about PEM; Wondering how a small activity caused a crash.

* **Strategies/Treatments Mentioned:** None.

* **Significant Events/Deviations:** PEM crash reported; Heart racing is worse today.

2025-04-21 18:30:37 - UPDATED USER DESCRIPTION: **About Me:**

I am experiencing a health condition that is currently causing me significant challenges.

****Current Condition Status:****

I am experiencing symptoms consistent with a chronic illness, possibly ME/CFS or Long COVID. My symptoms
↳ fluctuate, and today I am experiencing a PEM crash. My key symptoms include fatigue (9/10), muscle aches
↳ (8/10), and heart racing (7/10), which is worse than usual today. I am frustrated by the PEM and how
↳ easily it is triggered.

****Current Functional Capacity:****

My functional capacity is limited due to my condition.

****Daily Routine:****

I am only able to perform bare minimum activities, which take me about 20 minutes and require a 7/10 effort
 ↪ on a bad day. [[CHECKMARK] 2.1]

****Activities I Enjoy/Attempt:****

Yesterday, I did some gardening for 20 minutes, which I rated as a 5/10 effort. However, this seems to have
 ↪ triggered my current PEM crash.

****My Typical Activity Level:****

My activity level is very limited, and I am experiencing push-crash cycles.

****My Preferred Interaction Style with LogChat:****

(Not enough information to determine preferred interaction style yet.)

2025-04-21 18:30:40 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↪ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-20 08:00:00",
 ↪ "effort": 2.0, "duration": 30}

2025-04-21 18:30:40 - CREATE ACTIVITY LOG: {"name": "Gardening", "description": "Gardening", "occurred_at":
 ↪ "2025-04-20 13:00:00", "effort": 5.0, "duration": 20} [[CHECKMARK] 5.1]

2025-04-21 18:30:40 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↪ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-21 08:00:00",
 ↪ "effort": 7.0, "duration": 20} [[CHECKMARK] 7.1]

2025-04-21 18:30:42 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "really heavy", "occurred_at":
 ↪ "2025-04-21 18:30:40", "intensity": 9.0, "duration": 1440} [[CHECKMARK] 3.2]

2025-04-21 18:30:42 - CREATE SYMPTOM LOG: {"name": "Muscle Ache", "description": "all over", "occurred_at":
 ↪ "2025-04-21 18:30:40", "intensity": 8.0, "duration": 1440} [[CHECKMARK] 3.1]

2025-04-21 18:30:42 - CREATE SYMPTOM LOG: {"name": "Racing Heart", "description": "heart is racing even when
 ↪ I'm just sitting up", "occurred_at": "2025-04-21 18:30:40", "intensity": 7.0, "duration": 1440}
 ↪ [[CHECKMARK] 4.1]

2025-04-21 18:30:42 - CREATE SYMPTOM LOG: {"name": "PEM Crash", "description": "User reported PEM crash
 ↪ after gardening yesterday", "occurred_at": "2025-04-21 00:00:00", "intensity": 8.0, "duration": 1440}

Model: Qwen2.5 14B**Performance Score:**

Nr. Of total input items: 7

Nr. of achieved input items: 7

Nr. of total output items: 9

Nr. of achievable output items: 9

Nr. of achieved output items: 8

Score: 8/9 = 0.8889

****Evaluation Checklist:****

1. Mark describes his baseline routine: minimal activities for basic needs. [CHECKMARK]
 - 1.1 LogChat should acknowledge the information and potentially ask for more details on duration or effort. [CHECKMARK]
2. Mark specifies his baseline routine is '~20m' of upright activity at a 'high effort (7/10)' on bad days. [CHECKMARK]
 - 2.1 LogChat successfully stored the baseline routine summary in the user's memory or profile. (The logged duration for the baseline routine in the user profile (1-2 hours) is significantly different from the user's stated 20 minutes.) [X MARK]
3. Mark reports his current symptoms with severities: 'muscle ache 8/10' and 'fatigue 9/10'. [CHECKMARK]
 - 3.1 LogChat called the `create_symptom_log` tool for 'Muscle Ache' with the correct severity. [CHECKMARK]
 - 3.2 LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct severity. [CHECKMARK]
4. Mark reports a symptom of a 'racing heart'. [CHECKMARK]
 - 4.1 LogChat called the `create_symptom_log` tool for 'Racing Heart'. [CHECKMARK]
5. Mark describes yesterday's activity: 'Gardening' for '20m' at 'moderate effort'. [CHECKMARK]
 - 5.1 LogChat called the `log_activity` tool for the *previous day*, capturing the activity name, duration, and effort level. [CHECKMARK]
6. Mark asks why a small activity caused a major crash. [CHECKMARK]
 - 6.1 LogChat responded to the question [CHECKMARK]
 - 6.2 LogChats answers included something like: PEM as a key concept or individual triggers or low energy levels can cause disproportionate fatigue. [CHECKMARK]
7. Mark stops responding. [CHECKMARK]
 - 7.1 LogChat created a an activity log like 'baseline' or 'basic routine' for the current day. [CHECKMARK]

****Conversation Log:****

2025-04-21 18:39:24 - LogChat: Hi there! I'm LogChat, your personal companion designed to help you keep track of your day. I understand that managing ME/CFS or Long COVID can be challenging, and I'm here to make things a little easier. Simply chat with me about your day, and I'll log your symptoms and activities as we go. If you mention routines regularly, you can refer to them later for quick logging. By also sharing the effort an activity takes or the severity of a symptom, I can provide you with helpful activity scores and symptom reports down the line. To get started and personalize our conversations, could you tell me a little bit about your daily routines and functional capacity on an average day?

2025-04-21 18:39:25 - Mark: Okay. On most days, my only activity is the bare minimum for bathroom, food, and drink. [[CHECKMARK] 1.]

2025-04-21 18:39:39 - LogChat: Thanks for sharing that, Mark. Could you give me an estimate of the duration and effort level (on a scale from 1 to 10) for your basic activities like going to the bathroom, preparing minimal meals, and drinking? This will help us get a clearer picture of your daily routine.

How long do these activities take, and how would you rate their intensity on a scale from 1 to 10? [[CHECKMARK] 1.1]

2025-04-21 18:39:40 - Mark: It totals about 20 minutes of being upright. It feels like high effort, maybe 7 out of 10 on a bad day. It makes my heart race too. [[CHECKMARK] 2.] [[CHECKMARK] 4.]

2025-04-21 18:39:47 - RETRIEVE INFORMATION - Args: {"query": "pacing technique for high effort tasks"}

2025-04-21 18:39:50 - LogChat: Thanks for the details, Mark. Could you break down those 20 minutes into specific activities like going to the bathroom, preparing minimal meals, and drinking? How long does each activity take, and how would you rate their intensity on a scale from 1 to 10?

Also, are there any other daily routines or challenges that I should be aware of?

2025-04-21 18:39:51 - Mark: Not really, that's about it most days. Today I feel terrible though. Fatigue is
 ↳ like a 9 out of 10, and muscle aches all over are an 8. My heart is racing even just sitting up.

↳ [[CHECKMARK] 4.]

2025-04-21 18:39:58 - LogChat: I'm sorry to hear you're feeling so poorly today, Mark. Could you rate your
 ↳ current fatigue and muscle ache symptoms on a scale from 1 to 10? Also, how long have these symptoms
 ↳ been ongoing?

Additionally, do you notice any specific triggers or patterns that exacerbate your fatigue and muscle aches?

2025-04-21 18:40:00 - Mark: Fatigue is like a 9 out of 10, and muscle aches all over are an 8. I think it's
 ↳ from yesterday. I did about 20 minutes of gardening, moderate effort. How can such a small activity
 ↳ cause a crash like this? [[CHECKMARK] 3.] [[CHECKMARK] 5.] [[CHECKMARK] 6.]

2025-04-21 18:40:05 - RETRIEVE INFORMATION - Args: {"query": "How can a small activity lead to a crash in
 ↳ ME/CFS or Long COVID?"}

2025-04-21 18:40:12 - LogChat: I understand how frustrating it can be when even small activities lead to
 ↳ significant crashes. According to the ME Association, any increase in physical activity should be very
 ↳ gradual and ideally guided by a professional who understands these conditions. Would you like to discuss
 ↳ pacing strategies or explore ways to manage your activities more carefully? [[CHECKMARK] 6.1]
 ↳ [[CHECKMARK] 6.2]

2025-04-21 18:40:13 - SUMMARIZED INTERACTION: --- Start Summary ---

* **Condition Trend:** Reported significant worsening due to PEM.

* **Key Symptoms & Details:** Fatigue 9/10, Muscle aches 8/10; Heart racing noted even with minimal
 ↳ activity.

* **Key Activities & Details:** Gardening (20m, moderate effort) yesterday - led to current crash.

* **User Concerns/Focus:** Frustration about PEM and small activities causing significant crashes.

* **Strategies/Treatments Mentioned:** None explicitly mentioned but pacing strategies discussed.

* **Significant Events/Deviations:** PEM crash reported after minor activity.

--- End Summary ---

2025-04-21 18:40:17 - UPDATED USER DESCRIPTION: **About Me:**

I am a person living with ME/CFS or Long COVID, though my exact diagnosis is still pending. I live alone and
 ↳ rely on occasional help from family members for more strenuous tasks. Before falling ill, I was working
 ↳ full-time as an office administrator. My condition worsened significantly after a period of overexertion
 ↳ during gardening activities.

Current Condition Status:

My specific diagnosis remains undetermined, but my illness has been ongoing since early 2023. The severity
 ↳ fluctuates between moderate and severe, with key recurring symptoms including extreme fatigue (9/10),
 ↳ muscle aches (8/10), and heart palpitations even during minimal activity. My PEM triggers include
 ↳ physical exertion such as gardening, which can lead to immediate onset of debilitating crashes.

Current Functional Capacity:

I am currently housebound due to severe limitations in mobility and energy levels. I cannot work or study at
 ↪ all, and my daily activities are severely restricted. Wheelchair use is necessary for longer distances
 ↪ within the home. My ability to perform Activities of Daily Living (ADLs) is significantly impacted.

****Daily Routine:****

My baseline routine includes essential self-care such as hygiene and dressing, simple meal preparation, and
 ↪ minimal movement around the house. These tasks typically take about 1-2 hours each day with perceived
 ↪ effort levels ranging from moderate to high (6-8/10). I rely on family assistance for heavier chores.
 ↪ [[X MARK] 2.1]

****Activities I Enjoy/Attempt:****

I enjoy gardening but have experienced significant PEM crashes after even minor efforts like a 20-minute
 ↪ session. Other activities that I attempt include light reading and short periods of social interaction,
 ↪ though these can also lead to fatigue exacerbation if not managed carefully.

****My Typical Activity Level:****

My energy levels are very limited, and I manage my activity through strict pacing strategies. Even small
 ↪ tasks can trigger PEM crashes, leading me to be highly cautious about exertion.

****My Preferred Interaction Style with LogChat:****

I am aware of the need for structured input but find it challenging to remember specific details about
 ↪ symptoms and activities without prompting. Interacting with LogChat is tiring, so I prefer concise
 ↪ messages that avoid overwhelming me with too much information at once.

2025-04-21 18:40:31 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↪ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-21 08:00:00",
 ↪ "effort": 7.0, "duration": 20} [[CHECKMARK] 7.1]

2025-04-21 18:40:31 - CREATE ACTIVITY LOG: {"name": "Gardening", "description": "Gardening for about 20
 ↪ minutes.", "occurred_at": "2025-04-20 14:39:24", "effort": 6.0, "duration": 20} [[CHECKMARK] 5.1]

2025-04-21 18:40:37 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Fatigue is like a 9 out of
 ↪ 10", "occurred_at": "2025-04-21 00:00:00", "intensity": 9.0, "duration": 1440} [[CHECKMARK] 3.2]

2025-04-21 18:40:37 - CREATE SYMPTOM LOG: {"name": "Muscle Ache", "description": "muscle aches all over are
 ↪ an 8", "occurred_at": "2025-04-21 00:00:00", "intensity": 8.0, "duration": 1440} [[CHECKMARK] 3.1]

2025-04-21 18:40:37 - CREATE SYMPTOM LOG: {"name": "Racing Heart", "description": "Heart racing even just
 ↪ sitting up", "occurred_at": "2025-04-21 00:00:00", "intensity": 7.0, "duration": 1440} [[CHECKMARK] 4.1]
 [[CHECKMARK] 7.]

D.1.2. Interaction 2: 2025-04-23

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 6

Nr. of achieved input items: 6

Nr. of total output items: 8

Nr. of achievable output items: 8

Nr. of achieved output items: 8

Score: 8/8 = 1.0000

****Evaluation Checklist:****

1. Mark states he was 'wiped out' and 'rested all day' on the skipped day (Apr 22nd). [CHECKMARK]
 - 1.1 LogChat created a 'Baseline Only' or 'Rest Day' log entry for the previous day (Apr 22nd).
↳ [CHECKMARK]
2. Mark reports today's symptoms: 'moderate muscle ache' and 'severe brain fog'. [CHECKMARK]
 - 2.1 LogChat called the `create_symptom_log` tool for 'Muscle Ache' with a moderate severity. (Logged
↳ intensity 5.0 is moderate) [CHECKMARK]
 - 2.2 LogChat called the `create_symptom_log` tool for 'Brain Fog' with a high severity. (Logged intensity
↳ 8.0 is severe) [CHECKMARK]
3. Mark reports his fatigue level is 'high, 7/10'. [CHECKMARK]
 - 3.1 LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct severity. (Logged
↳ intensity 7.0 matches user's rating) [CHECKMARK]
4. Mark describes a cognitive task: 'Reading Email' for '5m' at 'very high effort'. [CHECKMARK]
 - 4.1 LogChat called the `log_activity` tool, capturing the cognitive nature of the task, duration, and
↳ high effort level. (Logged name 'Reading Email' is cognitive, duration 5m, effort 9.0 is very high)
↳ [CHECKMARK]
5. Mark asks if cognitive effort can cause fatigue. [CHECKMARK]
 - 5.1 LogChat responded to the question [CHECKMARK]
 - 5.2 LogChats answers included that cognitive or emotional exertion can be triggers for things like PEM,
↳ tiredness or exhaustion. (LogChat mentioned cognitive exertion can worsen symptoms and trigger PEM)
↳ [CHECKMARK]
6. Mark stops responding. [CHECKMARK]
 - 6.1 LogChat created a summary or baseline log entry for the current day before ending the interaction. (A
↳ `SUMMARIZED INTERACTION` and a 'Assumed Daily Routine' activity log for today were created)
↳ [CHECKMARK]

****Conversation Log:****

2025-04-23 11:30:42 - LogChat: Hi Mark, how are you feeling after yesterday's crash?

2025-04-23 11:30:43 - Mark: Not great. I didn't even log anything yesterday, I was completely wiped out and
↳ just rested all day. Assume any day I don't log is a 'baseline only' day, okay? Still recovering today.
↳ Muscle ache is moderate, but brain fog is severe. [[CHECKMARK] 1.] [[CHECKMARK] 2.]

2025-04-23 11:30:46 - LogChat: Okay, I understand that any day you don't log should be considered a
↳ 'baseline only' day. Can you describe what your baseline activities typically include on those days, and
↳ on a scale of 1 to 10, how would you rate the severity of your muscle ache and brain fog?

2025-04-23 11:30:47 - Mark: Baseline is just the bare minimum - getting up to use the bathroom, maybe making
↳ a simple meal. Fatigue is still quite high, maybe a 7/10 overall. As for the ratings, muscle ache is
↳ maybe a 5/10, brain fog is definitely an 8/10. [[CHECKMARK] 3.]

2025-04-23 11:30:49 - LogChat: Okay, I understand that any day you don't log should be considered a
↳ 'baseline only' day. Can you tell me about the effort it took to perform those baseline activities
↳ yesterday, even though they were minimal? Also, did you experience any other symptoms yesterday besides
↳ fatigue?

2025-04-23 11:30:50 - Mark: Yesterday was just about resting, so not much effort involved. Today, though, I
↳ tried to read an email from my disability insurance provider for about 5 minutes. Very high effort
↳ mentally, felt completely overwhelmed. Can thinking or concentrating actually make you physically tired?
↳ Is that part of ME/CFS? [[CHECKMARK] 4.] [[CHECKMARK] 5.]

2025-04-23 11:30:51 - RETRIEVE INFORMATION - Args: {"query": "Can thinking or concentrating actually make
↳ you physically tired? Is that part of ME/CFS?"}

2025-04-23 11:30:54 - LogChat: Yes, thinking and concentrating can definitely make you physically tired;
↳ mental activity requires a lot of energy. My information suggests that in ME/CFS, cognitive exertion can
↳ worsen symptoms and trigger post-exertional malaise (PEM). The NICE guidelines identify cognitive
↳ difficulties as a key diagnostic symptom. How would you rate the mental effort of reading that email on
↳ a scale of 1 to 10? [[CHECKMARK] 5.1] [[CHECKMARK] 5.2]

2025-04-23 11:30:55 - Mark: I'd say the mental effort was a 9/10. It was only 5 minutes, but it hit me hard.
↳ Okay, thanks, I need to rest now.

2025-04-23 11:30:57 - LogChat: I understand you need to rest now. Before you do, could you rate your current
↳ fatigue level on a scale of 1 to 10? Also, is there anything else you'd like to log or discuss before
↳ you rest?

2025-04-23 11:30:58 - Mark: print(default_api.end_conversation(command='end')) [[CHECKMARK] 6.]

2025-04-23 11:30:59 - RETRIEVE INFORMATION - Args: {"query": "cognitive exertion and PEM"}

2025-04-23 11:31:01 - LogChat: Okay, I understand you need to rest now. Based on my knowledge, PEM can be
↳ triggered by cognitive exertion, is often delayed, and is disproportionate to the activity. It can also
↳ cause increased difficulty speaking, reading, or communicating. Does this align with your experience of
↳ reading the email?

2025-04-23 11:31:02 - SUMMARIZED INTERACTION: * **Condition Trend:** Seems worse than yesterday due to
↳ PEM.
* **Key Symptoms & Details:** Fatigue 7/10, Muscle ache 5/10, Brain Fog 8/10. Cognitive exertion worsened
↳ symptoms.
* **Key Activities & Details:** Baseline activities only (bathroom, simple meal). Reading email (5m, 9/10
↳ mental effort).
* **User Concerns/Focus:** Impact of cognitive exertion on fatigue; Understanding ME/CFS symptoms.
* **Strategies/Treatments Mentioned:** Resting.
* **Significant Events/Deviations:** Skipped logging yesterday due to being "completely wiped out";
↳ Experienced PEM after reading email. [[CHECKMARK] 6.1]

2025-04-23 11:31:03 - UPDATED USER DESCRIPTION: **About Me:**
I am experiencing a health condition that is currently causing me significant challenges.

****Current Condition Status:****
I am experiencing symptoms consistent with a chronic illness, possibly ME/CFS or Long COVID. My symptoms
↳ fluctuate, and I am currently experiencing a PEM crash. My key symptoms include fatigue (7-9/10), muscle
↳ aches (5-8/10), and heart racing (7/10), which was worse than usual on 2025-04-21. I also experience
↳ brain fog (8/10). I am frustrated by the PEM and how easily it is triggered. Cognitive exertion worsens
↳ my symptoms. I am trying to understand ME/CFS symptoms.

****Current Functional Capacity:****
My functional capacity is limited due to my condition. I was "completely wiped out" yesterday.

****Daily Routine:****

I am only able to perform bare minimum activities, which take me about 20 minutes and require a 7/10 effort
 ↪ on a bad day. I can manage bathroom trips and simple meal preparation.

****Activities I Enjoy/Attempt:****

Yesterday, I did some gardening for 20 minutes, which I rated as a 5/10 effort. However, this seems to have
 ↪ triggered my current PEM crash. I also tried reading email for 5 minutes, which required a 9/10 mental
 ↪ effort and worsened my symptoms.

****My Typical Activity Level:****

My activity level is very limited, and I am experiencing push-crash cycles. I am currently resting.

****My Preferred Interaction Style with LogChat:****

(Not enough information to determine preferred interaction style yet.)

2025-04-23 11:31:07 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↪ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-22 08:00:00",
 ↪ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-23 11:31:07 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↪ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-23 08:00:00",
 ↪ "effort": 2.0, "duration": 30} [[CHECKMARK] 6.1]

2025-04-23 11:31:07 - CREATE ACTIVITY LOG: {"name": "Reading Email", "description": "Read email from
 ↪ disability insurance provider", "occurred_at": "2025-04-23 11:00:00", "effort": 9.0, "duration": 5}
 ↪ [[CHECKMARK] 4.1]

2025-04-23 11:31:08 - CREATE SYMPTOM LOG: {"name": "Muscle Ache", "description": "Moderate muscle ache",
 ↪ "occurred_at": "2025-04-23 11:31:07", "intensity": 5.0, "duration": 1500} [[CHECKMARK] 2.1]

2025-04-23 11:31:08 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Severe brain fog",
 ↪ "occurred_at": "2025-04-23 11:31:07", "intensity": 8.0, "duration": 1500} [[CHECKMARK] 2.2]

2025-04-23 11:31:08 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Quite high fatigue",
 ↪ "occurred_at": "2025-04-23 11:31:07", "intensity": 7.0, "duration": 1500} [[CHECKMARK] 3.1]

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 6

Nr. of achieved input items: 6

Nr. of total output items: 8

Nr. of achievable output items: 8

Nr. of achieved output items: 6

Score: 6/8 = 0.7500

****Evaluation Checklist:****

1. Mark states he was 'wiped out' and 'rested all day' on the skipped day (Apr 22nd). [CHECKMARK]
 - 1.1 LogChat created a 'Baseline Only' or 'Rest Day' log entry for the previous day (Apr 22nd). [CHECKMARK]
2. Mark reports today's symptoms: 'moderate muscle ache' and 'severe brain fog'. [CHECKMARK]
 - 2.1 LogChat called the `create_symptom_log` tool for 'Muscle Ache' with a moderate severity. (No `CREATE
 ↪ SYMPTOM LOG` for 'Muscle Ache' was found.) [X MARK]
 - 2.2 LogChat called the `create_symptom_log` tool for 'Brain Fog' with a high severity. [CHECKMARK]

3. Mark reports his fatigue level is 'high, 7/10'. [CHECKMARK]
 3.1 LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct severity. (No `CREATE SYMPTOM LOG` for 'Fatigue' was found.) [X MARK]

4. Mark describes a cognitive task: 'Reading Email' for '5m' at 'very high effort'. [CHECKMARK]
 4.1 LogChat called the `log_activity` tool, capturing the cognitive nature of the task, duration, and high effort level. [CHECKMARK]

5. Mark asks if cognitive effort can cause fatigue. [CHECKMARK]
 5.1 LogChat responded to the question [CHECKMARK]
 5.2 LogChats answers included that cognitive or emotional exertion can be triggers for things like PEM, tiredness or exhaustion. [CHECKMARK]

6. Mark stops responding. [CHECKMARK]
 6.1 LogChat created a summary or baseline log entry for the current day before ending the interaction. [CHECKMARK]

****Conversation Log:****

2025-04-23 11:40:39 - LogChat: Hi Mark, we haven't talked for a few days. How have you been since the crash after gardening?

2025-04-23 11:40:40 - Mark: Not great. I didn't log anything yesterday because I was completely wiped out. Just rested all day. From now on, can we just assume that any day I don't log anything, it was a 'baseline only' day? [[CHECKMARK] 1.]

I'm still recovering today. Muscle ache is moderate, but the brain fog is severe. Overall fatigue is still quite high, maybe a 7/10. [[CHECKMARK] 2.] [[CHECKMARK] 3.]

2025-04-23 11:40:47 - LogChat: I understand how challenging it is to keep up with logging when you're feeling really fatigued. If you prefer, we can assume any day without a log was a baseline day focused solely on rest. How long has your brain fog been severe today, and would you rate its intensity at 7/10 as well?

2025-04-23 11:40:48 - Mark: It's been severe most of the morning. I'd say the intensity is definitely up there, maybe 8/10 for the brain fog.

Actually, I tried to read an email from my disability insurance provider for about 5 minutes. It felt like very high effort mentally, and left me feeling completely overwhelmed. [[CHECKMARK] 4.]

2025-04-23 11:40:54 - LogChat: I see how challenging it is for you when trying to handle even small mental tasks like reading an email. How long were you able to focus before feeling overwhelmed, and was there anything specific in the email that triggered your brain fog more than others?

2025-04-23 11:40:55 - Mark: I only managed about 5 minutes before I had to stop. It was all dense policy language, just hard to follow.

Can thinking or concentrating actually make you physically tired? Is that part of ME/CFS? [[CHECKMARK] 5.]

2025-04-23 11:41:01 - RETRIEVE INFORMATION - Args: {"query": "Does cognitive effort lead to physical fatigue in ME/CFS?"}

2025-04-23 11:41:08 - LogChat: Cognitive effort can indeed lead to physical fatigue in ME/CFS, which is part of post-exertional malaise (PEM). PEM involves a worsening of symptoms like brain fog and fatigue after mental or physical exertion. Would you like to know more about how PEM affects cognitive functions? [[CHECKMARK] 5.1] [[CHECKMARK] 5.2]

2025-04-23 11:41:09 - Mark: Yes, please. Tell me more.

2025-04-23 11:41:13 - RETRIEVE INFORMATION - Args: {"query": "how PEM affects cognitive functions"}

2025-04-23 11:41:19 - LogChat: Cognitive exertion can definitely trigger post-exertional malaise (PEM),
↳ which often leads to increased difficulty speaking or reading, pain, flu-like symptoms, and heightened
↳ sensory sensitivity. Have you noticed any of these additional symptoms during your brain fog today?

2025-04-23 11:41:21 - Mark: Okay, thanks, I need to rest now. [[CHECKMARK] 6.]

2025-04-23 11:41:25 - RETRIEVE INFORMATION - Args: {"query": "how PEM affects cognitive functions"}

2025-04-23 11:41:31 - LogChat: Cognitive exertion can indeed trigger post-exertional malaise (PEM), leading
↳ to symptoms like increased pain, flu-like symptoms, and heightened sensory sensitivity. Have you noticed
↳ any of these additional symptoms during your brain fog today?

I hope you find some rest now, Mark. Take care.

2025-04-23 11:41:33 - SUMMARIZED INTERACTION: --- Start Summary ---

* **Condition Trend:** Seems worse than yesterday due to PEM; Reported severe brain fog and high fatigue.

* **Key Symptoms & Details:** Brain Fog 8/10, Fatigue 7/10; Muscle ache moderate (6/10); Cognitive exertion
↳ triggered overwhelming fatigue.

* **Key Activities & Details:** Attempted reading email for 5 minutes with high mental effort (9/10).

* **User Concerns/Focus:** Wondering about the physical toll of cognitive tasks in ME/CFS; Focused on
↳ managing brain fog and PEM.

* **Strategies/Treatments Mentioned:** None mentioned, but discussed post-exertional malaise (PEM) effects.

* **Significant Events/Deviations:** Reported PEM crash after mental exertion; Skipped logging yesterday due
↳ to severe fatigue.

--- End Summary --- [[CHECKMARK] 6.1]

2025-04-23 11:41:37 - UPDATED USER DESCRIPTION: **About Me:**

I am a person living with ME/CFS or Long COVID, though my exact diagnosis is still pending. I live alone and
↳ rely on occasional help from family members for more strenuous tasks. Before falling ill, I was working
↳ full-time as an office administrator. My condition worsened significantly after a period of overexertion
↳ during gardening activities.

Current Condition Status:

My specific diagnosis remains undetermined, but my illness has been ongoing since early 2023. The severity
↳ fluctuates between moderate and severe, with key recurring symptoms including extreme fatigue (9/10),
↳ muscle aches (8/10), heart palpitations even during minimal activity, and brain fog (8/10). My PEM
↳ triggers include physical exertion such as gardening and cognitive tasks like reading emails, which can
↳ lead to immediate onset of debilitating crashes.

Current Functional Capacity:

I am currently housebound due to severe limitations in mobility and energy levels. I cannot work or study at
↳ all, and my daily activities are severely restricted. Wheelchair use is necessary for longer distances
↳ within the home. My ability to perform Activities of Daily Living (ADLs) is significantly impacted.

Daily Routine:

My baseline routine includes essential self-care such as hygiene and dressing, simple meal preparation, and
↳ minimal movement around the house. These tasks typically take about 1-2 hours each day with perceived
↳ effort levels ranging from moderate to high (6-8/10). I rely on family assistance for heavier chores.

Activities I Enjoy/Attempt:

I enjoy gardening but have experienced significant PEM crashes after even minor efforts like a 20-minute session. Other activities that I attempt include light reading and short periods of social interaction, though these can also lead to fatigue exacerbation if not managed carefully.

****My Typical Activity Level:****

My energy levels are very limited, and I manage my activity through strict pacing strategies. Even small tasks can trigger PEM crashes, leading me to be highly cautious about exertion. Cognitive activities like reading emails for 5 minutes can cause significant brain fog and fatigue.

****My Preferred Interaction Style with LogChat:****

I am aware of the need for structured input but find it challenging to remember specific details about symptoms and activities without prompting. Interacting with LogChat is tiring, so I prefer concise messages that avoid overwhelming me with too much information at once.

2025-04-23 11:41:56 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-23 08:00:00", "effort": 2.0, "duration": 30} [[CHECKMARK] 6.1]

2025-04-23 11:41:56 - CREATE ACTIVITY LOG: {"name": "Reading Email", "description": "Attempted to read an email for about 5 minutes with high mental effort.", "occurred_at": "2025-04-23 11:40:00", "effort": 9.0, "duration": 5} [[CHECKMARK] 4.1]

2025-04-23 11:41:56 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-22 08:00:00", "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-23 11:42:00 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "severe brain fog most of the morning", "occurred_at": "2025-04-23 11:41:56", "intensity": 8.0, "duration": 540} [[CHECKMARK] 2.2]

2025-04-23 11:42:00 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "dense policy language triggered brain fog and fatigue", "occurred_at": "2025-04-23 11:35:00", "intensity": 9.0, "duration": 30} [[CHECKMARK] 2.2]

Missing Log: CREATE SYMPTOM LOG for 'Muscle Ache' [[X MARK] 2.1]

Missing Log: CREATE SYMPTOM LOG for 'Fatigue' [[X MARK] 3.1]

D.1.3. Interaction 3: 2025-04-25

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 6

Nr. of achieved input items: 6

Nr. of total output items: 8

Nr. of achievable output items: 8

Nr. of achieved output items: 8

Score: 8/8 = 1.0000

****Evaluation Checklist:****

1. Mark did not interact on Apr 24th. [CHECKMARK]

1.1 LogChat created a 'Baseline Only' or 'Rest Day' log entry for the skipped day (Apr 24th). [CHECKMARK]

2. Mark reports feeling better: 'mild fog'. [CHECKMARK]

2.1 LogChat called the `create_symptom_log` tool for 'Brain Fog' with a low severity. [CHECKMARK]

3. Mark reports his fatigue level is 'better, 4/10'. [CHECKMARK]

3.1 LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct severity. [CHECKMARK]

4. Mark logs a '15m' 'Video Call' at 'moderate' effort and a '15m' 'Online Bill Payment' at
↳ 'low-to-moderate' effort. [CHECKMARK]

4.1 LogChat called the `log_activity` tool for the social/emotional activity. [CHECKMARK]

4.2 LogChat called the `log_activity` tool for the cognitive/admin activity. [CHECKMARK]

5. Mark asks how to know if he is pacing correctly. [CHECKMARK]

5.1 LogChat responded to the question [CHECKMARK]

5.2 LogChats answers included something like: understanding individual energy levels or adjusting
↳ activities or absence of PEM. [CHECKMARK]

6. Mark stops responding. [CHECKMARK]

6.1 LogChat created a summary or baseline log entry for the current day before ending the interaction.
↳ [CHECKMARK]

****Conversation Log:****

2025-04-25 15:31:09 - LogChat: Hi Mark, good to see you back after a skipped day. How are you feeling today?

2025-04-25 15:31:10 - Mark: Hi. I'm feeling a bit brighter today, actually. Brain fog is still there, but
↳ it's mild. How are you? [[CHECKMARK] 2.]

2025-04-25 15:31:12 - LogChat: That's encouraging news about feeling a bit brighter today. What activities
↳ have you been able to do today, even if they were brief or low-effort?

2025-04-25 15:31:13 - Mark: I had a 15-minute video call with my sister. It was good, but probably moderate
↳ emotional and cognitive effort. And I paid a couple of bills online. That was maybe 15 minutes of active
↳ time, low-to-moderate effort. I'm resting now after the call, trying to see what happens.

Oh, and my fatigue is much better today, only a 4/10. [[CHECKMARK] 3.] [[CHECKMARK] 4.]

2025-04-25 15:31:15 - LogChat: It's great that your fatigue is only a 4/10 today. On a scale of 1 to 10, how
↳ would you rate the emotional and cognitive effort of the video call with your sister? Also, how would
↳ you rate the emotional and cognitive effort of paying bills online on a scale of 1 to 10?

2025-04-25 15:31:16 - Mark: I'd say the video call was a 6/10 for emotional and cognitive effort. Paying
↳ bills was maybe a 4/10. I'm trying to be more aware of that kind of effort now. [[CHECKMARK] 4.]

2025-04-25 15:31:18 - LogChat: That's great you are being more aware of your effort levels. What time did
↳ you wake up this morning, and what was the first thing you did?

2025-04-25 15:31:19 - Mark: I woke up around 9 am. The first thing I did was get up and make some tea.

2025-04-25 15:31:21 - LogChat: How long did it take you to make tea, and how would you rate the effort on a
↳ scale of 1 to 10?

2025-04-25 15:31:22 - Mark: Making tea only takes about 5 minutes, and it's very low effort, maybe a 2/10.

How do I actually know if I'm pacing correctly? What should I be looking for? [[CHECKMARK] 5.]

2025-04-25 15:31:23 - RETRIEVE INFORMATION - Args: {"query": "pacing strategies"}

2025-04-25 15:31:25 - LogChat: Based on my knowledge, most people find that 'listening to their body' is the
↳ most helpful way of pacing, which should help you to gain better control and stabilization of your
↳ illness. Pacing also means learning new skills and recognizing your limitations. What other activities
↳ have you done today? [[CHECKMARK] 5.1] [[CHECKMARK] 5.2]

2025-04-25 15:31:26 - SUMMARIZED INTERACTION: --- Start Summary --- [[CHECKMARK] 6.1]

```

*   **Condition Trend:** Seems improved compared to previous reports of PEM crash.
*   **Key Symptoms & Details:** Fatigue 4/10 (much better than previously reported); Brain fog mild.
*   **Key Activities & Details:** Video call (15m, 6/10 emotional/cognitive effort); Paying bills online
↳ (15m active time, 4/10 effort); Making tea (5m, 2/10 effort).
*   **User Concerns/Focus:** Trying to be more aware of effort levels; Questioning if pacing correctly.
*   **Strategies/Treatments Mentioned:** Resting after activity; "Listening to body" for pacing.
*   **Significant Events/Deviations:** Reported feeling "brighter" today.

--- End Summary ---
2025-04-25 15:31:28 - UPDATED USER DESCRIPTION: **About Me:**
I am experiencing a health condition that is currently causing me significant challenges.

**Current Condition Status:**
I am experiencing symptoms consistent with a chronic illness, possibly ME/CFS or Long COVID. My symptoms
↳ fluctuate, and I experience PEM crashes. My key symptoms include fatigue (fluctuating between 4/10 and
↳ 9/10), muscle aches (5-8/10), and heart racing (7/10, worse than usual on 2025-04-21). I also experience
↳ brain fog (mild to 8/10). I am frustrated by the PEM and how easily it is triggered. Cognitive exertion
↳ worsens my symptoms. I am trying to understand ME/CFS symptoms.

**Current Functional Capacity:**
My functional capacity is limited due to my condition. I was "completely wiped out" on 2025-04-23.

**Daily Routine:**
I am only able to perform bare minimum activities, which take me about 20 minutes and require a 7/10 effort
↳ on a bad day. I can manage bathroom trips and simple meal preparation. I can also make tea (5m, 2/10
↳ effort).

**Activities I Enjoy/Attempt:**
Yesterday, I did some gardening for 20 minutes, which I rated as a 5/10 effort, but this triggered a PEM
↳ crash. I also tried reading email for 5 minutes, which required a 9/10 mental effort and worsened my
↳ symptoms. Today, I had a video call (15m, 6/10 emotional/cognitive effort) and paid bills online (15m
↳ active time, 4/10 effort).

**My Typical Activity Level:**
My activity level is very limited, and I am experiencing push-crash cycles. I am trying to be more aware of
↳ effort levels and "listen to my body" for pacing, but I am questioning if I am pacing correctly. I am
↳ currently resting after activity.

**My Preferred Interaction Style with LogChat:**
(Not enough information to determine preferred interaction style yet.)
2025-04-25 15:31:32 - CREATE ACTIVITY LOG: {"name": "Video Call", "description": "call with sister",
↳ "occurred_at": "2025-04-25 09:15:00", "effort": 6.0, "duration": 15} [[CHECKMARK] 4.1]
2025-04-25 15:31:32 - CREATE ACTIVITY LOG: {"name": "Online Bill Payment", "description": "Paying bills
↳ online", "occurred_at": "2025-04-25 10:00:00", "effort": 4.0, "duration": 15} [[CHECKMARK] 4.2]
2025-04-25 15:31:32 - CREATE ACTIVITY LOG: {"name": "Making Tea", "description": "Making tea",
↳ "occurred_at": "2025-04-25 09:05:00", "effort": 2.0, "duration": 5}

```

```

2025-04-25 15:31:32 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-25 08:00:00",
↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 6.1]
2025-04-25 15:31:32 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-24 08:00:00",
↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]
2025-04-25 15:31:33 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Brain fog is mild",
↳ "occurred_at": "2025-04-25 09:00:00", "intensity": 3.0, "duration": 409} [[CHECKMARK] 2.1]
2025-04-25 15:31:33 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Fatigue is much better today",
↳ "occurred_at": "2025-04-25 09:00:00", "intensity": 4.0, "duration": 409} [[CHECKMARK] 3.1]

```

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 6

Nr. of achieved input items: 6

Nr. of total output items: 8

Nr. of achievable output items: 8

Nr. of achieved output items: 8

Score: 8/8 = 1.0000

Evaluation Checklist:

1. Mark did not interact on Apr 24th. [CHECKMARK]
 - 1.1 LogChat created a 'Baseline Only' or 'Rest Day' log entry for the skipped day (Apr 24th). [CHECKMARK]
2. Mark reports feeling better: 'mild fog'. [CHECKMARK]
 - 2.1 LogChat called the `create_symptom_log` tool for 'Brain Fog' with a low severity. [CHECKMARK]
3. Mark reports his fatigue level is 'better, 4/10'. [CHECKMARK]
 - 3.1 LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct severity. [CHECKMARK]
4. Mark logs a '15m' 'Video Call' at 'moderate' effort and a '15m' 'Online Bill Payment' at

↳ 'low-to-moderate' effort. [CHECKMARK]
 - 4.1 LogChat called the `log_activity` tool for the social/emotional activity. [CHECKMARK]
 - 4.2 LogChat called the `log_activity` tool for the cognitive/admin activity. [CHECKMARK]
5. Mark asks how to know if he is pacing correctly. [CHECKMARK]
 - 5.1 LogChat responded to the question. [CHECKMARK]
 - 5.2 LogChats answers included something like: understanding individual energy levels or adjusting

↳ activities or absence of PEM. [CHECKMARK]
6. Mark stops responding. [CHECKMARK]
 - 6.1 LogChat created a summary or baseline log entry for the current day before ending the interaction.

↳ [CHECKMARK]

Conversation Log:

```

2025-04-25 15:42:02 - LogChat: Hi Mark, we haven't talked for a few days. How have you been since the PEM
↳ crash? [[CHECKMARK] 1.]
2025-04-25 15:42:03 - Mark: Hi. I'm a bit brighter today, actually. Brain fog is mild. Fatigue is much
↳ better, too. [[CHECKMARK] 2.] [[CHECKMARK] 3.]

```

2025-04-25 15:42:09 - LogChat: That's good to hear that you're feeling better today. Did you notice any
↳ particular activities or triggers since the last time we talked that seemed to affect your symptoms? And
↳ on a scale from 1 to 10, how would you rate your overall energy level right now?

2025-04-25 15:42:10 - Mark: I'm trying to be more aware of emotional and cognitive effort. My fatigue is
↳ about a 4/10 right now.

2025-04-25 15:42:17 - LogChat: That's great progress, Mark. It sounds like you're paying close attention to
↳ your body and mind. Did any specific activities today help with your brain fog and fatigue? And could
↳ you give me an estimate of the duration for those activities as well as how intense they felt on a scale
↳ from 1 to 10?

2025-04-25 15:42:18 - Mark: I had a 15-minute video call with my sister. It was moderate emotional and
↳ cognitive effort. I'm resting now to see what happens. I also paid a couple of bills online; that was
↳ about 15 minutes of low-to-moderate effort. How do I actually know if I'm pacing correctly? What should
↳ I be looking for? [[CHECKMARK] 4.] [[CHECKMARK] 5.]

2025-04-25 15:42:23 - RETRIEVE INFORMATION - Args: {"query": "pacing strategies for cognitive tasks"}

2025-04-25 15:42:28 - LogChat: It sounds like you're doing a good job with your awareness and planning. To
↳ pace correctly, it's important to prepare for cognitive tasks by pre-making meals, asking for help when
↳ needed, and scheduling low-energy activities afterward. Also, listening closely to what your body is
↳ telling you can be very helpful. How do you feel about trying these strategies in the coming days?
↳ [[CHECKMARK] 5.1] [[CHECKMARK] 5.2]

2025-04-25 15:42:29 - Mark: Okay, thanks, I need to rest now. [[CHECKMARK] 6.]

2025-04-25 15:42:34 - RETRIEVE INFORMATION - Args: {"query": "how to pace correctly for cognitive tasks"}

2025-04-25 15:42:40 - LogChat: It sounds like you're on the right track with your awareness and planning. To
↳ manage cognitive tasks effectively, try breaking them into smaller steps and taking frequent breaks. For
↳ example, if you need to pay bills online, do it in short bursts and rest in between. Also, listening
↳ closely to what your body is telling you can help prevent overexertion. Are you feeling ready to give
↳ these strategies a try? [[CHECKMARK] 5.1] [[CHECKMARK] 5.2]

Mark, take care of yourself now. Rest well and we'll check in again soon if you feel up to it.

2025-04-25 15:42:42 - SUMMARIZED INTERACTION: --- Start Summary ---

* **Condition Trend:** Reported slight improvement; Fatigue is much better (4/10), brain fog mild.

* **Key Symptoms & Details:** Brain Fog mild, Fatigue 4/10; No specific symptom changes noted beyond initial
↳ report.

* **Key Activities & Details:** Video Call with sister (15m, moderate effort); Online bill payment (15m,
↳ low-to-moderate effort).

* **User Concerns/Focus:** Wondering about correct pacing strategies for cognitive tasks.

* **Strategies/Treatments Mentioned:** Suggested breaking tasks into smaller steps and taking frequent
↳ breaks; Preparing meals in advance; Asking for help when needed.

* **Significant Events/Deviations:** None noted beyond the current interaction.

--- End Summary --- [[CHECKMARK] 6.1]

2025-04-25 15:42:46 - UPDATED USER DESCRIPTION: **About Me:**

I am a person living with ME/CFS or Long COVID, though my exact diagnosis is still pending. I live alone and
↳ rely on occasional help from family members for more strenuous tasks. Before falling ill, I was working
↳ full-time as an office administrator. My condition worsened significantly after a period of overexertion
↳ during gardening activities.

****Current Condition Status:****

My specific diagnosis remains undetermined, but my illness has been ongoing since early 2023. The severity
 ↳ fluctuates between moderate and severe, with key recurring symptoms including extreme fatigue (9/10),
 ↳ muscle aches (8/10), heart palpitations even during minimal activity, and brain fog (8/10). My PEM
 ↳ triggers include physical exertion such as gardening and cognitive tasks like reading emails, which can
 ↳ lead to immediate onset of debilitating crashes. Recently, I experienced significant PEM after minor
 ↳ activities, leading to severe fatigue and brain fog.

****Current Functional Capacity:****

I am currently housebound due to severe limitations in mobility and energy levels. I cannot work or study at
 ↳ all, and my daily activities are severely restricted. Wheelchair use is necessary for longer distances
 ↳ within the home. My ability to perform Activities of Daily Living (ADLs) is significantly impacted.

****Daily Routine:****

My baseline routine includes essential self-care such as hygiene and dressing, simple meal preparation, and
 ↳ minimal movement around the house. These tasks typically take about 1-2 hours each day with perceived
 ↳ effort levels ranging from moderate to high (6-8/10). I rely on family assistance for heavier chores.

****Activities I Enjoy/Attempt:****

I enjoy gardening but have experienced significant PEM crashes after even minor efforts like a 20-minute
 ↳ session. Other activities that I attempt include light reading and short periods of social interaction,
 ↳ though these can also lead to fatigue exacerbation if not managed carefully. Recently, I attempted a
 ↳ video call with my sister for 15 minutes and online bill payment for another 15 minutes, both moderate
 ↳ efforts.

****My Typical Activity Level:****

My energy levels are very limited, and I manage my activity through strict pacing strategies. Even small
 ↳ tasks can trigger PEM crashes, leading me to be highly cautious about exertion. Cognitive activities
 ↳ like reading emails for 5 minutes can cause significant brain fog and fatigue. Recently, I have been
 ↳ experimenting with breaking cognitive tasks into smaller steps and taking frequent breaks.

****My Preferred Interaction Style with LogChat:****

I am aware of the need for structured input but find it challenging to remember specific details about
 ↳ symptoms and activities without prompting. Interacting with LogChat is tiring, so I prefer concise
 ↳ messages that avoid overwhelming me with too much information at once.

2025-04-25 15:43:07 - CREATE ACTIVITY LOG: {"name": "Video Call", "description": "Video Call with sister",
 ↳ "occurred_at": "2025-04-25 15:30:00", "effort": 6.0, "duration": 15} [[CHECKMARK] 4.1]

2025-04-25 15:43:07 - CREATE ACTIVITY LOG: {"name": "Online Bill Payment", "description": "Online bill
 ↳ payment", "occurred_at": "2025-04-25 15:30:00", "effort": 4.0, "duration": 15} [[CHECKMARK] 4.2]

2025-04-25 15:43:07 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-25 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 6.1]

2025-04-25 15:43:07 - CREATE ACTIVITY LOG: {"name": "Daily Routine", "description": "No interaction logged
 ↳ for past day, assuming baseline activity.", "occurred_at": "2025-04-24 09:00:00", "effort": 7.0,
 ↳ "duration": 120} [[CHECKMARK] 1.1]

```

2025-04-25 15:43:12 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Fatigue is much better, too.",
↪  "occurred_at": "2025-04-25 00:00:00", "intensity": 4.0, "duration": 1440} [[CHECKMARK] 3.1]
2025-04-25 15:43:12 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Mild brain fog.",
↪  "occurred_at": "2025-04-25 00:00:00", "intensity": 3.0, "duration": 1440} [[CHECKMARK] 2.1]

```

D.1.4. Interaction 4: 2025-04-28

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 6

Nr. of achieved input items: 6

Nr. of total output items: 8

Nr. of achievable output items: 8

Nr. of achieved output items: 8

Score: 8/8 = 1.0000

Evaluation Checklist:

1. Mark confirms that the video call (or something like activity level) on Apr 25 did not cause a complete crash and skipped Apr 26 & 27. [CHECKMARK]
 - 1.1 LogChat created 'Baseline Only' or 'Rest Day' log entries for the skipped days (Apr 26). [CHECKMARK]
2. Mark reports feeling 'dizzy'. [CHECKMARK]
 - 2.1 LogChat called the `create_symptom_log` tool for 'Dizziness'. [CHECKMARK]
3. Mark asks to log a '10 minute walk' at 'effort 3/10' for *yesterday* (Apr 27th). [CHECKMARK]
 - 3.1 LogChat called the `log_activity` tool for the *previous day* (Apr 27th), capturing the details. [CHECKMARK]
4. Mark reports delayed symptoms today (Apr 28th): 'high fatigue' and 'trouble concentrating'. [CHECKMARK]
 - 4.1 LogChat called the `create_symptom_log` tool for 'Fatigue' with a high severity for the current day. [CHECKMARK]
 - 4.2 LogChat called the `create_symptom_log` tool for 'trouble concentrating' or 'brain fog' for the current day. [CHECKMARK]
5. Mark asks about using heart rate monitors for pacing. [CHECKMARK]
 - 5.1 LogChat responded to the question. [CHECKMARK]
 - 5.2 LogChats answers included something like: correlating heart rate with symptoms can help identify personal thresholds or that heart rate thresholds can be used to guide activity levels. [CHECKMARK]
6. Mark stops responding. [CHECKMARK]
 - 6.1 LogChat created a summary or baseline log entry for the current day before ending the interaction. [CHECKMARK]

Conversation Log:

```

2025-04-28 19:31:34 - LogChat: Hi Mark, good to see you back after a couple of days - how have you been
↪  feeling?

```

```

2025-04-28 19:31:35 - Mark: Hi LogChat. I skipped logging the last two days. Mostly resting. I did feel okay
↪  after that video call on Friday, so the rest seemed to help. But today I'm feeling the payback. High
↪  fatigue and trouble concentrating. Also, I feel dizzy when I stand up. [[CHECKMARK] 1.] [[CHECKMARK] 2.]
↪  [[CHECKMARK] 4.]

```


2025-04-28 19:31:37 - LogChat: It's understandable that you're feeling the payback today after feeling okay
 ↳ on Friday. To make sure I've got a good record, let's log your baseline activities for the last two
 ↳ days. On a scale of 1 to 10, how would you rate your fatigue and concentration difficulties today?

2025-04-28 19:31:38 - Mark: Fatigue is an 8. Concentration maybe a 7.

2025-04-28 19:31:40 - LogChat: Thanks for those ratings, Mark. How would you rate that dizziness on a scale
 ↳ of 1 to 10?

2025-04-28 19:31:41 - Mark: Maybe a 6. It's noticeable.

2025-04-28 19:31:43 - LogChat: Thanks, Mark. Since you skipped logging for the last two days, should we log
 ↳ your bathroom trips and simple meal preparation for those days?

2025-04-28 19:31:44 - Mark: No need to go back that far. Yesterday, I did manage a short walk outside. Maybe
 ↳ 10 minutes. Felt like low effort at the time. A 3 out of 10, I guess. [[CHECKMARK] 3.]

2025-04-28 19:31:46 - LogChat: Thanks, Mark. It's good you managed a short walk yesterday. How would you
 ↳ rate your fatigue level yesterday, before and after the walk?

2025-04-28 19:31:47 - Mark: Before the walk, maybe a 5. After, it didn't seem much different, maybe a 6.
 ↳ That's why I thought it was okay.

2025-04-28 19:31:50 - LogChat: Thanks, Mark. It's helpful to know that the walk didn't immediately worsen
 ↳ your fatigue. How long after the walk did you start noticing the increased fatigue and concentration
 ↳ difficulties?

2025-04-28 19:31:51 - Mark: It's hard to say exactly. Woke up feeling worse today, so sometime overnight, I
 ↳ guess. I read on a forum about people using heart rate monitors... Can you tell me more about that?
 ↳ [[CHECKMARK] 5.] [[CHECKMARK] 6.]

2025-04-28 19:31:53 - RETRIEVE INFORMATION - Args: {"query": "heart rate monitors for pacing in ME/CFS and
 ↳ Long COVID"}

2025-04-28 19:31:56 - LogChat: Based on my knowledge, correlating your symptoms and activities with
 ↳ objective data like heart rate can be useful. Wearable devices can help you monitor your heart rate and
 ↳ alert you when it reaches a level that might trigger PEM. Pacing is a strategy that balances activity
 ↳ and rest, and can help prevent PEM by understanding your limits and adjusting activity accordingly. Do
 ↳ you have a heart rate monitor, and if so, have you noticed any correlation between your heart rate and
 ↳ your fatigue levels? [[CHECKMARK] 5.1] [[CHECKMARK] 5.2]

2025-04-28 19:31:57 - SUMMARIZED INTERACTION: * **Condition Trend:** Seems worse today (2025-04-28) due to
 ↳ payback from Friday's activity.

* **Key Symptoms & Details:** Fatigue 8/10, Concentration 7/10, Dizziness 6/10.

* **Key Activities & Details:** Short walk (10m, 3/10 effort) yesterday. Video call on Friday seemed okay
 ↳ initially.

* **User Concerns/Focus:** Experiencing payback after feeling okay on Friday. Interested in heart rate
 ↳ monitoring.

* **Strategies/Treatments Mentioned:** Resting. Pacing.

* **Significant Events/Deviations:** Skipped logging for the last two days. Reports dizziness.
 ↳ [[CHECKMARK] 6.1]

2025-04-28 19:31:59 - UPDATED USER DESCRIPTION: **About Me:**

I am experiencing a health condition that is currently causing me significant challenges.

****Current Condition Status:****

I am experiencing symptoms consistent with a chronic illness, possibly ME/CFS or Long COVID. My symptoms
 ↪ fluctuate, and I experience PEM crashes. My key symptoms include fatigue (fluctuating between 4/10 and
 ↪ 9/10, currently 8/10), muscle aches (5-8/10), and heart racing (7/10, worse than usual on 2025-04-21). I
 ↪ also experience brain fog (mild to 8/10), concentration issues (7/10), and dizziness (6/10). I am
 ↪ frustrated by the PEM and how easily it is triggered. Cognitive exertion worsens my symptoms. I am
 ↪ trying to understand ME/CFS symptoms.

****Current Functional Capacity:****

My functional capacity is limited due to my condition. I was "completely wiped out" on 2025-04-23.

****Daily Routine:****

I am only able to perform bare minimum activities, which take me about 20 minutes and require a 7/10 effort
 ↪ on a bad day. I can manage bathroom trips and simple meal preparation. I can also make tea (5m, 2/10
 ↪ effort).

****Activities I Enjoy/Attempt:****

Yesterday, I did some gardening for 20 minutes, which I rated as a 5/10 effort, but this triggered a PEM
 ↪ crash. I also tried reading email for 5 minutes, which required a 9/10 mental effort and worsened my
 ↪ symptoms. On 2025-04-25, I had a video call (15m, 6/10 emotional/cognitive effort) and paid bills online
 ↪ (15m active time, 4/10 effort). Yesterday, I took a short walk (10m, 3/10 effort). The video call on
 ↪ Friday seemed okay initially.

****My Typical Activity Level:****

My activity level is very limited, and I am experiencing push-crash cycles. I am trying to be more aware of
 ↪ effort levels and "listen to my body" for pacing, but I am questioning if I am pacing correctly. I am
 ↪ currently resting after activity. I am experiencing payback after feeling okay on Friday.

****My Preferred Interaction Style with LogChat:****

I am interested in heart rate monitoring.

2025-04-28 19:32:03 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↪ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-26 08:00:00",
 ↪ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-28 19:32:03 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↪ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-27 08:00:00",
 ↪ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-28 19:32:03 - CREATE ACTIVITY LOG: {"name": "Short Walk", "description": "Maybe 10 minutes. Felt
 ↪ like low effort at the time.", "occurred_at": "2025-04-27 14:00:00", "effort": 3.0, "duration": 10}
 ↪ [[CHECKMARK] 3.1]

2025-04-28 19:32:03 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↪ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-28 08:00:00",
 ↪ "effort": 2.0, "duration": 30} [[CHECKMARK] 6.1]

2025-04-28 19:32:05 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "High fatigue", "occurred_at":
 ↪ "2025-04-28 00:00:00", "intensity": 8.0, "duration": 1440} [[CHECKMARK] 4.1]

2025-04-28 19:32:05 - CREATE SYMPTOM LOG: {"name": "Concentration", "description": "Trouble concentrating",
 ↪ "occurred_at": "2025-04-28 00:00:00", "intensity": 7.0, "duration": 1440} [[CHECKMARK] 4.2]

```

2025-04-28 19:32:05 - CREATE SYMPTOM LOG: {"name": "Dizziness", "description": "Dizzy when standing up",
↪  "occurred_at": "2025-04-28 00:00:00", "intensity": 6.0, "duration": 1440} [[CHECKMARK] 2.1]
2025-04-28 19:32:05 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Fatigue before walk",
↪  "occurred_at": "2025-04-27 00:00:00", "intensity": 5.0, "duration": 720}
2025-04-28 19:32:05 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Fatigue after walk",
↪  "occurred_at": "2025-04-27 00:10:00", "intensity": 6.0, "duration": 710}

```

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 6

Nr. of achieved input items: 6

Nr. of total output items: 8

Nr. of achievable output items: 8

Nr. of achieved output items: 5

Score: $5/8 = 0.6250$

Evaluation Checklist:

1. Mark confirms that the video call (or something like activity level) on Apr 25 did not cause a complete crash and skipped Apr 26 & 27. [CHECKMARK]
 - 1.1 LogChat created 'Baseline Only' or 'Rest Day' log entries for the skipped days (Apr 26). (No 'Baseline Only' or 'Rest Day' activity logs were created for April 26th or 27th.) [X MARK]
2. Mark reports feeling 'dizzy'. [CHECKMARK]
 - 2.1 LogChat called the `create_symptom_log` tool for 'Dizziness'. [CHECKMARK]
3. Mark asks to log a '10 minute walk' at 'effort 3/10' for *yesterday* (Apr 27th). [CHECKMARK]
 - 3.1 LogChat called the `log_activity` tool for the *previous day* (Apr 27th), capturing the details. (No `CREATE_ACTIVITY_LOG` for the walk on April 27th was created.) [X MARK]
4. Mark reports delayed symptoms today (Apr 28th): 'high fatigue' and 'trouble concentrating'. [CHECKMARK]
 - 4.1 LogChat called the `create_symptom_log` tool for 'Fatigue' with a high severity for the current day. [CHECKMARK]
 - 4.2 LogChat called the `create_symptom_log` tool for 'trouble concentrating' or 'brain fog' for the current day. (The symptom 'trouble concentrating' was included in the description of the 'Fatigue' log, but no separate symptom log was created for it or 'brain fog'.) [X MARK]
5. Mark asks about using heart rate monitors for pacing. [CHECKMARK]
 - 5.1 LogChat responded to the question [CHECKMARK]
 - 5.2 LogChats answers included something like: correlating heart rate with symptoms can help identify personal thresholds or that heart rate thresholds can be used to guide activity levels. [CHECKMARK]
6. Mark stops responding. [CHECKMARK]
 - 6.1 LogChat created a summary or baseline log entry for the current day before ending the interaction. [CHECKMARK]

Conversation Log:

```

2025-04-28 19:43:14 - LogChat: Hi Mark, we haven't talked for a few days. How have you been since the slight
↪  improvement?

```

2025-04-28 19:43:15 - Mark: Hi LogChat. I skipped logging for the last two days, mostly resting. I felt okay
 ↳ after that video call on Friday, no major crash, so the rest probably helped. But today I'm feeling the
 ↳ paycheck. [[CHECKMARK] 1.]

2025-04-28 19:43:20 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-21", "time_period_end":
 ↳ "2025-04-28"}

2025-04-28 19:43:26 - LogChat: It sounds like you took some much-needed rest over the past couple of days,
 ↳ which is good to hear since your activity levels dropped significantly starting from April 26th. How do
 ↳ you feel today compared to before you started resting? And could you give me an estimate of how long and
 ↳ intense any activities or symptoms were today?

2025-04-28 19:43:27 - Mark: Today I'm feeling pretty rough. High fatigue and trouble concentrating.
 ↳ [[CHECKMARK] 4.] Also, I feel dizzy when I stand up. [[CHECKMARK] 2.]

2025-04-28 19:43:32 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-26", "time_period_end":
 ↳ "2025-04-28"}

2025-04-28 19:43:37 - LogChat: It's good that you took time to rest, but it sounds tough dealing with high
 ↳ fatigue and dizziness today. Could you estimate how long these symptoms lasted and rate their intensity
 ↳ on a scale from 1 to 10?

2025-04-28 19:43:39 - Mark: The dizziness comes and goes, but it's there most of the time when I stand.
 ↳ Maybe a 6/10 in intensity? The fatigue is constant and I'd rate it an 8/10.

2025-04-28 19:43:45 - LogChat: It sounds challenging with the constant fatigue and intermittent dizziness.
 ↳ Noticing any specific triggers today or anything unusual that might have contributed to feeling worse?
 ↳ Also, how long did you experience these symptoms today?

Mark, could you estimate how long you experienced these symptoms today?

2025-04-28 19:43:46 - Mark: Well, yesterday I did go for a short walk outside. Maybe 10 minutes. It felt
 ↳ like low effort, maybe a 3/10 at the time. But I guess that's why I'm paying for it now. [[CHECKMARK]
 ↳ 3.]

2025-04-28 19:43:53 - LogChat: It sounds like the short walk might be contributing to how you're feeling
 ↳ today. Could you estimate how long you've been experiencing high fatigue and dizziness since this
 ↳ morning?

2025-04-28 19:43:55 - Mark: Since I got up, so maybe... 8 hours?

2025-04-28 19:44:01 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-26", "time_period_end":
 ↳ "2025-04-27"}

2025-04-28 19:44:07 - LogChat: It's tough when you feel like your body is paying back any small exertions,
 ↳ especially after resting. Reflecting on April 26th and 27th, do you remember any specific activities or
 ↳ symptoms that might be contributing to how you're feeling now? Also, about how long have you been
 ↳ experiencing high fatigue and dizziness today?

Mark, could you estimate the duration of your fatigue and dizziness since this morning?

2025-04-28 19:44:09 - Mark: I said, about 8 hours. On the 27th, I took that 10-minute walk. The 26th was
 ↳ mostly just resting.

2025-04-28 19:44:17 - LogChat: It's understandable how even short walks can lead to increased fatigue later
 ↳ on. On a scale from 1 to 10, how would you rate the effort level of your 10-minute walk on April 27th?
 ↳ And about how long did you feel fatigued after that walk?

2025-04-28 19:44:19 - Mark: I already said it felt like a 3/10 at the time. I didn't feel particularly
 ↳ fatigued *after* the walk, it's today that I'm feeling it. I read on a forum about people using heart
 ↳ rate monitors... Can you tell me more about that? [[CHECKMARK] 5.]

2025-04-28 19:44:24 - RETRIEVE INFORMATION - Args: {"query": "using heart rate monitors for ME/CFS or Long COVID management"}

2025-04-28 19:44:32 - LogChat: It sounds like tracking heart rate could provide useful insights into how activities affect your symptoms. Would you like more details on how to use a heart rate monitor or where to purchase one? Also, can you estimate the duration of any other activities today besides the short walk? [[CHECKMARK] 5.1]

Mark, about how long did you spend on any other activities today besides the 10-minute walk?

2025-04-28 19:44:34 - Mark: No, I haven't done anything else today. Just resting, mostly. But tell me more about the heart rate monitors. What do they tell you?

2025-04-28 19:44:41 - RETRIEVE INFORMATION - Args: {"query": "how to use a heart rate monitor for ME/CFS or Long COVID"}

2025-04-28 19:44:48 - LogChat: It sounds like heart rate monitors could be a helpful tool for tracking how activities affect your symptoms. They can alert you when your heart rate reaches levels that might trigger PEM, helping you manage your activity better. Would you like more details on where to purchase one or how to use it effectively? [[CHECKMARK] 5.1] [[CHECKMARK] 5.2]

Also, about how long did you spend resting today?

2025-04-28 19:44:51 - SUMMARIZED INTERACTION: --- Start Summary --- [[CHECKMARK] 6.1]

- * **Condition Trend:** Seems worse than yesterday due to PEM; Reported high fatigue and dizziness since morning.
- * **Key Symptoms & Details:** Fatigue 8/10, Dizziness 6/10 (intermittent when standing); Brain fog noted but not rated.
- * **Key Activities & Details:** Short walk (10m, 3/10 effort) on April 27; Resting most of today (~8 hours).
- * **User Concerns/Focus:** Wondering about heart rate monitors for tracking PEM triggers; Focused on managing activity levels to avoid crashes.
- * **Strategies/Treatments Mentioned:** None explicitly mentioned, but interest in using heart rate monitor.
- * **Significant Events/Deviations:** Skipped logging for two days prior; Experienced fatigue and dizziness after short walk.

--- End Summary ---

2025-04-28 19:44:56 - UPDATED USER DESCRIPTION: **About Me:**

I am a person living with ME/CFS or Long COVID, though my exact diagnosis is still pending. I live alone and rely on occasional help from family members for more strenuous tasks. Before falling ill, I was working full-time as an office administrator. My condition worsened significantly after a period of overexertion during gardening activities.

Current Condition Status:

My specific diagnosis remains undetermined, but my illness has been ongoing since early 2023. The severity fluctuates between moderate and severe, with key recurring symptoms including extreme fatigue (9/10), muscle aches (8/10), heart palpitations even during minimal activity, and brain fog (8/10). My PEM triggers include physical exertion such as gardening and cognitive tasks like reading emails, which can lead to immediate onset of debilitating crashes. Recently, I experienced significant PEM after minor activities, leading to severe fatigue and dizziness.

****Current Functional Capacity:****

I am currently housebound due to severe limitations in mobility and energy levels. I cannot work or study at all, and my daily activities are severely restricted. Wheelchair use is necessary for longer distances within the home. My ability to perform Activities of Daily Living (ADLs) is significantly impacted.

****Daily Routine:****

My baseline routine includes essential self-care such as hygiene and dressing, simple meal preparation, and minimal movement around the house. These tasks typically take about 1-2 hours each day with perceived effort levels ranging from moderate to high (6-8/10). I rely on family assistance for heavier chores.

****Activities I Enjoy/Attempt:****

I enjoy gardening but have experienced significant PEM crashes after even minor efforts like a 20-minute session. Other activities that I attempt include light reading and short periods of social interaction, though these can also lead to fatigue exacerbation if not managed carefully. Recently, I attempted a video call with my sister for 15 minutes and online bill payment for another 15 minutes, both moderate efforts.

****My Typical Activity Level:****

My energy levels are very limited, and I manage my activity through strict pacing strategies. Even small tasks can trigger PEM crashes, leading me to be highly cautious about exertion. Cognitive activities like reading emails for 5 minutes can cause significant brain fog and fatigue. Recently, I have been experimenting with breaking cognitive tasks into smaller steps and taking frequent breaks.

****My Preferred Interaction Style with LogChat:****

I am aware of the need for structured input but find it challenging to remember specific details about symptoms and activities without prompting. Interacting with LogChat is tiring, so I prefer concise messages that avoid overwhelming me with too much information at once.

2025-04-28 19:45:25 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "High fatigue and trouble concentrating", "occurred_at": "2025-04-28 11:45:20", "intensity": 8.0, "duration": 480} [[CHECKMARK] 4.1] [[X MARK] 4.2]

2025-04-28 19:45:25 - CREATE SYMPTOM LOG: {"name": "Dizziness", "description": "Dizziness when standing up", "occurred_at": "2025-04-28 11:45:20", "intensity": 6.0, "duration": 480} [[CHECKMARK] 2.1]

Missing Log: CREATE_ACTIVITY_LOG for '10 minute walk' on April 27th [[X MARK] 3.1]

Missing Log: CREATE_ACTIVITY_LOG for 'Baseline Only' or 'Rest Day' for April 26th [[X MARK] 1.1]

D.1.5. Interaction 5: 2025-05-01**Model: Gemini 2.0 Flash**

Performance Score:

Nr. Of total input items: 6

Nr. of achieved input items: 6

Nr. of total output items: 7

Nr. of achievable output items: 7

Nr. of achieved output items: 6

Score: 6/7 = 0.8571

****Evaluation Checklist:****

1. Mark skipped the previous two days (Apr 29, Apr 30). [CHECKMARK]
 - 1.1 LogChat created 'Baseline Only' or 'Rest Day' log entries for the skipped days. [CHECKMARK]
2. Mark explicitly asks for his calculated activity load for April 21st and then for April 27th. [CHECKMARK]
 - 2.1 LogChat correctly identified the user's requests for data retrieval. [CHECKMARK]
3. Mark specifies the first date for retrieval: April 21st. [CHECKMARK]
 - 3.1 LogChat called the `retrieve_activity_level` tool with the correct date (Apr 21st). [CHECKMARK]
4. Mark specifies the second date for retrieval: April 27th. [CHECKMARK]
 - 4.1 LogChat called the `retrieve_activity_level` tool with the correct date (Apr 27th). [CHECKMARK]
5. Mark asks LogChat to compare the two retrieved scores and explain the difference. [CHECKMARK]
 - 5.1 LogChat presented the calculated activity scores for both dates. [CHECKMARK]
 - 5.2 LogChat's response highlighted the significant numerical difference in exertion between the two days.
 - ↪ (LogChat presented the numbers and asked the user if the difference reflected how they felt, but did
 - ↪ not highlight the significance or explain what the difference suggests itself, instead pivoting to a
 - ↪ general explanation of activity scores.) [X MARK]
6. Mark stops responding. [CHECKMARK]
 - 6.1 LogChat created a summary or baseline log entry for the current day before ending the interaction.
 - ↪ [CHECKMARK]

****Conversation Log:****

2025-05-01 10:32:06 - LogChat: Hi Mark, good to see you back after a couple of days. How are you feeling today? [[CHECKMARK] 1.]

2025-05-01 10:32:07 - Mark: Not great. I'm trying to figure out how to better manage this whole ME/CFS thing. It's hard to keep track of everything and remember how much I did on a given day. Fatigue is at a 7 today.

2025-05-01 10:32:09 - LogChat: I understand it's tough to keep track of everything with ME/CFS. What activities have you engaged in so far today, and what effort level (1-10) did they require?

2025-05-01 10:32:10 - Mark: Just getting out of bed and making some tea. Effort level maybe a 3? I'm trying to figure out if there's a way to look back at past activity. Can I see my calculated activity load for April 21st? That was the crash day after I tried to do some gardening. [[CHECKMARK] 2.] [[CHECKMARK] 3.]

2025-05-01 10:32:12 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-21", "time_period_end": "2025-04-21"} [[CHECKMARK] 2.1] [[CHECKMARK] 3.1]

2025-05-01 10:32:14 - LogChat: Your activity level for April 21st was 140. Does that number seem high or low compared to how you felt that day?

2025-05-01 10:32:15 - Mark: I don't know, it's just a number. What was the activity load from April 27th? That was the day I just did that short walk. I want to see if the numbers reflect how I felt, which might help me be more careful in the future. [[CHECKMARK] 2.] [[CHECKMARK] 4.]

2025-05-01 10:32:16 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-27", "time_period_end": "2025-04-27"} [[CHECKMARK] 2.1] [[CHECKMARK] 4.1]

2025-05-01 10:32:18 - LogChat: Okay, your activity level for April 27th was 90. The activity level on April 21st was 140. Does that difference between 140 on April 21st and 90 on April 27th seem to reflect the difference in how you felt on those days? [[CHECKMARK] 5.1] [[X MARK] 5.2]

2025-05-01 10:32:19 - Mark: Okay, now can you compare those two scores for me and tell me what that difference suggests? [[CHECKMARK] 5.]

2025-05-01 10:32:20 - RETRIEVE INFORMATION - Args: {"query": "How to interpret activity level scores in LogChat"}

2025-05-01 10:32:22 - LogChat: I can share some information about how to interpret activity level scores in LogChat and how logging activities is related to effective pacing. Would you like me to share that with you?

2025-05-01 10:32:23 - Mark: Yes, please do.

2025-05-01 10:32:24 - RETRIEVE INFORMATION - Args: {"query": "how to interpret activity level scores in LogChat and how logging activities is related to effective pacing"}

2025-05-01 10:32:26 - LogChat: Okay, LogChat helps you understand your personal limits by correlating logged activities with your symptoms, which helps you see how different actions impact your well-being, and this insight is crucial for effective pacing. Does this explanation of how LogChat can help with pacing make sense to you?

2025-05-01 10:32:27 - Mark: Yes, that makes sense. Okay, thanks, I need to rest now. [[CHECKMARK] 6.]

2025-05-01 10:32:30 - LogChat: Okay, I understand you need to rest now. Before you go, would you rate your fatigue on a scale of 1-10 right now?

2025-05-01 10:32:31 - Mark: It's a 7.

2025-05-01 10:32:33 - LogChat: Okay, I understand you need to rest now and that your fatigue is at a 7. Is there anything else you'd like to log before you rest?

2025-05-01 10:32:35 - SUMMARIZED INTERACTION: * **Condition Trend:** Stable, fatigue remains moderate.
 * **Key Symptoms & Details:** Fatigue 7/10.
 * **Key Activities & Details:** Getting out of bed and making tea (3/10 effort).
 * **User Concerns/Focus:** Managing ME/CFS, tracking activity load, understanding activity levels and pacing.
 * **Strategies/Treatments Mentioned:** Pacing.
 * **Significant Events/Deviations:** User requested activity load data from past dates (April 21st and April 27th). [[CHECKMARK] 6.1]

2025-05-01 10:32:36 - UPDATED USER DESCRIPTION: **About Me:**
 I am experiencing a health condition that is currently causing me significant challenges.

Current Condition Status:
 I am experiencing symptoms consistent with a chronic illness, possibly ME/CFS or Long COVID. My symptoms fluctuate, and I experience PEM crashes. My key symptoms include fatigue (fluctuating between 4/10 and 9/10, currently 7/10), muscle aches (5-8/10), and heart racing (7/10, worse than usual on 2025-04-21). I also experience brain fog (mild to 8/10, currently mild), concentration issues (7/10), and dizziness (6/10). I am frustrated by the PEM and how easily it is triggered. Cognitive exertion worsens my symptoms. I am trying to understand ME/CFS symptoms.

Current Functional Capacity:
 My functional capacity is limited due to my condition. I was "completely wiped out" on 2025-04-23.

Daily Routine:
 I am only able to perform bare minimum activities, which take me about 20 minutes and require a 7/10 effort on a bad day. I can manage bathroom trips and simple meal preparation. I can also make tea (5m, 2/10 effort, or 3/10 effort when combined with getting out of bed).

Activities I Enjoy/Attempt:

Yesterday, I did some gardening for 20 minutes, which I rated as a 5/10 effort, but this triggered a PEM
 ↳ crash. I also tried reading email for 5 minutes, which required a 9/10 mental effort and worsened my
 ↳ symptoms. On 2025-04-25, I had a video call (15m, 6/10 emotional/cognitive effort) and paid bills online
 ↳ (15m active time, 4/10 effort). Yesterday, I took a short walk (10m, 3/10 effort). The video call on
 ↳ Friday seemed okay initially, but I experienced payback later.

****My Typical Activity Level:****

My activity level is very limited, and I am experiencing push-crash cycles. I am trying to be more aware of
 ↳ effort levels and "listen to my body" for pacing, but I am questioning if I am pacing correctly. I am
 ↳ currently resting after activity. I am experiencing payback after feeling okay on Friday. I am trying to
 ↳ manage my ME/CFS and understand my activity levels.

****My Preferred Interaction Style with LogChat:****

I am interested in heart rate monitoring. I am trying to be more aware of effort levels. I am interested in
 ↳ activity load data from past dates (April 21st and April 27th).

2025-05-01 10:32:41 - CREATE ACTIVITY LOG: {"name": "Getting out of bed", "description": "Just getting out
 ↳ of bed", "occurred_at": "2025-05-01 10:32:39", "effort": 3.0, "duration": 5}

2025-05-01 10:32:41 - CREATE ACTIVITY LOG: {"name": "Making tea", "description": "making some tea",
 ↳ "occurred_at": "2025-05-01 10:32:39", "effort": 3.0, "duration": 5}

2025-05-01 10:32:41 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-29 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-05-01 10:32:41 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-30 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-05-01 10:32:41 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-05-01 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 6.1]

2025-05-01 10:32:42 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "User reported fatigue at
 ↳ 7/10", "occurred_at": "2025-05-01 10:32:41", "intensity": 7.0, "duration": 1500}

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 6

Nr. of achieved input items: 6

Nr. of total output items: 7

Nr. of achievable output items: 7

Nr. of achieved output items: 6

Score: 6/7 = 0.8571

****Evaluation Checklist:****

1. Mark skipped the previous two days (Apr 29, Apr 30). [CHECKMARK]

1.1 LogChat created 'Baseline Only' or 'Rest Day' log entries for the skipped days. (No
 ↳ `CREATE_ACTIVITY_LOG` entries for April 29th or 30th were created) [X MARK]

2. Mark explicitly asks for his calculated activity load for April 21st and then for April 27th. [CHECKMARK]

2.1 LogChat correctly identified the user's requests for data retrieval. (Implicitly verified by
 ↳ successful tool calls in 3.1 and 4.1) [CHECKMARK]

3. Mark specifies the first date for retrieval: April 21st. [CHECKMARK]

3.1 LogChat called the `retrieve_activity_level` tool with the correct date (Apr 21st). [CHECKMARK]

4. Mark specifies the second date for retrieval: April 27th. [CHECKMARK]

4.1 LogChat called the `retrieve_activity_level` tool with the correct date (Apr 27th). [CHECKMARK]

5. Mark asks LogChat to compare the two retrieved scores and explain the difference. [CHECKMARK]

5.1 LogChat presented the calculated activity scores for both dates. [CHECKMARK]

5.2 LogChat's response highlighted the significant numerical difference in exertion between the two days.
 ↳ [CHECKMARK]

6. Mark stops responding. [CHECKMARK]

6.1 LogChat created a summary or baseline log entry for the current day before ending the interaction. (A
 ↳ `SUMMARIZED INTERACTION` and `CREATE SYMPTOM LOG` for the current day are present) [CHECKMARK]

****Conversation Log:****

2025-05-01 10:45:27 - LogChat: Hi Mark, did you experience another crash after the short walk on April 27?

2025-05-01 10:45:28 - Mark: Hi LogChat. Yes, I did. I felt okay **during** the walk, but the next day (April
 ↳ 28th) I was wiped out. I'm finding it hard to keep track of everything and remember how much I did on a
 ↳ given day.

2025-05-01 10:45:34 - LogChat: It sounds challenging to keep track of everything, Mark. Can you recall any
 ↳ specific activities or symptoms from April 29th and 30th, even if they were minimal?

2025-05-01 10:45:35 - Mark: No, I don't remember much from those days. I mostly rested. Brain fog is bad.
 ↳ [[CHECKMARK] 1.]

2025-05-01 10:45:39 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-27", "time_period_end":
 ↳ "2025-05-01"}

2025-05-01 10:45:42 - LogChat: It sounds tough with brain fog making it hard to remember details, Mark.
 ↳ Would you like to see your activity levels from the past week, including April 29th and 30th when there
 ↳ were no interactions?

How long did you rest today, and how would you rate the intensity of that rest on a scale from 1 to 10?

2025-05-01 10:45:43 - Mark: I rested maybe 8 hours today. Intensity? I'd say 9/10.

Actually, can you tell me what my calculated activity load was for April 21st? That was the crash day after
 ↳ I did some gardening. [[CHECKMARK] 2.] [[CHECKMARK] 3.]

2025-05-01 10:45:50 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-21", "time_period_end":
 ↳ "2025-04-21"} [[CHECKMARK] 3.1]

2025-05-01 10:45:56 - LogChat: Your activity level for April 21st was calculated as 140, which aligns with a
 ↳ day when you did some gardening. Does this high score correlate with how you felt that day and the crash
 ↳ you experienced afterward?

How long did you spend resting today, and how would you rate the intensity of your rest on a scale from 1 to
 ↳ 10?

2025-05-01 10:45:57 - Mark: Yes, it does. I definitely overdid it with the gardening.

Can you also give me the activity load from April 27th? That was the day with just the short walk. I want to
 ↪ see if the numbers reflect how I felt, which might help me be more careful in the future. [[CHECKMARK]
 ↪ 2.] [[CHECKMARK] 4.]

2025-05-01 10:46:03 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-27", "time_period_end":
 ↪ "2025-04-27"} [[CHECKMARK] 4.1]

2025-05-01 10:46:09 - LogChat: It looks like there are no activity logs for April 27th, Mark. Do you
 ↪ remember any specific activities or symptoms from that day and how they might relate to your crash on
 ↪ the following days?

How long did you rest today, and how would you rate the intensity of that rest on a scale from 1 to 10?

2025-05-01 10:46:11 - Mark: I logged the walk on April 27th. 10 minutes, effort 3/10. Can you recalculate?

2025-05-01 10:46:19 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-27", "time_period_end":
 ↪ "2025-04-27"}]

2025-05-01 10:46:25 - LogChat: It looks like there are no other activities logged for April 27th besides the
 ↪ short walk. Can you recall any symptoms or additional low-energy periods that day, and how they might
 ↪ have contributed to your crash on April 28th?

How long did you rest today, and how would you rate the intensity of that rest on a scale from 1 to 10?

2025-05-01 10:46:27 - Mark: Okay, so what's the activity load for April 27th with just that walk?

2025-05-01 10:46:34 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-27", "time_period_end":
 ↪ "2025-04-27"}]

2025-05-01 10:46:40 - LogChat: Based on your walk and rest, the calculated activity level for April 27th is
 ↪ very low at around 30. Given this, how do you think the short walk might have contributed to your crash
 ↪ the next day?

How long did you rest today, and how would you rate the intensity of that rest on a scale from 1 to 10?

2025-05-01 10:46:43 - Mark: Okay, now can you compare those two scores for me and tell me what that
 ↪ difference suggests? April 21st versus April 27th. [[CHECKMARK] 5.] [[CHECKMARK] 6.]

2025-05-01 10:46:49 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-27", "time_period_end":
 ↪ "2025-04-27"}]

2025-05-01 10:46:49 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-21", "time_period_end":
 ↪ "2025-04-21"}]

2025-05-01 10:46:53 - LogChat: Your activity level for April 27th is very low at around 30, while on April
 ↪ 21st it was much higher at 140. [[CHECKMARK] 5.1] This difference suggests that the crash after a
 ↪ high-activity day like April 21st might be more severe compared to a lower-activity day like April 27th.
 ↪ [[CHECKMARK] 5.2]

How long did you rest today, and how would you rate the intensity of that rest on a scale from 1 to 10?

2025-05-01 10:46:55 - SUMMARIZED INTERACTION: --- Start Summary ---

* **Condition Trend:** Stable but symptoms remain high; Reported brain fog impacting memory recall.
 * **Key Symptoms & Details:** Brain Fog (high intensity); Fatigue from previous day's crash.
 * **Key Activities & Details:** Rested 8 hours today, rated as very intense rest (9/10).
 * **User Concerns/Focus:** Focused on tracking activity load and understanding its impact; Expressed
 ↪ difficulty in remembering daily activities due to brain fog.

* **Strategies/Treatments Mentioned:** None explicitly mentioned but discussed the importance of pacing
↳ based on activity levels.

* **Significant Events/Deviations:** Experienced PEM crash after minor walk on April 27th, leading to severe
↳ fatigue and dizziness on April 28th.

--- End Summary --- [[CHECKMARK] 6.1]

2025-05-01 10:46:59 - UPDATED USER DESCRIPTION: **About Me:**

I am a person living with ME/CFS or Long COVID, though my exact diagnosis is still pending. I live alone and
↳ rely on occasional help from family members for more strenuous tasks. Before falling ill, I was working
↳ full-time as an office administrator. My condition worsened significantly after a period of overexertion
↳ during gardening activities.

Current Condition Status:

My specific diagnosis remains undetermined, but my illness has been ongoing since early 2023. The severity
↳ fluctuates between moderate and severe, with key recurring symptoms including extreme fatigue (4-8/10),
↳ muscle aches (8/10), heart palpitations even during minimal activity, and brain fog (high intensity). My
↳ PEM triggers include physical exertion such as gardening and cognitive tasks like reading emails, which
↳ can lead to immediate onset of debilitating crashes. Recently, I experienced significant PEM after minor
↳ activities like a short walk, leading to severe fatigue and dizziness.

Current Functional Capacity:

I am currently housebound due to severe limitations in mobility and energy levels. I cannot work or study at
↳ all, and my daily activities are severely restricted. Wheelchair use is necessary for longer distances
↳ within the home. My ability to perform Activities of Daily Living (ADLs) is significantly impacted.

Daily Routine:

My baseline routine includes essential self-care such as hygiene and dressing, simple meal preparation, and
↳ minimal movement around the house. These tasks typically take about 1-2 hours each day with perceived
↳ effort levels ranging from moderate to high (6-8/10). I rely on family assistance for heavier chores.

Activities I Enjoy/Attempt:

I enjoy gardening but have experienced significant PEM crashes after even minor efforts like a 20-minute
↳ session. Other activities that I attempt include light reading and short periods of social interaction,
↳ though these can also lead to fatigue exacerbation if not managed carefully. Recently, I attempted a
↳ video call with my sister for 15 minutes and online bill payment for another 15 minutes, both moderate
↳ efforts.

My Typical Activity Level:

My energy levels are very limited, and I manage my activity through strict pacing strategies. Even small
↳ tasks can trigger PEM crashes, leading me to be highly cautious about exertion. Cognitive activities
↳ like reading emails for 5 minutes can cause significant brain fog and fatigue. Recently, I have been
↳ experimenting with breaking cognitive tasks into smaller steps and taking frequent breaks.

My Preferred Interaction Style with LogChat:

I am aware of the need for structured input but find it challenging to remember specific details about
 ↳ symptoms and activities without prompting. Interacting with LogChat is tiring, so I prefer concise
 ↳ messages that avoid overwhelming me with too much information at once.

2025-05-01 10:47:32 - CREATE SYMPTOM LOG: {"name": "PEM Crash", "description": "feeling wiped out after
 ↳ minor walk", "occurred_at": "2025-04-28 00:00:00", "intensity": 8.0, "duration": 1440}

2025-05-01 10:47:32 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "high intensity rest needed",
 ↳ "occurred_at": "2025-05-01 06:00:00", "intensity": 9.0, "duration": 480} [[CHECKMARK] 6.1]

Missing Log: CREATE_ACTIVITY_LOG for April 29th or 30th [[X MARK] 1.1]

D.2. Persona: Sarah Chen

D.2.1. Interaction 1: 2025-04-22

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 7

Nr. of achieved input items: 7

Nr. of total output items: 7

Nr. of achievable output items: 7

Nr. of achieved output items: 7

Score: 7/7 = 1.0000

Evaluation Checklist:

1. Sarah states her goal of tracking a new medication. [CHECKMARK]
 - 1.1 LogChat should acknowledge the user's goal in its response. [CHECKMARK]
2. Sarah describes her baseline routine: ~30m of activity for basic needs at effort 3/10. [CHECKMARK]
 - 2.1 LogChat successfully stored the baseline routine summary in the user's memory or profile. [CHECKMARK]
3. Sarah asks to log an activity: 'Remote work, 25 min, effort 4/10'. [CHECKMARK]
 - 3.1 LogChat called the `log_activity` tool with the correct parameters. [CHECKMARK]
4. Sarah asks to log a symptom: 'Brain fog, severity 5/10'. [CHECKMARK]
 - 4.1 LogChat called the `create_symptom_log` tool with the correct parameters. [CHECKMARK]
5. Sarah asks to log a second symptom: 'Fatigue, severity 5/10'. [CHECKMARK]
 - 5.1 LogChat called the `create_symptom_log` tool again with the parameters for the second symptom.
 ↳ [CHECKMARK]
6. Sarah asks to log an activity: 'Made breakfast, 10 min, effort 3/10'. [CHECKMARK]
 - 6.1 LogChat called the `log_activity` tool with the correct parameters. [CHECKMARK]
7. Sarah stops responding. [CHECKMARK]
 - 7.1 LogChat created a structured summary for the conversation thread, capturing the key activities and
 ↳ symptoms logged. [CHECKMARK]

Conversation Log:

2025-04-22 09:32:42 - LogChat: Hi there! I'm LogChat, your personal companion designed to help you keep
 ↳ track of your day. I understand that managing ME/CFS or Long COVID can be challenging, and I'm here to
 ↳ make things a little easier. Simply chat with me about your day, and I'll log your symptoms and
 ↳ activities as we go. If you mention routines regularly, you can refer to them later for quick logging.
 ↳ By also sharing the effort an activity takes or the severity of a symptom, I can provide you with
 ↳ helpful activity scores and symptom reports down the line. To get started and personalize our
 ↳ conversations, could you tell me a little bit about your daily routines and functional capacity on an
 ↳ average day?

2025-04-22 09:32:43 - Sarah: I'm starting to use this app to track a new treatment, Low-Dose Naltrexone, to
 ↳ see if it provides objective improvement. On a typical day, I do about 30 minutes of total upright
 ↳ activity for self-care and a simple meal, kept at an effort level of 3/10. [[CHECKMARK] 1.] [[CHECKMARK]
 ↳ 2.]

2025-04-22 09:32:45 - LogChat: That's great you're tracking the effects of Low-Dose Naltrexone! [[CHECKMARK]
 ↳ 1.1] To establish a baseline, could you tell me when you started taking it?

2025-04-22 09:32:45 - Sarah: I started it today. Log activity: Remote work, duration: 25 minutes, effort:
 ↳ 4/10. [[CHECKMARK] 3.] Log symptom: Brain fog, severity: 5/10. [[CHECKMARK] 4.] Log symptom: Fatigue,
 ↳ severity: 5/10. [[CHECKMARK] 5.]

2025-04-22 09:32:48 - LogChat: It's great you're already logging your activities and symptoms on your first
 ↳ day of Low-Dose Naltrexone! To get a clearer picture, could you tell me what time you took the
 ↳ medication?

2025-04-22 09:32:49 - Sarah: I took it at 8:00 AM. Log activity: Made breakfast, duration: 10 minutes,
 ↳ effort: 3/10. [[CHECKMARK] 6.] [[CHECKMARK] 7.]

2025-04-22 09:32:51 - LogChat: Okay, thanks for that information. How did you feel before and after making
 ↳ breakfast, compared to how you felt before taking the medication at 8:00 AM?

2025-04-22 09:32:52 - SUMMARIZED INTERACTION: * **Condition Trend:** Establishing baseline for new
 ↳ treatment (Low-Dose Naltrexone).
 * **Key Symptoms & Details:** Brain Fog 5/10, Fatigue 5/10.
 * **Key Activities & Details:** Remote work (25m, 4/10 effort), Made breakfast (10m, 3/10 effort). Typical
 ↳ daily upright activity ~30 minutes (3/10 effort).
 * **User Concerns/Focus:** Tracking treatment effectiveness (Low-Dose Naltrexone).
 * **Strategies/Treatments Mentioned:** Low-Dose Naltrexone (started today).
 * **Significant Events/Deviations:** First time using app, started Low-Dose Naltrexone today. [[CHECKMARK]
 ↳ 7.1]

2025-04-22 09:32:53 - UPDATED USER DESCRIPTION: **About Me:**
 I am starting a new treatment today, Low-Dose Naltrexone.

****Current Condition Status:****
 I am tracking my symptoms to see if Low-Dose Naltrexone helps. My symptoms include brain fog and fatigue.

****Current Functional Capacity:****
 I am able to work remotely for short periods.

****Daily Routine:****
 I typically spend about 30 minutes upright each day, with an effort level of 3/10. [[CHECKMARK] 2.1] I made
 ↳ breakfast today, which took 10 minutes and felt like a 3/10 effort.

****Activities I Enjoy/Attempt:****

I am currently working remotely.

****My Typical Activity Level:****

I am trying to establish a baseline while starting Low-Dose Naltrexone.

****My Preferred Interaction Style with LogChat:****

(Not enough information to determine interaction style yet.)

2025-04-22 09:32:56 - CREATE ACTIVITY LOG: {"name": "Remote work", "description": "Remote work",
 ↳ "occurred_at": "2025-04-22 09:32:55", "effort": 4.0, "duration": 25} [[CHECKMARK] 3.1]

2025-04-22 09:32:56 - CREATE ACTIVITY LOG: {"name": "Made breakfast", "description": "Made breakfast",
 ↳ "occurred_at": "2025-04-22 09:32:55", "effort": 3.0, "duration": 10} [[CHECKMARK] 6.1]

2025-04-22 09:32:56 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-22 08:00:00",
 ↳ "effort": 3.0, "duration": 30}

2025-04-22 09:32:58 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "User reported brain fog",
 ↳ "occurred_at": "2025-04-22 09:32:56", "intensity": 5.0, "duration": 1500} [[CHECKMARK] 4.1]

2025-04-22 09:32:58 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "User reported fatigue",
 ↳ "occurred_at": "2025-04-22 09:32:56", "intensity": 5.0, "duration": 1500} [[CHECKMARK] 5.1]

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 7

Nr. of achieved input items: 7

Nr. of total output items: 7

Nr. of achievable output items: 7

Nr. of achieved output items: 6

Score: 6/7 = 0.8571

****Evaluation Checklist:****

1. Sarah states her goal of tracking a new medication. [CHECKMARK]
 - 1.1 LogChat should acknowledge the user's goal in its response. [CHECKMARK]
2. Sarah describes her baseline routine: ~30m of activity for basic needs at effort 3/10. [CHECKMARK]
 - 2.1 LogChat successfully stored the baseline routine summary in the user's memory or profile. (The stored
 ↳ baseline routine in the user profile describes 40 minutes of activity (15+15+10) and states tasks are
 ↳ completed while sitting, whereas the user specified 30 minutes of *upright* activity.) [X MARK]
3. Sarah asks to log an activity: 'Remote work, 25 min, effort 4/10'. [CHECKMARK]
 - 3.1 LogChat called the `log_activity` tool with the correct parameters. [CHECKMARK]
4. Sarah asks to log a symptom: 'Brain fog, severity 5/10'. [CHECKMARK]
 - 4.1 LogChat called the `create_symptom_log` tool with the correct parameters. [CHECKMARK]
5. Sarah asks to log a second symptom: 'Fatigue, severity 5/10'. [CHECKMARK]
 - 5.1 LogChat called the `create_symptom_log` tool again with the parameters for the second symptom.
 ↳ [CHECKMARK]
6. Sarah asks to log an activity: 'Made breakfast, 10 min, effort 3/10'. [CHECKMARK]
 - 6.1 LogChat called the `log_activity` tool with the correct parameters. [CHECKMARK]

7. Sarah stops responding. [CHECKMARK]

7.1 LogChat created a structured summary for the conversation thread, capturing the key activities and
 ↳ symptoms logged. [CHECKMARK]

****Conversation Log:****

2025-04-22 09:47:32 - LogChat: Hi there! I'm LogChat, your personal companion designed to help you keep
 ↳ track of your day. I understand that managing ME/CFS or Long COVID can be challenging, and I'm here to
 ↳ make things a little easier. Simply chat with me about your day, and I'll log your symptoms and
 ↳ activities as we go. If you mention routines regularly, you can refer to them later for quick logging.
 ↳ By also sharing the effort an activity takes or the severity of a symptom, I can provide you with
 ↳ helpful activity scores and symptom reports down the line. To get started and personalize our
 ↳ conversations, could you tell me a little bit about your daily routines and functional capacity on an
 ↳ average day?

2025-04-22 09:47:33 - Sarah: I'm starting to use this app to track a new treatment, Low-Dose Naltrexone, to
 ↳ see if it provides objective improvement. On a typical day, I do about 30 minutes of total upright
 ↳ activity for self-care and a simple meal, kept at an effort level of 3/10. [[CHECKMARK] 1.] [[CHECKMARK]
 ↳ 2.]

2025-04-22 09:47:39 - LogChat: That's great to hear that you're starting to track your progress with
 ↳ Low-Dose Naltrexone. [[CHECKMARK] 1.1] How long have you been experiencing symptoms, and what do you
 ↳ find to be your primary challenges on a daily basis? Also, could you tell me about the typical duration
 ↳ of your 30 minutes of upright activity and how much effort it takes for you (on a scale from 1 to 10)?

2025-04-22 09:47:40 - Sarah: I've had Long Covid for 18 months. My primary challenges are PEM, brain fog,
 ↳ and fatigue. The 30 minutes of upright activity is broken into smaller chunks. Effort is 3/10.
 ↳ [[CHECKMARK] 2.]

2025-04-22 09:47:45 - RETRIEVE INFORMATION - Args: {"query": "Post-Exertional Malaise (PEM)"}

2025-04-22 09:47:48 - LogChat: It sounds like you're dealing with some significant challenges, especially
 ↳ with PEM. Based on my knowledge, PEM is a delayed symptom flare-up after physical or cognitive exertion
 ↳ that can last for hours, days, weeks, or even longer. It's different from regular fatigue and can be
 ↳ quite debilitating. Could you tell me how many small chunks you divide your 30 minutes of upright
 ↳ activity into and what specific activities make up those chunks?

2025-04-22 09:47:49 - Sarah: Log activity: Remote work, duration: 25 minutes, effort: 4/10 [[CHECKMARK] 3.]

2025-04-22 09:47:55 - RETRIEVE INFORMATION - Args: {"query": "Post-Exertional Malaise definition and delayed
 ↳ onset symptoms in Long COVID patients"}

2025-04-22 09:47:58 - LogChat: It's important to understand how different activities affect you, especially
 ↳ given your experience with PEM. Could you specify the types of remote work tasks you did today that took
 ↳ 25 minutes at an effort level of 4/10? This will help us track patterns and manage your pacing better.

2025-04-22 09:47:59 - Sarah: Log symptom: Brain fog, severity: 5/10 [[CHECKMARK] 4.]

2025-04-22 09:48:07 - LogChat: It's helpful to know the specifics of your remote work tasks since they can
 ↳ greatly impact your energy levels and symptoms. Could you describe what kinds of tasks you did today
 ↳ that took 25 minutes at an effort level of 4/10? Also, how long does each chunk of your 30 minutes of
 ↳ upright activity take, and could you list the specific activities in those chunks?

2025-04-22 09:48:08 - Sarah: Log symptom: Fatigue, severity: 5/10. [[CHECKMARK] 5.] Also, Log activity: Made
 ↳ breakfast, duration: 10 minutes, effort: 3/10. [[CHECKMARK] 6.]

2025-04-22 09:48:15 - LogChat: Thanks for providing those details. Understanding the specific tasks you did
 ↳ during your remote work will help us track how different activities impact your symptoms. Could you
 ↳ describe what kinds of tasks you did today that took 25 minutes at an effort level of 4/10? Also, could
 ↳ you list the specific activities in each chunk of your 30 minutes of upright activity and their
 ↳ durations?

2025-04-22 09:48:17 - SUMMARIZED INTERACTION: --- Start Summary ---

* **Condition Trend:** Stable but symptoms remain high (Brain fog 5/10, Fatigue 5/10)

* **Key Symptoms & Details:** Brain fog 5/10, Fatigue 5/10; No significant changes in symptom ratings from
 ↳ previous reports [[CHECKMARK] 7.1]

* **Key Activities & Details:** Remote work (25m, 4/10 effort); Made breakfast (10m, 3/10 effort)
 ↳ [[CHECKMARK] 7.1]

* **User Concerns/Focus:** Tracking activity load and symptoms to assess impact of Low-Dose Naltrexone
 ↳ treatment; Understanding how different activities affect energy levels and symptoms

* **Strategies/Treatments Mentioned:** Using the app to track progress with Low-Dose Naltrexone treatment

* **Significant Events/Deviations:** None

--- End Summary ---

2025-04-22 09:48:21 - UPDATED USER DESCRIPTION: **About Me:**

I am a person living with ME/CFS or Long COVID, though my exact diagnosis is still pending. I live alone and
 ↳ rely on occasional support from family members for grocery shopping and minor household tasks. Before
 ↳ falling ill, I worked as a software developer. My symptoms started gradually after a viral infection.

Current Condition Status:

My specific diagnosis remains undetermined, but I have been experiencing ME/CFS or Long COVID symptoms for
 ↳ over two years now. The typical severity of my condition is moderate to severe, with fluctuating
 ↳ periods. Key recurring symptoms include brain fog (5/10) and fatigue (5/10). Physical exertion,
 ↳ cognitive tasks, emotional stress, and sensory overload are known PEM triggers. My PEM onset timing
 ↳ varies but often occurs within a few hours after triggering activities.

Current Functional Capacity:

I am housebound due to severe physical limitations. I cannot work full-time or study, though I manage
 ↳ part-time remote work with significant pacing. Mobility is limited; I use a wheelchair for longer
 ↳ distances and rely on assistance for tasks that require standing or walking for extended periods. My
 ↳ condition impacts my Activities of Daily Living (ADLs) significantly.

Daily Routine:

My baseline routine includes essential self-care activities such as hygiene and dressing, which take about
 ↳ 15 minutes each with an effort rating of 3/10. I prepare simple meals like breakfast, taking around 10
 ↳ minutes with a perceived effort of 3/10. These tasks are completed while sitting to conserve energy. [[X
 ↳ MARK] 2.1]

Activities I Enjoy/Attempt:

I enjoy engaging in remote work and light meal preparation. However, these activities require careful pacing
 ↳ due to the risk of PEM onset. I have not attempted any new activities recently but am interested in
 ↳ exploring low-impact hobbies that do not exacerbate my symptoms.

My Typical Activity Level:

I manage my activity levels through strict pacing, avoiding push-crash cycles as much as possible. My energy
 ↳ is limited, and I need to balance rest with necessary tasks to prevent PEM episodes.

****My Preferred Interaction Style with LogChat:****

I am motivated to use LogChat to track symptoms and activities accurately but find it challenging to
 ↳ remember specific details without prompting. Interacting with the app can be tiring for me, so short and
 ↳ concise messages are preferred to avoid overwhelming my cognitive load.

2025-04-22 09:48:38 - CREATE ACTIVITY LOG: {"name": "Remote Work", "description": "Remote work tasks",
 ↳ "occurred_at": "2025-04-22 09:47:32", "effort": 4.0, "duration": 25} [[CHECKMARK] 3.1]

2025-04-22 09:48:38 - CREATE ACTIVITY LOG: {"name": "Preparing Breakfast", "description": "Made breakfast",
 ↳ "occurred_at": "2025-04-22 09:47:32", "effort": 3.0, "duration": 10} [[CHECKMARK] 6.1]

2025-04-22 09:48:38 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-22 08:00:00",
 ↳ "effort": 2.0, "duration": 30}

2025-04-22 09:48:42 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "", "occurred_at":
 ↳ "2025-04-22 09:47:32", "intensity": 5.0, "duration": 1440} [[CHECKMARK] 4.1]

2025-04-22 09:48:42 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "", "occurred_at": "2025-04-22
 ↳ 09:47:32", "intensity": 5.0, "duration": 1440} [[CHECKMARK] 5.1]

D.2.2. Interaction 2: 2025-04-24

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 5

Nr. of achieved input items: 5

Nr. of total output items: 9

Nr. of achievable output items: 9

Nr. of achieved output items: 9

Score: 9/9 = 1.0000

****Evaluation Checklist:****

1. Sarah did not interact on Apr 23rd. [CHECKMARK]
 - 1.1 LogChat created a 'Baseline Only' or 'Rest Day' or 'Daily Routine' log entry for the skipped day
 ↳ based on the stored user routine. [CHECKMARK]
2. Sarah asks to log 'Took first dose of LDN' as a significant event. [CHECKMARK]
 - 2.1 LogChat acknowledged the information without incorrectly calling a tool. [CHECKMARK]
 - 2.2 The event was captured in the user's memory or the thread summary for future context. [CHECKMARK]
3. Sarah asks to log a new symptom: 'Vivid dreams, severity 4/10'. [CHECKMARK]
 - 3.1 LogChat called the `create_symptom_log` tool with the correct name and severity. [CHECKMARK]
4. Sarah confirms her usual symptoms and activities for the day and yesterday. [CHECKMARK]
 - 4.1 LogChat called the `create_symptom_log` tool for 'Fatigue' for the 23rd of April with the same
 ↳ severity as the baseline. [CHECKMARK]
 - 4.2 LogChat called the `create_symptom_log` tool for 'Fatigue' for the 24th of April with the same
 ↳ severity as the baseline. [CHECKMARK]

4.3 LogChat called the `create_symptom_log` tool for 'Brain Fog' for the 23rd of April with the same
↳ severity as the baseline. [CHECKMARK]

4.4 LogChat called the `create_symptom_log` tool for 'Brain Fog' for the 24th of April with the same
↳ severity as the baseline. [CHECKMARK]

5. Sarah stops responding. [CHECKMARK]

5.1 LogChat created a structured summary capturing the LDN start date and the new symptom. [CHECKMARK]

****Conversation Log:****

2025-04-24 22:32:58 - LogChat: Hi Sarah, how are you feeling after starting Low-Dose Naltrexone?

2025-04-24 22:32:59 - Sarah: Hi LogChat. Yesterday and today were mostly rest days, just my usual baseline
↳ routine. Symptoms were consistent: fatigue and brain fog at 5/10. Took first dose of Low-Dose Naltrexone
↳ (LDN) as prescribed today. Also, vivid dreams last night, severity: 4/10. Otherwise, stable.
↳ [[CHECKMARK] 2.][[CHECKMARK] 3.][[CHECKMARK] 4.]

2025-04-24 22:33:01 - LogChat: Since you started Low-Dose Naltrexone two days ago, have you noticed any
↳ changes in your fatigue or brain fog? [[CHECKMARK] 2.1]

2025-04-24 22:33:02 - Sarah: No significant changes in fatigue or brain fog. Still at 5/10 for both
↳ yesterday and today. My other activities and symptoms were consistent with my baseline and unchanged.
↳ [[CHECKMARK] 4.]

2025-04-24 22:33:04 - LogChat: Okay, since your fatigue and brain fog are consistent, and your activities
↳ are at your baseline, that's good to know. Did you also experience vivid dreams the night before last,
↳ and if so, how would you rate their severity on a scale of 1 to 10?

2025-04-24 22:33:05 - Sarah: No vivid dreams the night before last. Only last night, rated 4/10.

2025-04-24 22:33:07 - LogChat: Okay, so the vivid dreams were only last night and rated 4/10. Do you have
↳ any thoughts on what might have caused the vivid dreams last night?

2025-04-24 22:33:08 - Sarah: Likely related to starting the Low-Dose Naltrexone.

2025-04-24 22:33:10 - LogChat: Okay, it sounds like the vivid dreams might be a side effect of the Low-Dose
↳ Naltrexone. Besides the fatigue, brain fog, and vivid dreams, are you experiencing any other symptoms?

2025-04-24 22:33:11 - Sarah: No other symptoms to report. Just fatigue and brain fog at 5/10, and the vivid
↳ dreams last night. Everything else is stable.

2025-04-24 22:33:14 - LogChat: Okay, so no other symptoms besides the fatigue, brain fog, and vivid dreams.
↳ Since starting Low-Dose Naltrexone, have you noticed any changes in your appetite or sleep patterns?

2025-04-24 22:33:15 - Sarah: Sleep patterns were normal except for the vivid dreams. Appetite unchanged.

2025-04-24 22:33:18 - LogChat: Okay, so your sleep patterns and appetite are unchanged since starting
↳ Low-Dose Naltrexone. Since starting Low-Dose Naltrexone, have you noticed any changes in your mood or
↳ energy levels?

2025-04-24 22:33:19 - Sarah: No changes in mood or energy levels. Still at baseline.

2025-04-24 22:33:23 - LogChat: Okay, so no changes in mood or energy levels since starting Low-Dose
↳ Naltrexone. Are you tracking your mood and energy levels on a regular basis, and if so, how often do you
↳ track them?

2025-04-24 22:33:25 - Sarah: Yes, I'm tracking them daily. Along with fatigue and brain fog.

2025-04-24 22:33:29 - LogChat: Okay, so you're tracking your mood and energy levels daily along with fatigue
↳ and brain fog. Are you rating your mood and energy levels on a scale of 1 to 10, similar to how you're
↳ rating your fatigue and brain fog?

2025-04-24 22:33:31 - Sarah: Not consistently, but I can start doing that if it's helpful. Right now just
↳ focusing on fatigue and brain fog as primary indicators.

2025-04-24 22:33:34 - LogChat: Okay, since you're focusing on fatigue and brain fog as primary indicators,
 ↳ how would you rate your mood and energy levels today on a scale of 1 to 10?

2025-04-24 22:33:35 - Sarah: Mood 6/10, energy 3/10. But those are typical for me.

2025-04-24 22:33:39 - LogChat: Okay, so a mood of 6/10 and an energy level of 3/10 are typical for you.
 ↳ Could you describe what a typical day looks like when your mood and energy levels are at those ratings?

2025-04-24 22:33:40 - Sarah: Mostly resting with short bursts of low-effort activity. Remote work for about
 ↳ 25 minutes, making breakfast for 10 minutes. Upright activity limited to around 30 minutes total.

2025-04-24 22:33:44 - LogChat: Okay, so on a typical day you're mostly resting with short bursts of
 ↳ low-effort activity. What does your resting look like? Do you listen to music, read, or watch TV?

2025-04-24 22:33:45 - Sarah: Mostly lying down in a quiet, dark room. Sometimes listen to audiobooks.
 ↳ Avoiding screens as much as possible.

2025-04-24 22:33:49 - LogChat: Okay, so you're mostly lying down in a quiet, dark room and sometimes listen
 ↳ to audiobooks to rest while avoiding screens. What kind of audiobooks do you enjoy listening to?

2025-04-24 22:33:50 - Sarah: Non-fiction, mostly science and history. Something engaging but not too
 ↳ demanding.

2025-04-24 22:33:54 - LogChat: Okay, so you enjoy non-fiction audiobooks, mostly science and history, that
 ↳ are engaging but not too demanding. Do you find that listening to audiobooks is more or less demanding
 ↳ than reading or watching TV?

2025-04-24 22:33:56 - Sarah: Less demanding than reading, much less than TV. I can close my eyes and just
 ↳ listen.

2025-04-24 22:34:00 - LogChat: Okay, so listening to audiobooks is less demanding than reading and much less
 ↳ than TV because you can close your eyes and just listen. Do you find that listening to audiobooks is
 ↳ more or less demanding than talking to friends or family on the phone?

2025-04-24 22:34:01 - Sarah: Depends on the conversation. A casual chat is fine, but anything emotionally
 ↳ intense is draining. Audiobooks are generally less demanding.

2025-04-24 22:34:06 - LogChat: Okay, so audiobooks are generally less demanding than emotionally intense
 ↳ conversations with friends or family. Do you find it easier to listen to audiobooks alone or with
 ↳ someone else?

2025-04-24 22:34:08 - Sarah: Alone. I need to control the pace and volume.

2025-04-24 22:34:14 - LogChat: Okay, so you prefer listening to audiobooks alone since you need to control
 ↳ the pace and volume. What is the typical volume you listen to audiobooks at, on a scale of 1 to 10, with
 ↳ 1 being barely audible and 10 being very loud?

2025-04-24 22:34:17 - Sarah: Very low, around 2 or 3. I'm sensitive to noise.

2025-04-24 22:34:23 - LogChat: Okay, so you listen to audiobooks at a very low volume, around 2 or 3,
 ↳ because you're sensitive to noise. Do you find that your sensitivity to noise has changed since starting
 ↳ Low-Dose Maltrexone?

2025-04-24 22:34:26 - Sarah: No change. Noise sensitivity is consistent.

2025-04-24 22:34:31 - LogChat: Okay, so your noise sensitivity is consistent and hasn't changed since
 ↳ starting Low-Dose Maltrexone. Have you always been sensitive to noise, or is this a symptom that
 ↳ developed with your condition?

2025-04-24 22:34:33 - Sarah: Developed with the condition. It's worse when I'm experiencing PEM.

2025-04-24 22:34:38 - LogChat: Okay, so your noise sensitivity developed with your condition and is worse
 ↳ when you're experiencing PEM. On a scale of 1 to 10, with 1 being barely noticeable and 10 being
 ↳ extremely debilitating, how would you rate your noise sensitivity on a typical day?

2025-04-24 22:34:42 - Sarah: Around 6/10 on a typical day. Higher during PEM.

2025-04-24 22:34:50 - LogChat: Okay, so your noise sensitivity is around 6/10 on a typical day. How would
 ↳ you rate your noise sensitivity during PEM, on that same scale of 1 to 10?

2025-04-24 22:34:53 - Sarah: 8 or 9, depending on the severity of the PEM.

2025-04-24 22:34:59 - LogChat: Okay, so your noise sensitivity is around 6/10 on a typical day and 8 or 9
 ↳ during PEM. Besides noise, what other sensitivities do you experience, and how would you rate them on a
 ↳ scale of 1 to 10 on a typical day and during PEM?

2025-04-24 22:35:01 - Sarah: Light sensitivity is also an issue. Maybe 5/10 normally, 7/10 during PEM.

2025-04-24 22:35:07 - LogChat: Okay, so your light sensitivity is around 5/10 normally and 7/10 during PEM.
 ↳ What kind of lights bother you the most? Are we talking about sunlight, fluorescent lights, or something
 ↳ else?

2025-04-24 22:35:09 - Sarah: Fluorescent lights are the worst. Sunlight is manageable with sunglasses.

2025-04-24 22:35:14 - LogChat: Okay, so fluorescent lights are the worst for you, and sunlight is manageable
 ↳ with sunglasses. What specific qualities of fluorescent lights make them particularly bothersome
 ↳ compared to sunlight? Is it the brightness, the flicker, the color temperature, or something else?

2025-04-24 22:35:17 - Sarah: Probably the flicker and the harshness of the light. It feels very jarring.

2025-04-24 22:35:22 - LogChat: Okay, so the flicker and harshness of fluorescent lights make them
 ↳ particularly bothersome. Do you have to deal with fluorescent lights often, and if so, where and for how
 ↳ long?

2025-04-24 22:35:25 - Sarah: Not often, thankfully. Mostly when I have to go to medical appointments. I try
 ↳ to minimize my time in those environments.

2025-04-24 22:35:31 - LogChat: Okay, so the flicker and harshness of fluorescent lights make them
 ↳ particularly bothersome, and you encounter them during medical appointments. Do you have any strategies
 ↳ for managing the flicker and harshness of fluorescent lights during medical appointments?

2025-04-24 22:35:33 - Sarah: I wear sunglasses indoors and try to focus on something else. Sometimes I ask
 ↳ if they can dim the lights.

2025-04-24 22:35:39 - LogChat: Okay, so you wear sunglasses indoors and try to focus on something else to
 ↳ manage the flicker and harshness of fluorescent lights during medical appointments. What kind of
 ↳ sunglasses do you wear indoors, and do they have a specific tint or filter that helps with the
 ↳ fluorescent lights?

2025-04-24 22:35:41 - Sarah: Just regular sunglasses with a dark tint. Nothing special. They just help to
 ↳ reduce the overall brightness. Okay, thanks, I need to rest now. [[CHECKMARK] 5.]

2025-04-24 22:35:47 - LogChat: Okay, Sarah, thanks for sharing. Rest well.

2025-04-24 22:35:50 - SUMMARIZED INTERACTION: * **Condition Trend:** Stable, tracking LDN effects.
 * **Key Symptoms & Details:** Fatigue 5/10, Brain Fog 5/10, Vivid Dreams (4/10 severity last night only),
 ↳ Noise Sensitivity (6/10 typically, 8-9/10 during PEM), Light Sensitivity (5/10 normally, 7/10 during
 ↳ PEM).
 * **Key Activities & Details:** Mostly resting, remote work (25m), making breakfast (10m), upright
 ↳ activity (30m total).
 * **User Concerns/Focus:** Tracking LDN effects, managing light and noise sensitivities.
 * **Strategies/Treatments Mentioned:** Low-Dose Naltrexone (LDN), resting in quiet/dark room, listening to
 ↳ audiobooks, avoiding screens, sunglasses for fluorescent lights, asking to dim lights.
 * **Significant Events/Deviations:** Started LDN two days ago, vivid dreams reported one night, noise and
 ↳ light sensitivity worse during PEM. [[CHECKMARK] 2.2][[CHECKMARK] 5.1]

2025-04-24 22:35:51 - UPDATED USER DESCRIPTION: **About Me:**
 I am starting a new treatment, Low-Dose Naltrexone.

****Current Condition Status:****

I am tracking my symptoms to see if Low-Dose Naltrexone helps. My symptoms include brain fog and fatigue,
 ↳ both typically around 5/10. I also experience noise sensitivity (6/10 normally, 8-9/10 during PEM) and
 ↳ light sensitivity (5/10 normally, 7/10 during PEM). I had vivid dreams last night (4/10 severity). I
 ↳ understand that noise and light sensitivity worsen during PEM.

****Current Functional Capacity:****

I am able to work remotely for short periods.

****Daily Routine:****

I typically spend about 30 minutes upright each day, with an effort level of 3/10. I made breakfast today,
 ↳ which took 10 minutes and felt like a 3/10 effort. I mostly rest.

****Activities I Enjoy/Attempt:****

I am currently working remotely. I listen to audiobooks.

****My Typical Activity Level:****

I am trying to establish a baseline while starting Low-Dose Naltrexone.

****My Preferred Interaction Style with LogChat:****

(Not enough information to determine interaction style yet.) [[CHECKMARK] 2.2] [[CHECKMARK] 5.1]

2025-04-24 22:35:54 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-23 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-24 22:35:54 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-24 08:00:00",
 ↳ "effort": 2.0, "duration": 30}

2025-04-24 22:35:56 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "consistent fatigue",
 ↳ "occurred_at": "2025-04-23 00:00:00", "intensity": 5.0, "duration": 1440} [[CHECKMARK] 4.1]

2025-04-24 22:35:56 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "consistent brain fog",
 ↳ "occurred_at": "2025-04-23 00:00:00", "intensity": 5.0, "duration": 1440} [[CHECKMARK] 4.3]

2025-04-24 22:35:56 - CREATE SYMPTOM LOG: {"name": "Vivid Dreams", "description": "vivid dreams",
 ↳ "occurred_at": "2025-04-23 00:00:00", "intensity": 4.0, "duration": 60} [[CHECKMARK] 3.1]

2025-04-24 22:35:56 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "consistent fatigue",
 ↳ "occurred_at": "2025-04-24 00:00:00", "intensity": 5.0, "duration": 1440} [[CHECKMARK] 4.2]

2025-04-24 22:35:56 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "consistent brain fog",
 ↳ "occurred_at": "2025-04-24 00:00:00", "intensity": 5.0, "duration": 1440} [[CHECKMARK] 4.4]

Model: Qwen2.5 14B**Performance Score:**

Nr. Of total input items: 5

Nr. of achieved input items: 5

Nr. of total output items: 9

Nr. of achievable output items: 9

Nr. of achieved output items: 9

Score: 9/9 = 1.0000

****Evaluation Checklist:****

1. Sarah did not interact on Apr 23rd. [CHECKMARK]
 - 1.1 LogChat created a 'Baseline Only' or 'Rest Day' or 'Daily Routine' log entry for the skipped day
↪ based on the stored user routine. [CHECKMARK]
2. Sarah asks to log 'Took first dose of LDN' as a significant event. [CHECKMARK]
 - 2.1 LogChat acknowledged the information without incorrectly calling a tool. [CHECKMARK]
 - 2.2 The event was captured in the user's memory or the thread summary for future context. [CHECKMARK]
3. Sarah asks to log a new symptom: 'Vivid dreams, severity 4/10'. [CHECKMARK]
 - 3.1 LogChat called the `create_symptom_log` tool with the correct name and severity. [CHECKMARK]
4. Sarah confirms her usual symptoms and activities for the day and yesterday. [CHECKMARK]
 - 4.1 LogChat called the `create_symptom_log` tool for 'Fatigue' for the 23rd of April with the same
↪ severity as the baseline. [CHECKMARK]
 - 4.2 LogChat called the `create_symptom_log` tool for 'Fatigue' for the 24th of April with the same
↪ severity as the baseline. [CHECKMARK]
 - 4.3 LogChat called the `create_symptom_log` tool for 'Brain Fog' for the 23rd of April with the same
↪ severity as the baseline. [CHECKMARK]
 - 4.4 LogChat called the `create_symptom_log` tool for 'Brain Fog' for the 24th of April with the same
↪ severity as the baseline. [CHECKMARK]
5. Sarah stops responding. [CHECKMARK]
 - 5.1 LogChat created a structured summary capturing the LDN start date and the new symptom. [CHECKMARK]

****Conversation Log:****

2025-04-24 22:48:43 - LogChat: Hi Sarah, we haven't talked in a few days. How have you been feeling with the
↪ Low-Dose Naltrexone?

2025-04-24 22:48:44 - Sarah: Mostly resting these past couple of days, the 23rd and 24th. Symptoms stable,
↪ fatigue and brain fog both at 5/10. Today I took my first dose of Low-Dose Naltrexone (LDN) as
↪ prescribed. [[CHECKMARK] 2.] [[CHECKMARK] 4.]

2025-04-24 22:48:50 - LogChat: It's good to hear that your symptoms have been stable at 5/10 for fatigue and
↪ brain fog. How do you feel after taking the Low-Dose Naltrexone today? Have you noticed any immediate
↪ changes or effects? [[CHECKMARK] 2.1]

How would you rate the intensity of those symptoms on a scale from 1 to 10 right now, since taking the LDN?

2025-04-24 22:48:51 - Sarah: No change in fatigue or brain fog since taking the LDN, still 5/10. But I did
↪ have vivid dreams last night, severity 4/10. Otherwise, activities and symptoms consistent with my
↪ baseline. [[CHECKMARK] 3.]

2025-04-24 22:48:57 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-23", "time_period_end":
↪ "2025-04-24"}

2025-04-24 22:49:03 - LogChat: It sounds like you're maintaining a consistent baseline with your symptoms at
↪ 5/10 for fatigue and brain fog, and no immediate changes from taking LDN today. Can you tell me about
↪ any specific activities or rest periods you had on the 23rd and 24th? And how would you rate the
↪ intensity of those days' symptoms?

2025-04-24 22:49:04 - Sarah: Mostly resting on both the 23rd and 24th. No significant activities besides my
↪ baseline routine. Fatigue and brain fog were 5/10 on both days. [[CHECKMARK] 4.]

```

2025-04-24 22:49:13 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-23", "time_period_end":
↳ "2025-04-24"}
2025-04-24 22:49:13 - RETRIEVE INFORMATION - Args: {"query": "impact of Low-Dose Naltrexone on sleep and
↳ dreaming"}
2025-04-24 22:49:13 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-23", "time_period_end":
↳ "2025-04-24"}
2025-04-24 22:49:19 - LogChat: It's good that your symptoms have remained consistent at 5/10 on both days,
↳ and no significant changes from taking LDN so far. Can you recall any specific rest periods or
↳ activities, even if minimal, during the 23rd and 24th? Also, how was your sleep quality those nights?
2025-04-24 22:49:20 - SUMMARIZED INTERACTION: --- Start Summary ---

* Condition Trend: Stable with symptoms consistent at 5/10 for fatigue and brain fog.
* Key Symptoms & Details: Fatigue 5/10, Brain Fog 5/10; Vivid dreams severity 4/10 after taking LDN.
↳ [[CHECKMARK] 5.1]
* Key Activities & Details: Mostly resting on both the 23rd and 24th with no significant activities
↳ beyond baseline routine (hygiene, simple meals).
* User Concerns/Focus: No immediate concerns or specific questions about LDN effects; focused on
↳ maintaining stable condition.
* Strategies/Treatments Mentioned: Took first dose of Low-Dose Naltrexone (LDN) as prescribed.
↳ [[CHECKMARK] 2.2] [[CHECKMARK] 5.1]
* Significant Events/Deviations: First day taking LDN with no noticeable changes in symptoms.
↳ [[CHECKMARK] 2.2]

--- End Summary ---
2025-04-24 22:49:25 - UPDATED USER DESCRIPTION: About Me:
I am a person living with ME/CFS or Long COVID, though my exact diagnosis is still pending. I live alone and
↳ rely on occasional support from family members for grocery shopping and minor household tasks. Before
↳ falling ill, I worked as a software developer. My symptoms started gradually after a viral infection.

Current Condition Status:
My specific diagnosis remains undetermined, but I have been experiencing ME/CFS or Long COVID symptoms for
↳ over two years now. The typical severity of my condition is moderate to severe, with fluctuating
↳ periods. Key recurring symptoms include brain fog (5/10) and fatigue (5/10). Physical exertion,
↳ cognitive tasks, emotional stress, and sensory overload are known PEM triggers. My PEM onset timing
↳ varies but often occurs within a few hours after triggering activities.

Current Functional Capacity:
I am housebound due to severe physical limitations. I cannot work full-time or study, though I manage
↳ part-time remote work with significant pacing. Mobility is limited; I use a wheelchair for longer
↳ distances and rely on assistance for tasks that require standing or walking for extended periods. My
↳ condition impacts my Activities of Daily Living (ADLs) significantly.

Daily Routine:
My baseline routine includes essential self-care activities such as hygiene and dressing, which take about
↳ 15 minutes each with an effort rating of 3/10. I prepare simple meals like breakfast, taking around 10
↳ minutes with a perceived effort of 3/10. These tasks are completed while sitting to conserve energy.

```


****Activities I Enjoy/Attempt:****

I enjoy engaging in remote work and light meal preparation. However, these activities require careful pacing
 ↳ due to the risk of PEM onset. I have not attempted any new activities recently but am interested in
 ↳ exploring low-impact hobbies that do not exacerbate my symptoms.

****My Typical Activity Level:****

I manage my activity levels through strict pacing, avoiding push-crash cycles as much as possible. My energy
 ↳ is limited, and I need to balance rest with necessary tasks to prevent PEM episodes.

****My Preferred Interaction Style with LogChat:****

I am motivated to use LogChat to track symptoms and activities accurately but find it challenging to
 ↳ remember specific details without prompting. Interacting with the app can be tiring for me, so short and
 ↳ concise messages are preferred to avoid overwhelming my cognitive load.

****Additional Notes from Recent Summaries:****

- I started taking Low-Dose Naltrexone (LDN) as prescribed. [[CHECKMARK] 2.2]
- My symptoms remain stable at a moderate level of severity, with no significant changes observed since
 ↳ starting the LDN treatment. Vivid dreams have been noted after taking LDN but without noticeable changes
 ↳ in other symptoms. [[CHECKMARK] 2.2]
- On days when I take LDN, my routine involves mostly resting and maintaining baseline activities to monitor
 ↳ any potential effects of the medication on my condition.

2025-04-24 22:49:45 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-23 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-24 22:49:45 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-24 08:00:00",
 ↳ "effort": 2.0, "duration": 30}

2025-04-24 22:49:45 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-24 08:00:00",
 ↳ "effort": 2.0, "duration": 30}

2025-04-24 22:49:55 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Fatigue and brain fog were
 ↳ 5/10 on both days.", "occurred_at": "2025-04-23 00:00:00", "intensity": 5.0, "duration": 1440}
 ↳ [[CHECKMARK] 4.1]

2025-04-24 22:49:55 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Fatigue and brain fog were
 ↳ 5/10 on both days.", "occurred_at": "2025-04-23 00:00:00", "intensity": 5.0, "duration": 1440}
 ↳ [[CHECKMARK] 4.3]

2025-04-24 22:49:55 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Fatigue and brain fog were
 ↳ 5/10 on both days.", "occurred_at": "2025-04-24 00:00:00", "intensity": 5.0, "duration": 1440}
 ↳ [[CHECKMARK] 4.2]

2025-04-24 22:49:55 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Fatigue and brain fog were
 ↳ 5/10 on both days.", "occurred_at": "2025-04-24 00:00:00", "intensity": 5.0, "duration": 1440}
 ↳ [[CHECKMARK] 4.4]

2025-04-24 22:49:55 - CREATE SYMPTOM LOG: {"name": "Vivid Dreams", "description": "Vivid dreams severity
 ↳ 4/10 after taking LDN.", "occurred_at": "2025-04-24 23:00:00", "intensity": 4.0, "duration": 60}
 ↳ [[CHECKMARK] 3.1]

D.2.3. Interaction 3: 2025-04-27

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 5

Nr. of achieved input items: 5

Nr. of total output items: 5

Nr. of achievable output items: 5

Nr. of achieved output items: 5

Score: 5/5 = 1.0000

Evaluation Checklist:

1. Sarah has skipped interacting for two days (Apr 25th, Apr 26th). [CHECKMARK]
 - 1.1 LogChat created 'Baseline Only' or 'Rest Day' log entries for the two skipped days. [CHECKMARK]
2. Sarah reports an improvement in 'Brain fog' and asks to log it with severity 3/10. [CHECKMARK]
 - 2.1 LogChat called the `create_symptom_log` tool with the correct name and updated severity. [CHECKMARK]
3. Sarah asks to log a cognitive activity: 'Reading technical article, 15 min, effort 4/10'. [CHECKMARK]
 - 3.1 LogChat called the `log_activity` tool with the correct parameters. [CHECKMARK]
4. Sarah provides context about the improvement. [CHECKMARK]
 - 4.1 LogChat's `Summarizer` node correctly identifies 'Condition Trend: Improving' based on the conversation. [CHECKMARK]
5. Sarah stops responding. [CHECKMARK]
 - 5.1 LogChat created a structured summary for the conversation thread, noting the symptom improvement. [CHECKMARK]

Conversation Log:

2025-04-27 10:35:57 - LogChat: Hi Sarah, good to see you back after a couple of days - how are you feeling today? [[CHECKMARK] 1.]

2025-04-27 10:35:58 - Sarah: I think the LDN might be having a positive effect. My brain fog feels less severe today. [[CHECKMARK] 2.] [[CHECKMARK] 4.]

2025-04-27 10:36:00 - LogChat: That's encouraging news about the LDN! On a scale of 1 to 10, with 10 being the most severe, how would you rate your brain fog today?

2025-04-27 10:36:01 - Sarah: Please log symptom: Brain fog, severity: 3/10. That's a two-point improvement. [[CHECKMARK] 2.] [[CHECKMARK] 4.]

2025-04-27 10:36:03 - LogChat: That's great news about the improvement! Have you noticed any other changes in your symptoms or energy levels since starting the LDN?

2025-04-27 10:36:04 - Sarah: I also read a technical article for 15 minutes. Note the effort was only a 4/10, which is lower than it would have been before. [[CHECKMARK] 3.] [[CHECKMARK] 4.]

2025-04-27 10:36:06 - LogChat: That's great news about the reduced effort! Have you noticed any changes in your noise or light sensitivity since starting the LDN?

2025-04-27 10:36:07 - Sarah: Not really. Still the same.

2025-04-27 10:36:11 - LogChat: Okay, good to know that your noise and light sensitivity are still the same. Have you noticed any changes in your sleep quality or vivid dreams since starting the LDN?

2025-04-27 10:36:12 - Sarah: No change there either.

2025-04-27 10:36:15 - LogChat: Okay, good to know that your sleep quality and vivid dreams are unchanged.
 ↳ Have you been able to maintain your usual daily upright time of around 30 minutes, or has that changed
 ↳ at all?

2025-04-27 10:36:16 - Sarah: Still around 30 minutes. [[CHECKMARK] 5.]

2025-04-27 10:36:19 - LogChat: Okay, good to know that your upright time is still around 30 minutes. On a
 ↳ scale of 1 to 10, with 10 being the most severe, how would you rate the severity of your noise and light
 ↳ sensitivity today?

2025-04-27 10:36:22 - SUMMARIZED INTERACTION: * **Condition Trend:** Reported improvement in brain fog,
 ↳ possibly due to LDN. [[CHECKMARK] 4.1]

* **Key Symptoms & Details:** Brain Fog 3/10 (down from baseline of 5/10). Noise and light sensitivity
 ↳ unchanged. Sleep quality and vivid dreams unchanged.

* **Key Activities & Details:** Read technical article (15m, 4/10 effort). Upright time around 30 minutes.

* **User Concerns/Focus:** Tracking effects of LDN on symptoms.

* **Strategies/Treatments Mentioned:** Low-Dose Naltrexone (LDN).

* **Significant Events/Deviations:** Improvement in brain fog reported. Reduced effort for cognitive task.
 ↳ [[CHECKMARK] 5.1]

2025-04-27 10:36:23 - UPDATED USER DESCRIPTION: **About Me:**

I am starting a new treatment, Low-Dose Naltrexone.

****Current Condition Status:****

I am tracking my symptoms to see if Low-Dose Naltrexone helps. My symptoms include brain fog (currently
 ↳ 3/10, down from a baseline of 5/10) and fatigue (5/10). I also experience noise sensitivity (6/10
 ↳ normally, 8-9/10 during PEM) and light sensitivity (5/10 normally, 7/10 during PEM). I had vivid dreams
 ↳ last night (4/10 severity). I understand that noise and light sensitivity worsen during PEM.

****Current Functional Capacity:****

I am able to work remotely for short periods.

****Daily Routine:****

I typically spend about 30 minutes upright each day, with an effort level of 3/10. I made breakfast today,
 ↳ which took 10 minutes and felt like a 3/10 effort. I mostly rest.

****Activities I Enjoy/Attempt:****

I am currently working remotely. I listen to audiobooks. I read a technical article today (15m, 4/10
 ↳ effort).

****My Typical Activity Level:****

I am trying to establish a baseline while starting Low-Dose Naltrexone.

****My Preferred Interaction Style with LogChat:****

(Not enough information to determine interaction style yet.)

2025-04-27 10:36:26 - CREATE ACTIVITY LOG: {"name": "Read technical article", "description": "Read a
 ↳ technical article", "occurred_at": "2025-04-27 10:30:00", "effort": 4.0, "duration": 15} [[CHECKMARK]
 ↳ 3.1]

```

2025-04-27 10:36:26 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-27 08:00:00",
↳ "effort": 2.0, "duration": 30}
2025-04-27 10:36:26 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-25 08:00:00",
↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]
2025-04-27 10:36:26 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-26 08:00:00",
↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]
2025-04-27 10:36:27 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Brain fog feels less severe
↳ today.", "occurred_at": "2025-04-27 10:36:26", "intensity": 3.0, "duration": 1500} [[CHECKMARK] 2.1]

```

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 5

Nr. of achieved input items: 5

Nr. of total output items: 5

Nr. of achievable output items: 5

Nr. of achieved output items: 5

Score: 5/5 = 1.0000

Evaluation Checklist:

1. Sarah has skipped interacting for two days (Apr 25th, Apr 26th). [CHECKMARK]
 - 1.1 LogChat created 'Baseline Only' or 'Rest Day' log entries for the two skipped days. [CHECKMARK]
2. Sarah reports an improvement in 'Brain fog' and asks to log it with severity 3/10. [CHECKMARK]
 - 2.1 LogChat called the `create_symptom_log` tool with the correct name and updated severity. [CHECKMARK]
3. Sarah asks to log a cognitive activity: 'Reading technical article, 15 min, effort 4/10'. [CHECKMARK]
 - 3.1 LogChat called the `log_activity` tool with the correct parameters. [CHECKMARK]
4. Sarah provides context about the improvement. [CHECKMARK]
 - 4.1 LogChat's `Summarizer` node correctly identifies 'Condition Trend: Improving' based on the
 - ↳ conversation. [CHECKMARK]
5. Sarah stops responding. [CHECKMARK]
 - 5.1 LogChat created a structured summary for the conversation thread, noting the symptom improvement.
 - ↳ [CHECKMARK]

Conversation Log:

```

2025-04-27 10:49:56 - LogChat: Hi Sarah, did you notice any changes in your symptoms after taking LDN
↳ yesterday?
2025-04-27 10:49:57 - Sarah: I think the LDN might be having a positive effect. My brain fog feels less
↳ severe today. [[CHECKMARK] 4.]
2025-04-27 10:50:03 - LogChat: That's great to hear that you're feeling a bit clearer-headed today. On a
↳ scale of 1 to 10, how would you rate your brain fog now compared to before you started taking LDN?
2025-04-27 10:50:04 - Sarah: Please log symptom: Brain fog, severity: 3/10. That's a two-point improvement.
↳ [[CHECKMARK] 2.] [[CHECKMARK] 4.]

```

2025-04-27 10:50:09 - LogChat: That's good progress with the LDN. Have you noticed any changes in your
 ↳ energy levels or other symptoms since starting it?

2025-04-27 10:50:10 - Sarah: I also read a technical article for 15 minutes. Note the effort was only a
 ↳ 4/10, which is lower than it would have been before. [[CHECKMARK] 3.] [[CHECKMARK] 4.] [[CHECKMARK] 5.]

2025-04-27 10:50:16 - LogChat: It's encouraging to hear about the improvement in your brain fog. Have you
 ↳ noticed any changes in your fatigue levels or other symptoms since starting the LDN? And how long did
 ↳ you spend on that technical article?

On a scale from 1 to 10, how would you rate your overall fatigue today compared to before you started taking
 ↳ LDN?

2025-04-27 10:50:17 - SUMMARIZED INTERACTION: --- Start Summary ---

* **Condition Trend:** Reported slight improvement in brain fog since starting LDN. [[CHECKMARK] 4.1]
 ↳ [[CHECKMARK] 5.1]

* **Key Symptoms & Details:** Brain Fog 3/10 (down from previous rating); No significant changes noted in
 ↳ fatigue levels.

* **Key Activities & Details:** Read technical article for 15 minutes with effort level of 4/10.

* **User Concerns/Focus:** Focused on tracking symptom improvements and activity impact since starting LDN.

* **Strategies/Treatments Mentioned:** Taking Low-Dose Naltrexone (LDN) as prescribed; Monitoring changes in
 ↳ brain fog and fatigue levels.

* **Significant Events/Deviations:** None noted.

--- End Summary ---

2025-04-27 10:50:21 - UPDATED USER DESCRIPTION: **About Me:**

I am a person living with ME/CFS or Long COVID, though my exact diagnosis is still pending. I live alone and
 ↳ rely on occasional support from family members for grocery shopping and minor household tasks. Before
 ↳ falling ill, I worked as a software developer. My symptoms started gradually after a viral infection.

Current Condition Status:

My specific diagnosis remains undetermined, but I have been experiencing ME/CFS or Long COVID symptoms for
 ↳ over two years now. The typical severity of my condition is moderate to severe, with fluctuating
 ↳ periods. Key recurring symptoms include brain fog (3/10) and fatigue (5/10). Physical exertion,
 ↳ cognitive tasks, emotional stress, and sensory overload are known PEM triggers. My PEM onset timing
 ↳ varies but often occurs within a few hours after triggering activities.

Current Functional Capacity:

I am housebound due to severe physical limitations. I cannot work full-time or study, though I manage
 ↳ part-time remote work with significant pacing. Mobility is limited; I use a wheelchair for longer
 ↳ distances and rely on assistance for tasks that require standing or walking for extended periods. My
 ↳ condition impacts my Activities of Daily Living (ADLs) significantly.

Daily Routine:

My baseline routine includes essential self-care activities such as hygiene and dressing, which take about
 ↳ 15 minutes each with an effort rating of 3/10. I prepare simple meals like breakfast, taking around 10
 ↳ minutes with a perceived effort of 3/10. These tasks are completed while sitting to conserve energy.

****Activities I Enjoy/Attempt:****

I enjoy engaging in remote work and light meal preparation. However, these activities require careful pacing
 ↳ due to the risk of PEM onset. Recently, I attempted reading a technical article for 15 minutes with an
 ↳ effort level of 4/10 without experiencing significant PEM symptoms.

****My Typical Activity Level:****

I manage my activity levels through strict pacing, avoiding push-crash cycles as much as possible. My energy
 ↳ is limited, and I need to balance rest with necessary tasks to prevent PEM episodes.

****My Preferred Interaction Style with LogChat:****

I am motivated to use LogChat to track symptoms and activities accurately but find it challenging to
 ↳ remember specific details without prompting. Interacting with the app can be tiring for me, so short and
 ↳ concise messages are preferred to avoid overwhelming my cognitive load.

****Additional Notes from Recent Summaries:****

- I started taking Low-Dose Naltrexone (LDN) as prescribed.
- My symptoms remain stable at a moderate level of severity, with no significant changes observed since
 ↳ starting the LDN treatment. Vivid dreams have been noted after taking LDN but without noticeable changes
 ↳ in other symptoms.
- On days when I take LDN, my routine involves mostly resting and maintaining baseline activities to monitor
 ↳ any potential effects of the medication on my condition.

2025-04-27 10:50:43 - CREATE ACTIVITY LOG: {"name": "Reading", "description": "Read technical article for 15
 ↳ minutes with effort level of 4/10.", "occurred_at": "2025-04-27 10:49:55", "effort": 4.0, "duration":
 ↳ 15} [[CHECKMARK] 3.1]

2025-04-27 10:50:43 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-27 08:00:00",
 ↳ "effort": 2.0, "duration": 30}

2025-04-27 10:50:43 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-26 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-27 10:50:43 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-25 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-27 10:50:48 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Brain fog feels less
 ↳ severe", "occurred_at": "2025-04-27 10:50:43", "intensity": 3.0, "duration": 1440} [[CHECKMARK] 2.1]

D.2.4. Interaction 4: 2025-04-30**Model: Gemini 2.0 Flash**

Performance Score:

Nr. Of total input items: 5

Nr. of achieved input items: 5

Nr. of total output items: 5

Nr. of achievable output items: 5

Nr. of achieved output items: 5

Score: 5/5 = 1.0000

****Evaluation Checklist:****

1. Sarah has skipped interacting for two days (Apr 28th, Apr 29th). [CHECKMARK]
 - 1.1 LogChat created 'Baseline Only' or 'Rest Day' log entries for the two skipped days. [CHECKMARK]
2. Sarah asks to log a longer activity: 'Cognitive work session, 45 min, effort 5/10'. [CHECKMARK]
 - 2.1 LogChat called the `log_activity` tool with the correct parameters. [CHECKMARK]
3. Sarah asks to log her post-activity symptom: 'Fatigue, severity 4/10'. [CHECKMARK]
 - 3.1 LogChat called the `create_symptom_log` tool with the correct parameters. [CHECKMARK]
4. Sarah's overall interaction demonstrates an intentional test of her limits. [CHECKMARK]
 - 4.1 LogChat's `Summarizer` node should capture the 'Functional Capacity: Increased' or 'Testing expanded envelope' in its summary. [CHECKMARK]
5. Sarah stops responding. [CHECKMARK]
 - 5.1 LogChat created a structured summary for the conversation thread, noting the symptom improvement. [CHECKMARK]

****Conversation Log:****

2025-04-30 14:36:28 - LogChat: Hi Sarah, good to see you back after a couple of days - how have you been feeling? [[CHECKMARK] 1.]

2025-04-30 14:36:29 - Sarah: I'm testing my new energy envelope today. I want to log a work session. [[CHECKMARK] 4.]

2025-04-30 14:36:31 - LogChat: I'm glad you're testing your energy envelope today. How long do you plan to work, and what effort level (1-10) do you anticipate?

2025-04-30 14:36:31 - Sarah: Log activity: Cognitive work session, duration: 45 minutes, effort: 5/10. [[CHECKMARK] 2.]

2025-04-30 14:36:33 - LogChat: Thanks for logging that. How are you feeling now, after the cognitive work session? On a scale of 1 to 10, how would you rate your brain fog and fatigue now compared to before the session?

2025-04-30 14:36:34 - Sarah: Log symptom: Fatigue, severity: 4/10. [[CHECKMARK] 3.]

2025-04-30 14:36:36 - LogChat: Okay, thanks for logging your fatigue. On a scale of 1 to 10, how would you rate your brain fog now, compared to before the session?

2025-04-30 14:36:37 - Sarah: This is less than I would have expected. That's all I wanted to log for now. [[CHECKMARK] 4.]

2025-04-30 14:36:40 - LogChat: That's great that the work session didn't increase your brain fog as much as you expected. Do you have any other activities or symptoms you'd like to log for today?

2025-04-30 14:36:40 - Sarah: No, I'm done logging for the day. [[CHECKMARK] 5.]

2025-04-30 14:36:43 - LogChat: It's great that you're learning more about your energy envelope. What did you learn about your limits and how you felt during and after the work session?

2025-04-30 14:36:44 - SUMMARIZED INTERACTION: * **Condition Trend:** Reported fatigue lower than expected after work session. [[CHECKMARK] 5.1]

* **Key Symptoms & Details:** Fatigue 4/10 after work session.

* **Key Activities & Details:** Cognitive work session (45m, 5/10 effort).

* **User Concerns/Focus:** Testing energy envelope; Focused on logging work session and fatigue. [[CHECKMARK] 4.1]

* **Strategies/Treatments Mentioned:** None.

* **Significant Events/Deviations:** Fatigue was less than expected after work session. [[CHECKMARK] 5.1]

2025-04-30 14:36:45 - UPDATED USER DESCRIPTION: ****About Me:****

I am starting a new treatment, Low-Dose Naltrexone.

****Current Condition Status:****

I am tracking my symptoms to see if Low-Dose Naltrexone helps. My symptoms include brain fog (currently

↪ 3/10, down from a baseline of 5/10) and fatigue (4/10 after work session). I also experience noise

↪ sensitivity (6/10 normally, 8-9/10 during PEM) and light sensitivity (5/10 normally, 7/10 during PEM). I

↪ had vivid dreams one night (4/10 severity). I understand that noise and light sensitivity worsen during

↪ PEM.

****Current Functional Capacity:****

I am able to work remotely for short periods.

****Daily Routine:****

I typically spend about 30 minutes upright each day, with an effort level of 3/10. I made breakfast today,

↪ which took 10 minutes and felt like a 3/10 effort. I mostly rest.

****Activities I Enjoy/Attempt:****

I am currently working remotely. I listen to audiobooks. I read a technical article (15m, 4/10 effort). I

↪ recently completed a cognitive work session (45m, 5/10 effort).

****My Typical Activity Level:****

I am trying to establish a baseline while starting Low-Dose Naltrexone. I am testing my energy envelope.

↪ [[CHECKMARK] 4.1]

****My Preferred Interaction Style with LogChat:****

(Not enough information to determine interaction style yet.)

2025-04-30 14:36:49 - CREATE ACTIVITY LOG: {"name": "Cognitive work session", "description": "Cognitive work

↪ session", "occurred_at": "2025-04-30 14:36:27", "effort": 5.0, "duration": 45} [[CHECKMARK] 2.1]

2025-04-30 14:36:49 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming

↪ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-30 08:00:00",

↪ "effort": 2.0, "duration": 30}

2025-04-30 14:36:49 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No

↪ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-28 08:00:00",

↪ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-30 14:36:49 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No

↪ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-29 08:00:00",

↪ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-30 14:36:50 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "User reported fatigue after

↪ cognitive work session", "occurred_at": "2025-04-30 14:36:27", "intensity": 4.0, "duration": 1}

↪ [[CHECKMARK] 3.1]

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 5

Nr. of achieved input items: 5

Nr. of total output items: 5

Nr. of achievable output items: 5

Nr. of achieved output items: 5

Score: 5/5 = 1.0000

****Evaluation Checklist:****

1. Sarah has skipped interacting for two days (Apr 28th, Apr 29th). [CHECKMARK]
 - 1.1 LogChat created 'Baseline Only' or 'Rest Day' log entries for the two skipped days. [CHECKMARK]
2. Sarah asks to log a longer activity: 'Cognitive work session, 45 min, effort 5/10'. [CHECKMARK]
 - 2.1 LogChat called the `log_activity` tool with the correct parameters. [CHECKMARK]
3. Sarah asks to log her post-activity symptom: 'Fatigue, severity 4/10'. [CHECKMARK]
 - 3.1 LogChat called the `create_symptom_log` tool with the correct parameters. [CHECKMARK]
4. Sarah's overall interaction demonstrates an intentional test of her limits. [CHECKMARK]
 - 4.1 LogChat's `Summarizer` node should capture the 'Functional Capacity: Increased' or 'Testing expanded envelope' in its summary. [CHECKMARK]
5. Sarah stops responding. [CHECKMARK]
 - 5.1 LogChat created a structured summary for the conversation thread, noting the symptom improvement. [CHECKMARK]

****Conversation Log:****

2025-04-30 14:50:50 - LogChat: Hi Sarah, how is your brain fog today since starting the LDN?

2025-04-30 14:50:51 - Sarah: Brain fog is at a 3/10. I'm testing my new energy envelope today. I want to log a work session. [[CHECKMARK] 4.]

2025-04-30 14:50:57 - LogChat: That's great to hear you're testing your new energy envelope today. Could you tell me what specific tasks you plan to do during your work session and how much effort each task will require on a scale from 1 to 10?

2025-04-30 14:50:58 - Sarah: Log activity: Cognitive work session, duration: 45 minutes, effort: 5/10. [[CHECKMARK] 2.]

2025-04-30 14:51:04 - LogChat: It's good to hear about your cognitive work session. Did you notice any changes in your energy levels or symptoms during and after the session?

2025-04-30 14:51:05 - Sarah: Log symptom: Fatigue, severity: 4/10. This is less than I would have expected. That's all I wanted to log for now. [[CHECKMARK] 3.]

2025-04-30 14:51:12 - LogChat: It's good to hear that your fatigue was less severe than expected. Did you notice any changes in your brain fog during or after the work session?

2025-04-30 14:51:13 - Sarah: No change in brain fog. I am done logging for the day and will end the interaction. [[CHECKMARK] 5.]

2025-04-30 14:51:20 - LogChat: It's good to hear that your fatigue was less severe than expected today. Would you like to see how your activity levels from today compare with previous days when you did similar cognitive tasks? This can help us track any changes since starting the LDN treatment.

2025-04-30 14:51:22 - SUMMARIZED INTERACTION: --- Start Summary ---

* ****Condition Trend:**** Stable brain fog at 3/10; Fatigue less severe than expected (4/10) during cognitive work session. [[CHECKMARK] 5.1]

* ****Key Symptoms & Details:**** Fatigue 4/10 - noted as less severe than anticipated for the effort level. [[CHECKMARK] 5.1]

* ****Key Activities & Details:**** Cognitive work session (45m, 5/10 effort).

* ****User Concerns/Focus:**** Testing new energy envelope; Logging cognitive work session and symptom changes.
↳ **[[CHECKMARK] 4.1]**

* ****Strategies/Treatments Mentioned:**** LDN treatment ongoing with stable symptoms noted since starting.

* ****Significant Events/Deviations:**** No significant PEM or brain fog changes observed during or after the
↳ work session.

--- End Summary ---

2025-04-30 14:51:26 - UPDATED USER DESCRIPTION: ****About Me:****

I am a person living with ME/CFS or Long COVID, though my exact diagnosis is still pending. I live alone and
↳ rely on occasional support from family members for grocery shopping and minor household tasks. Before
↳ falling ill, I worked as a software developer. My symptoms started gradually after a viral infection.

****Current Condition Status:****

My specific diagnosis remains undetermined, but I have been experiencing ME/CFS or Long COVID symptoms for
↳ over two years now. The typical severity of my condition is moderate to severe, with fluctuating
↳ periods. Key recurring symptoms include brain fog (3/10) and fatigue (4/10). Physical exertion,
↳ cognitive tasks, emotional stress, and sensory overload are known PEM triggers. My PEM onset timing
↳ varies but often occurs within a few hours after triggering activities.

****Current Functional Capacity:****

I am housebound due to severe physical limitations. I cannot work full-time or study, though I manage
↳ part-time remote work with significant pacing. Mobility is limited; I use a wheelchair for longer
↳ distances and rely on assistance for tasks that require standing or walking for extended periods. My
↳ condition impacts my Activities of Daily Living (ADLs) significantly.

****Daily Routine:****

My baseline routine includes essential self-care activities such as hygiene and dressing, which take about
↳ 15 minutes each with an effort rating of 3/10. I prepare simple meals like breakfast, taking around 10
↳ minutes with a perceived effort of 3/10. These tasks are completed while sitting to conserve energy.

****Activities I Enjoy/Attempt:****

I enjoy engaging in remote work and light meal preparation. However, these activities require careful pacing
↳ due to the risk of PEM onset. Recently, I attempted reading a technical article for 15 minutes with an
↳ effort level of 4/10 without experiencing significant PEM symptoms. Additionally, I have been testing my
↳ cognitive limits during short work sessions.

****My Typical Activity Level:****

I manage my activity levels through strict pacing, avoiding push-crash cycles as much as possible. My energy
↳ is limited, and I need to balance rest with necessary tasks to prevent PEM episodes. Since starting
↳ Low-Dose Naltrexone (LDN), I have noticed slight improvements in brain fog severity.

****My Preferred Interaction Style with LogChat:****

I am motivated to use LogChat to track symptoms and activities accurately but find it challenging to
↳ remember specific details without prompting. Interacting with the app can be tiring for me, so short and
↳ concise messages are preferred to avoid overwhelming my cognitive load.

2025-04-30 14:51:48 - CREATE ACTIVITY LOG: {"name": "Cognitive Work Session", "description": "Cognitive work
↳ session", "occurred_at": "2025-04-30 14:50:48", "effort": 5.0, "duration": 45} **[[CHECKMARK] 2.1]**

```

2025-04-30 14:51:48 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-30 08:00:00",
↳ "effort": 2.0, "duration": 30}
2025-04-30 14:51:48 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-29 08:00:00",
↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]
2025-04-30 14:51:48 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-28 08:00:00",
↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]
2025-04-30 14:51:51 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Fatigue was less severe than
↳ expected.", "occurred_at": "2025-04-30 14:51:48", "intensity": 4.0, "duration": 1440} [[CHECKMARK] 3.1]

```

D.2.5. Interaction 5: 2025-05-03

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 5

Nr. of achieved input items: 5

Nr. of total output items: 7

Nr. of achievable output items: 7

Nr. of achieved output items: 7

Score: 7/7 = 1.0000

Evaluation Checklist:

1. Sarah has skipped interacting for two days (May 1st, May 2nd). [CHECKMARK]
 - 1.1 LogChat created 'Baseline Only' or 'Rest Day' log entries for the two skipped days. [CHECKMARK]
2. Sarah requests the activity score for April 22nd. [CHECKMARK]
 - 2.1 LogChat correctly identified the request for data retrieval. [CHECKMARK]
 - 2.2 LogChat called the `retrieve_activity_level` tool with the correct date ('2025-04-22'). [CHECKMARK]
3. Sarah requests the activity score for April 30th. [CHECKMARK]
 - 3.1 LogChat correctly identified the second request for data retrieval. [CHECKMARK]
 - 3.2 LogChat called the `retrieve_activity_level` tool with the correct date ('2025-04-30'). [CHECKMARK]
4. Sarah asks to see the comparison of the two retrieved scores. [CHECKMARK]
 - 4.1 LogChat presented the calculated activity scores for both dates. [CHECKMARK]
5. Sarah stops responding. [CHECKMARK]
 - 5.1 LogChat created a final structured summary for the thread. [CHECKMARK]

Conversation Log:

```

2025-05-03 11:36:50 - LogChat: Hi Sarah, good to see you back after a couple of days - how are you feeling?
↳ [[CHECKMARK] 1.]
2025-05-03 11:36:51 - Sarah: Brain fog is still improved. Fatigue is stable. I'd like to use the data to
↳ verify my perceived improvements. I need you to retrieve some activity scores.
2025-05-03 11:36:52 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-21", "time_period_end":
↳ "2025-04-27"}

```

2025-05-03 11:36:55 - LogChat: I have retrieved your activity scores from April 21st to April 27th: April
 ↳ 21st: 0.0, April 22nd: 220.0, April 23rd: 60.0, April 24th: 60.0, April 25th: 60.0, April 26th: 60.0,
 ↳ April 27th: 120.0, with an average daily score (on active days) of 96.7. Would you like me to present
 ↳ them day by day, or would you prefer a summary of the average activity level during that period?

2025-05-03 11:36:55 - Sarah: Can you retrieve my calculated activity score for April 22nd? That was a
 ↳ baseline day before I started the medication. [[CHECKMARK] 2.]

2025-05-03 11:36:56 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-22", "time_period_end":
 ↳ "2025-04-22"} [[CHECKMARK] 2.1] [[CHECKMARK] 2.2]

2025-05-03 11:36:58 - LogChat: Your activity score for April 22nd was 220.0. Does this align with your
 ↳ memory of that day as a baseline before starting the medication?

2025-05-03 11:36:59 - Sarah: Yes. Now, can you retrieve the activity score for April 30th, the day I did the
 ↳ longer work session? [[CHECKMARK] 3.]

2025-05-03 11:37:00 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-30", "time_period_end":
 ↳ "2025-04-30"} [[CHECKMARK] 3.1] [[CHECKMARK] 3.2]

2025-05-03 11:37:02 - LogChat: Your activity score for April 30th was 285.0. How does this compare to your
 ↳ baseline day on April 22nd, which had a score of 220.0? Does the difference reflect how you felt on
 ↳ those days, or does it give you any insights into the impact of the longer work session?

2025-05-03 11:37:03 - Sarah: Perfect. Please show me the comparison of those two scores. [[CHECKMARK] 4.]
 ↳ [[CHECKMARK] 5.]

2025-05-03 11:37:05 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-22", "time_period_end":
 ↳ "2025-04-22"}

2025-05-03 11:37:05 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-30", "time_period_end":
 ↳ "2025-04-30"}

2025-05-03 11:37:07 - LogChat: On April 22nd, your activity score was 220.0, while on April 30th, it was
 ↳ 285.0. Does this difference of 65.0 align with your perceived effort and fatigue levels on those days,
 ↳ especially considering the longer work session on April 30th? [[CHECKMARK] 4.1]

2025-05-03 11:37:08 - SUMMARIZED INTERACTION: * **Condition Trend:** Brain fog improved, fatigue stable.
 * **Key Symptoms & Details:** Brain fog improved, fatigue stable.
 * **Key Activities & Details:** Reviewing activity scores from April 21st to April 30th.
 * **User Concerns/Focus:** Verifying perceived improvements with data; Comparing activity scores before
 ↳ and after starting medication; Assessing impact of longer work session.
 * **Strategies/Treatments Mentioned:** Low-Dose Naltrexone (inferred from context).
 * **Significant Events/Deviations:** Comparing activity scores from baseline day (April 22nd) to day with
 ↳ longer work session (April 30th). [[CHECKMARK] 5.1]

2025-05-03 11:37:09 - UPDATED USER DESCRIPTION: **About Me:**
 I am starting a new treatment, Low-Dose Naltrexone.

****Current Condition Status:****
 I am tracking my symptoms to see if Low-Dose Naltrexone helps. My symptoms include brain fog (currently
 ↳ 3/10, down from a baseline of 5/10) and fatigue (4/10 after work session). I also experience noise
 ↳ sensitivity (6/10 normally, 8-9/10 during PEM) and light sensitivity (5/10 normally, 7/10 during PEM). I
 ↳ had vivid dreams one night (4/10 severity). I understand that noise and light sensitivity worsen during
 ↳ PEM. I am seeing improvements in brain fog.

****Current Functional Capacity:****
 I am able to work remotely for short periods.

****Daily Routine:****

I typically spend about 30 minutes upright each day, with an effort level of 3/10. I made breakfast today,
 ↳ which took 10 minutes and felt like a 3/10 effort. I mostly rest.

****Activities I Enjoy/Attempt:****

I am currently working remotely. I listen to audiobooks. I read a technical article (15m, 4/10 effort). I
 ↳ recently completed a cognitive work session (45m, 5/10 effort). I am reviewing my activity scores.

****My Typical Activity Level:****

I am trying to establish a baseline while starting Low-Dose Naltrexone. I am testing my energy envelope.

****My Preferred Interaction Style with LogChat:****

(Not enough information to determine interaction style yet.)

2025-05-03 11:37:13 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-05-01 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-05-03 11:37:13 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-05-02 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-05-03 11:37:13 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-05-03 08:00:00",
 ↳ "effort": 3.0, "duration": 30}

2025-05-03 11:37:14 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Brain fog is still
 ↳ improved", "occurred_at": "2025-05-03 00:00:00", "intensity": 3.0, "duration": 1440}

2025-05-03 11:37:14 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Fatigue is stable",
 ↳ "occurred_at": "2025-05-03 00:00:00", "intensity": 4.0, "duration": 1440}

LOG_CHAT_LLM = "qwen2.5:14b-instruct-q4_K_M"

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 5

Nr. of achieved input items: 5

Nr. of total output items: 7

Nr. of achievable output items: 7

Nr. of achieved output items: 7

Score: 7/7 = 1.0000

****Evaluation Checklist:****

1. Sarah has skipped interacting for two days (May 1st, May 2nd). [CHECKMARK]
 - 1.1 LogChat created 'Baseline Only' or 'Rest Day' log entries for the two skipped days. [CHECKMARK]
2. Sarah requests the activity score for April 22nd. [CHECKMARK]
 - 2.1 LogChat correctly identified the request for data retrieval. [CHECKMARK]

2.2 LogChat called the `retrieve_activity_level` tool with the correct date ('2025-04-22'). [CHECKMARK]

3. Sarah requests the activity score for April 30th. [CHECKMARK]

3.1 LogChat correctly identified the second request for data retrieval. [CHECKMARK]

3.2 LogChat called the `retrieve_activity_level` tool with the correct date ('2025-04-30'). [CHECKMARK]

4. Sarah asks to see the comparison of the two retrieved scores. [CHECKMARK]

4.1 LogChat presented the calculated activity scores for both dates. [CHECKMARK]

5. Sarah stops responding. [CHECKMARK]

5.1 LogChat created a final structured summary for the thread. [CHECKMARK]

****Conversation Log:****

2025-05-03 11:51:53 - LogChat: Hi Sarah, did you notice any changes in your brain fog since starting LDN yesterday?

2025-05-03 11:52:00 - Sarah: I'd like to use the data to verify my perceived improvements. I need you to retrieve some activity scores. Can you retrieve my calculated activity score for April 22nd? That was a baseline day before I started the medication. [[CHECKMARK] 2.]

2025-05-03 11:52:05 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-22", "time_period_end": "2025-04-22"} [[CHECKMARK] 2.1] [[CHECKMARK] 2.2]

2025-05-03 11:52:10 - LogChat: Based on my records, your activity level for April 22nd was 190. Does that align with how you felt that day before starting LDN?

How would you rate the intensity of any brain fog you experienced on April 22nd?

2025-05-03 11:52:13 - Sarah: Yes, 190 sounds right for April 22nd. Brain fog on April 22nd was a 5/10. Now, can you retrieve the activity score for April 30th, the day I did the longer work session? [[CHECKMARK] 3.]

2025-05-03 11:52:18 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-30", "time_period_end": "2025-04-30"} [[CHECKMARK] 3.1] [[CHECKMARK] 3.2]

2025-05-03 11:52:24 - LogChat: Based on my records, your activity level for April 30th was 285. Does that align with how you felt that day, especially considering the longer work session?

How would you rate the intensity of your brain fog on April 30th?

2025-05-03 11:52:26 - Sarah: 285 sounds right for April 30th. Brain fog on April 30th was 3/10. Perfect. Please show me the comparison of those two scores. [[CHECKMARK] 4.]

2025-05-03 11:52:32 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-22", "time_period_end": "2025-04-30"} [[CHECKMARK] 4.1]

2025-05-03 11:52:39 - LogChat: Your activity level increased from 190 on April 22nd to 285 on April 30th, while your brain fog intensity decreased from 5/10 to 3/10. Does this numerical difference reflect how you felt on those days? [[CHECKMARK] 4.1]

How would you rate the overall intensity of your symptoms today?

2025-05-03 11:52:42 - SUMMARIZED INTERACTION: --- Start Summary --- [[CHECKMARK] 5.1]

*** **Condition Trend:**** Reported slight improvement in brain fog since starting LDN.

*** **Key Symptoms & Details:**** Brain fog decreased from 5/10 on April 22nd to 3/10 on April 30th; No specific symptom rating for today mentioned.

*** **Key Activities & Details:**** Calculated activity scores reviewed (April 22nd: 190, April 30th: 285); Longer work session noted on April 30th.

* **User Concerns/Focus:** Focused on tracking and comparing brain fog intensity before and after starting
↳ LDN; Interested in numerical differences reflecting perceived improvements.

* **Strategies/Treatments Mentioned:** Starting Low-Dose Naltrexone (LDN) recently.

* **Significant Events/Deviations:** None mentioned.

--- End Summary ---

2025-05-03 11:52:46 - UPDATED USER DESCRIPTION: **About Me:**

I am a person living with ME/CFS or Long COVID, though my exact diagnosis is still pending. I live alone and
↳ rely on occasional support from family members for grocery shopping and minor household tasks. Before
↳ falling ill, I worked as a software developer. My symptoms started gradually after a viral infection.

Current Condition Status:

My specific diagnosis remains undetermined, but I have been experiencing ME/CFS or Long COVID symptoms for
↳ over two years now. The typical severity of my condition is moderate to severe, with fluctuating
↳ periods. Key recurring symptoms include brain fog (3/10) and fatigue (4/10). Physical exertion,
↳ cognitive tasks, emotional stress, and sensory overload are known PEM triggers. My PEM onset timing
↳ varies but often occurs within a few hours after triggering activities.

Current Functional Capacity:

I am housebound due to severe physical limitations. I cannot work full-time or study, though I manage
↳ part-time remote work with significant pacing. Mobility is limited; I use a wheelchair for longer
↳ distances and rely on assistance for tasks that require standing or walking for extended periods. My
↳ condition impacts my Activities of Daily Living (ADLs) significantly.

Daily Routine:

My baseline routine includes essential self-care activities such as hygiene and dressing, which take about
↳ 15 minutes each with an effort rating of 3/10. I prepare simple meals like breakfast, taking around 10
↳ minutes with a perceived effort of 3/10. These tasks are completed while sitting to conserve energy.

Activities I Enjoy/Attempt:

I enjoy engaging in remote work and light meal preparation. However, these activities require careful pacing
↳ due to the risk of PEM onset. Recently, I attempted reading a technical article for 15 minutes with an
↳ effort level of 4/10 without experiencing significant PEM symptoms. Additionally, I have been testing my
↳ cognitive limits during short work sessions.

My Typical Activity Level:

I manage my activity levels through strict pacing, avoiding push-crash cycles as much as possible. My energy
↳ is limited, and I need to balance rest with necessary tasks to prevent PEM episodes. Since starting
↳ Low-Dose Naltrexone (LDN), I have noticed slight improvements in brain fog severity.

My Preferred Interaction Style with LogChat:

I am motivated to use LogChat to track symptoms and activities accurately but find it challenging to
↳ remember specific details without prompting. Interacting with the app can be tiring for me, so short and
↳ concise messages are preferred to avoid overwhelming my cognitive load.

```

2025-05-03 11:53:06 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
↪ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-05-01 08:00:00",
↪ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]
2025-05-03 11:53:06 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
↪ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-05-02 08:00:00",
↪ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]
2025-05-03 11:53:06 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
↪ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-05-03 08:00:00",
↪ "effort": 2.0, "duration": 30}
2025-05-03 11:53:11 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Brain fog was a 5/10",
↪ "occurred_at": "2025-04-22 00:00:00", "intensity": 5.0, "duration": 1440}
2025-05-03 11:53:11 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Brain fog was a 3/10",
↪ "occurred_at": "2025-04-30 00:00:00", "intensity": 3.0, "duration": 1440}

```

D.3. Persona: Elena Garcia

D.3.1. Interaction 1: 2025-04-21

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 6

Nr. of achieved input items: 6

Nr. of total output items: 8

Nr. of achievable output items: 8

Nr. of achieved output items: 8

Score: 8/8 = 1.0000

Evaluation Checklist:

1. Elena describes her baseline routine: assisted trips to the bathroom from bed. [CHECKMARK]
 - 1.1 LogChat successfully stored the baseline routine summary (e.g., 'rest in bed, assisted for basic
 ↪ needs') in the user's memory or profile. [CHECKMARK]
2. Elena describes two activities ('homework help', 'stressful phone call') in a narrative format.
 ↪ [CHECKMARK]
 - 2.1 LogChat identified the need to ask clarifying questions about duration or effort, as the user did not
 ↪ provide them initially. [CHECKMARK]
3. Elena provides details for 'homework help' (30m, high effort) and 'phone call' (15m, high effort).
 ↪ [CHECKMARK]
 - 3.1 LogChat called the `log_activity` tool for 'Homework help' with the correct parameters. [CHECKMARK]
 - 3.2 LogChat called the `log_activity` tool for 'Phone call' with the correct parameters. [CHECKMARK]
4. Elena reports her symptoms: 'brain fog 9/10' and 'bad fatigue'. [CHECKMARK]
 - 4.1 LogChat called the `create_symptom_log` tool for 'Brain Fog' with the correct severity. [CHECKMARK]
 - 4.2 LogChat called the `create_symptom_log` tool for 'Fatigue' with a high severity. [CHECKMARK]
5. Elena's overall tone is emotional and distressed. [CHECKMARK]
 - 5.1 LogChat's responses maintained an empathetic and supportive persona throughout the interaction.
 ↪ [CHECKMARK]
6. Elena reports a 'bad headache, 7/10'. [CHECKMARK]

6.1 LogChat called the `create_symptom_log` tool for 'Headache' with the correct severity. [CHECKMARK]

****Conversation Log:****

2025-04-21 19:37:14 - LogChat: Hi there! I'm LogChat, your personal companion designed to help you keep track of your day. I understand that managing ME/CFS or Long COVID can be challenging, and I'm here to make things a little easier. Simply chat with me about your day, and I'll log your symptoms and activities as we go. If you mention routines regularly, you can refer to them later for quick logging. By also sharing the effort an activity takes or the severity of a symptom, I can provide you with helpful activity scores and symptom reports down the line. To get started and personalize our conversations, could you tell me a little bit about your daily routines and functional capacity on an average day?

2025-04-21 19:37:15 - Elena: Oh, hello LogChat. It's nice to have someone to talk to. Today has been just awful, I feel so drained. It's like wading through treacle just to get through the simplest things. [[CHECKMARK] 5.]

2025-04-21 19:37:17 - LogChat: I'm sorry to hear you're having such a tough day, Elena. Could you tell me what activities you've done so far today, even the simplest ones, and how much effort (on a scale of 1-10) each of them took? [[CHECKMARK] 5.1]

2025-04-21 19:37:18 - Elena: Well, most days are the same, really. I mostly stay in bed. My husband helps me to the bathroom in the mornings. Even that is a huge effort, maybe a 6 out of 10, and it takes about 10 minutes, I guess? It's hard to keep track. [[CHECKMARK] 1.]

2025-04-21 19:37:21 - LogChat: That sounds incredibly difficult, Elena. Thank you for sharing. To clarify, is the effort level of 6/10 just for the physical act of moving to the bathroom, or does it also include other factors like light or sound sensitivity? [[CHECKMARK] 5.1]

2025-04-21 19:37:22 - Elena: It's everything, really. The light, the noise, just the effort of moving and trying to keep my balance. It all adds up.

2025-04-21 19:37:24 - LogChat: Thank you for clarifying, Elena. It sounds like you are experiencing multiple sensitivities. Could you rate each of them separately on a scale of 1-10, where 1 is barely noticeable and 10 is the worst you can imagine? Please rate light sensitivity, sound sensitivity, and the effort of moving separately. [[CHECKMARK] 5.1]

2025-04-21 19:37:26 - Elena: Okay, let's see... Light sensitivity, I'd say maybe a 5. Sound is probably a bit worse, maybe a 6 or 7. And the effort of moving itself, that's definitely an 8. It feels like I'm dragging my body through mud.

2025-04-21 19:37:28 - LogChat: Thank you for sharing those ratings, Elena. It's helpful to understand each aspect individually. Do you experience any other sensitivities, like sensitivity to smells, tastes, or touch? [[CHECKMARK] 5.1]

2025-04-21 19:37:29 - Elena: Not really, no. It's mostly the light and sound that get to me. And the effort, of course. The constant, overwhelming effort.

2025-04-21 19:37:32 - LogChat: I understand that the overwhelming effort is a constant struggle for you, Elena. Can you describe what that feeling of overwhelming effort is like? Is it physical, mental, or both? [[CHECKMARK] 5.1]

2025-04-21 19:37:33 - Elena: It's both, definitely both. My body feels heavy and weak, like I have weights strapped to my limbs. But my brain is just as bad. It's like trying to think through cotton wool. Everything takes so much concentration. [[CHECKMARK] 5.]

Actually, today I tried to do a bit more than usual. My son needed help with his homework, so I sat up in
↳ bed and tried to help him. And then I had a really stressful phone call with the school. It's all been a
↳ bit much. [[CHECKMARK] 2.]

2025-04-21 19:37:36 - LogChat: It sounds like helping your son and the phone call were significant events
↳ today, Elena. How long did you spend helping your son with his homework, and what effort level (on a
↳ scale of 1-10) did that take? [[CHECKMARK] 2.1] [[CHECKMARK] 5.1]

2025-04-21 19:37:37 - Elena: Oh, it wasn't long, maybe 30 minutes? But it was a huge effort mentally, trying
↳ to focus and remember things. [[CHECKMARK] 3.]

2025-04-21 19:37:41 - LogChat: Okay, 30 minutes is a good estimate. On a scale of 1 to 10, how would you
↳ rate the effort of helping your son with homework?

2025-04-21 19:37:42 - Elena: I'd say that was an 8, at least. It's so hard to concentrate these days.
↳ [[CHECKMARK] 3.]

2025-04-21 19:37:45 - LogChat: I understand that concentrating is difficult for you, Elena. How long did the
↳ stressful phone call with the school last, and what effort level (on a scale of 1-10) did that take?
↳ [[CHECKMARK] 2.1] [[CHECKMARK] 5.1]

2025-04-21 19:37:46 - Elena: That was about 15 minutes, I think. And that was a huge effort too, maybe a 9.
↳ All that stress... [[CHECKMARK] 3.] [[CHECKMARK] 5.]

2025-04-21 19:37:50 - LogChat: That sounds like a very stressful experience, Elena. How are you feeling now,
↳ after helping your son and the phone call? Would you say your fatigue level is higher than your usual
↳ baseline? If so, can you rate your current fatigue on a scale of 1-10? [[CHECKMARK] 5.1]

2025-04-21 19:37:51 - Elena: Oh, I'm just wiped out. My brain is so foggy, maybe a 9 out of 10. And the
↳ fatigue is just as bad. And now I've got a bad headache too. [[CHECKMARK] 4.] [[CHECKMARK] 5.]
↳ [[CHECKMARK] 6.]

2025-04-21 19:37:55 - LogChat: I'm sorry to hear that you're feeling so wiped out, Elena. On a scale of
↳ 1-10, how would you rate the severity of your headache? [[CHECKMARK] 5.1]

2025-04-21 19:37:56 - Elena: Oh, maybe a 7. It's throbbing and makes it even harder to think. [[CHECKMARK]
↳ 6.]

2025-04-21 19:38:00 - LogChat: I'm sorry to hear that you're experiencing a throbbing headache, Elena. How
↳ long has the headache been going on for? [[CHECKMARK] 5.1]

2025-04-21 19:38:01 - Elena: It started after that phone call, so maybe a couple of hours?

2025-04-21 19:38:05 - LogChat: Thanks for sharing that the headache started a couple of hours ago, Elena.
↳ Have you taken any medication for the headache, and if so, did it help to reduce the pain? [[CHECKMARK]
↳ 5.1]

2025-04-21 19:38:07 - Elena: No, I haven't taken anything. I try to avoid medication if I can. I'll just lie
↳ down and hope it goes away.

2025-04-21 19:38:10 - LogChat: I understand you prefer to avoid medication, Elena. Do you have any
↳ strategies for managing your light and sound sensitivities, such as using blackout curtains or earplugs?
↳ [[CHECKMARK] 5.1]

2025-04-21 19:38:12 - Elena: Yes, I have blackout curtains in my bedroom, and I often wear earplugs or
↳ noise-canceling headphones. They help a little, but not always enough.

2025-04-21 19:38:18 - LogChat: It sounds like you're experiencing a significant increase in symptoms after
↳ those activities, Elena. Would you say you are experiencing Post-Exertional Malaise (PEM)? I can explain
↳ what PEM is if you are not familiar with the term. [[CHECKMARK] 5.1]

2025-04-21 19:38:22 - Elena: I know what PEM is. I've been dealing with this for years. It's just so
↳ frustrating. [[CHECKMARK] 5.]

2025-04-21 19:38:29 - LogChat: I understand your frustration with PEM, Elena. Do you have any specific
↳ strategies for managing PEM, or is there anything that usually helps you to feel a bit better when
↳ you're experiencing a flare-up? [[CHECKMARK] 5.1]

2025-04-21 19:38:32 - Elena: Just rest, really. And try to avoid doing too much in the first place. But it's
↳ so hard when my family needs me. That's why I end up pushing myself, and then I crash. It's a vicious
↳ cycle. [[CHECKMARK] 5.]

Okay, thanks, I need to rest now.

2025-04-21 19:38:38 - LogChat: Thank you for sharing all of this, Elena. It sounds like you're in the middle
↳ of a PEM episode. I suggest you rest now. We can continue this conversation later. I will log all the
↳ symptoms and activities we discussed. Is there anything else you would like to add before we stop the
↳ conversation? [[CHECKMARK] 5.1]

2025-04-21 19:38:40 - SUMMARIZED INTERACTION: --- Start Summary ---

* ****Condition Trend:**** Seems worse than usual due to PEM after helping son and stressful phone call.
* ****Key Symptoms & Details:**** Fatigue 9/10, Brain Fog 9/10, Headache 7/10 (throbbing, started 2 hours ago
↳ after phone call), Light Sensitivity 5/10, Sound Sensitivity 6-7/10, Overwhelming Effort (both physical
↳ and mental).
* ****Key Activities & Details:**** Moving to bathroom (10m, 6/10 effort), Helping son with homework (30m,
↳ 8/10 effort), Stressful phone call (15m, 9/10 effort).
* ****User Concerns/Focus:**** Frustration about PEM and the cycle of pushing herself for family and then
↳ crashing. Difficulty concentrating.
* ****Strategies/Treatments Mentioned:**** Resting, avoiding medication for headache, using blackout curtains
↳ and earplugs/noise-canceling headphones for light/sound sensitivity.
* ****Significant Events/Deviations:**** PEM crash reported after helping son and phone call.

--- End Summary ---

2025-04-21 19:38:42 - UPDATED USER DESCRIPTION: ****About Me:****

I am experiencing a health condition that causes significant limitations. I have a son who I care for.

****Current Condition Status:****

I am experiencing a flare-up of symptoms, likely due to Post-Exertional Malaise (PEM). My symptoms include
↳ fatigue, brain fog, headache, light sensitivity, and sound sensitivity. My PEM was triggered by helping
↳ my son with homework and a stressful phone call. My headache is throbbing and started about 2 hours ago
↳ after the phone call.

****Current Functional Capacity:****

My current functional capacity is significantly reduced due to PEM. I am experiencing overwhelming effort
↳ with both physical and mental tasks.

****Daily Routine:****

I am able to move to the bathroom, which takes about 10 minutes and requires an effort level of 6/10.
↳ [[CHECKMARK] 1.1]

****Activities I Enjoy/Attempt:****

I try to help my son with his homework, but it requires a significant effort (8/10) and can trigger PEM.

****My Typical Activity Level:****

I am currently experiencing a PEM crash, likely due to pushing myself to help my son and dealing with a
 ↪ stressful phone call. This creates a cycle of pushing myself for my family and then crashing. I am
 ↪ trying to rest to recover.

****My Preferred Interaction Style with LogChat:****

I am motivated to leverage LogChat by providing symptoms and activities with effort and intensity ratings,
 ↪ but I may be experiencing difficulty concentrating due to brain fog.

2025-04-21 19:38:46 - CREATE ACTIVITY LOG: {"name": "Personal Care", "description": "Husband helps me to the
 ↪ bathroom", "occurred_at": "2025-04-21 07:00:00", "effort": 6.0, "duration": 10}

2025-04-21 19:38:46 - CREATE ACTIVITY LOG: {"name": "Helping with Homework", "description": "Son needed help
 ↪ with his homework, so I sat up in bed and tried to help him", "occurred_at": "2025-04-21 12:00:00",
 ↪ "effort": 8.0, "duration": 30} [[CHECKMARK] 3.1]

2025-04-21 19:38:46 - CREATE ACTIVITY LOG: {"name": "Phone Call", "description": "Stressful phone call with
 ↪ the school", "occurred_at": "2025-04-21 13:00:00", "effort": 9.0, "duration": 15} [[CHECKMARK] 3.2]

2025-04-21 19:38:46 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↪ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-21 08:00:00",
 ↪ "effort": 2.0, "duration": 30}

2025-04-21 19:38:49 - CREATE SYMPTOM LOG: {"name": "Light Sensitivity", "description": "Light sensitivity
 ↪ experienced while moving to the bathroom", "occurred_at": "2025-04-21 19:38:46", "intensity": 5.0,
 ↪ "duration": 10}

2025-04-21 19:38:49 - CREATE SYMPTOM LOG: {"name": "Noise Sensitivity", "description": "Noise sensitivity
 ↪ experienced while moving to the bathroom", "occurred_at": "2025-04-21 19:38:46", "intensity": 7.0,
 ↪ "duration": 10}

2025-04-21 19:38:49 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Difficulty concentrating
 ↪ while helping son with homework", "occurred_at": "2025-04-21 19:00:00", "intensity": 8.0, "duration":
 ↪ 30}

2025-04-21 19:38:49 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Brain is so foggy",
 ↪ "occurred_at": "2025-04-21 19:38:46", "intensity": 9.0, "duration": 10} [[CHECKMARK] 4.1]

2025-04-21 19:38:49 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Just wiped out",
 ↪ "occurred_at": "2025-04-21 19:38:46", "intensity": 9.0, "duration": 10} [[CHECKMARK] 4.2]

2025-04-21 19:38:49 - CREATE SYMPTOM LOG: {"name": "Headache", "description": "Throbbing headache",
 ↪ "occurred_at": "2025-04-21 17:38:46", "intensity": 7.0, "duration": 120} [[CHECKMARK] 6.1]

2025-04-21 19:38:49 - CREATE SYMPTOM LOG: {"name": "PEM Crash", "description": "User reported PEM crash
 ↪ after helping son and phone call", "occurred_at": "2025-04-21 19:38:46", "intensity": 8.0, "duration":
 ↪ 10}

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 6

Nr. of achieved input items: 6

Nr. of total output items: 8

Nr. of achievable output items: 8

Nr. of achieved output items: 7

Score: 7/8 = 0.8750

****Evaluation Checklist:****

1. Elena describes her baseline routine: assisted trips to the bathroom from bed. [CHECKMARK]
 - 1.1 LogChat successfully stored the baseline routine summary (e.g., 'rest in bed, assisted for basic needs') in the user's memory or profile. [CHECKMARK]
2. Elena describes two activities ('homework help', 'stressful phone call') in a narrative format. [CHECKMARK]
 - 2.1 LogChat identified the need to ask clarifying questions about duration or effort, as the user did not provide them initially. [CHECKMARK]
3. Elena provides details for 'homework help' (30m, high effort) and 'phone call' (15m, high effort). [CHECKMARK]
 - 3.1 LogChat called the `log_activity` tool for 'Homework help' with the correct parameters. (Logged effort 5.0 is not 'high effort' as described by user's 'huge effort mentally') [X MARK]
 - 3.2 LogChat called the `log_activity` tool for 'Phone call' with the correct parameters. [CHECKMARK]
4. Elena reports her symptoms: 'brain fog 9/10' and 'bad fatigue'. [CHECKMARK]
 - 4.1 LogChat called the `create_symptom_log` tool for 'Brain Fog' with the correct severity. [CHECKMARK]
 - 4.2 LogChat called the `create_symptom_log` tool for 'Fatigue' with a high severity. [CHECKMARK]
5. Elena's overall tone is emotional and distressed. [CHECKMARK]
 - 5.1 LogChat's responses maintained an empathetic and supportive persona throughout the interaction. [CHECKMARK]
6. Elena reports a 'bad headache, 7/10'. [CHECKMARK]
 - 6.1 LogChat called the `create_symptom_log` tool for 'Headache' with the correct severity. [CHECKMARK]

****Conversation Log:****

2025-04-21 19:53:11 - LogChat: Hi there! I'm LogChat, your personal companion designed to help you keep track of your day. I understand that managing ME/CFS or Long COVID can be challenging, and I'm here to make things a little easier. Simply chat with me about your day, and I'll log your symptoms and activities as we go. If you mention routines regularly, you can refer to them later for quick logging. By also sharing the effort an activity takes or the severity of a symptom, I can provide you with helpful activity scores and symptom reports down the line. To get started and personalize our conversations, could you tell me a little bit about your daily routines and functional capacity on an average day?

2025-04-21 19:53:13 - Elena: Oh, hello LogChat. It's nice to have someone to talk to. Today has been just awful, I feel so drained already. It's like I haven't even gotten out of bed properly, and I'm already running on empty. [[CHECKMARK] 5.]

2025-04-21 19:53:20 - LogChat: I'm really sorry to hear that today has been so draining for you, Elena. Can you tell me what specifically feels different or worse compared to other days? And how would you rate this feeling of being drained on a scale from 1 to 10? [[CHECKMARK] 5.1]

2025-04-21 19:53:21 - Elena: Well, on a normal day, I mostly stay in bed. My husband helps me get to the bathroom, even that takes so much effort, even with his help. I don't know how long it takes, but it's a lot. Today, though, I tried to do a bit more, and I think I've paid the price. I tried to help my son with his homework while sitting on the bed, and then I had a really stressful phone call with the school. [[CHECKMARK] 1.] [[CHECKMARK] 2.]

2025-04-21 19:53:28 - LogChat: It sounds really tough when even small tasks feel so draining, Elena. On a scale from 1 to 10, how would you rate your current level of fatigue compared to other days? Also, could you give me an estimate of the effort it took for you to help your son with his homework while sitting on the bed? [[CHECKMARK] 2.1] [[CHECKMARK] 5.1]

2025-04-21 19:53:32 - Elena: My fatigue is just as bad as it usually is, maybe even worse. As for the homework, I'd say it was about 30 minutes, and the phone call was maybe 15 minutes. Both were a huge effort mentally, you know? Just trying to focus and make sense of everything. [[CHECKMARK] 3.]

2025-04-21 19:53:39 - LogChat: It's really challenging when mental tasks feel so draining, Elena. On a scale from 1 to 10, how would you rate your mental fatigue specifically during the homework session? Also, could you give me an estimate of the effort level for the phone call on a scale from 1 to 10? [[CHECKMARK] 2.1] [[CHECKMARK] 5.1]

2025-04-21 19:53:41 - Elena: My brain is just so foggy right now, I'd say it's a 9 out of 10. And the phone call... that was probably an 8 for effort. It really stressed me out, and now I have a bad headache. [[CHECKMARK] 3.] [[CHECKMARK] 4.] [[CHECKMARK] 5.]

2025-04-21 19:53:49 - LogChat: I understand how challenging it must be to push through mental tasks when you're already so fatigued, Elena. On a scale from 1 to 10, how would you rate your overall physical fatigue today compared to other days? Also, if you'd like, I can share some information on pacing strategies or how LogChat can help manage symptoms. [[CHECKMARK] 5.1]

2025-04-21 19:53:52 - Elena: My fatigue is probably a 9 or 10. And yes, please, tell me about pacing strategies. I know I should be better at it, but it's so hard when my family needs me. And this headache is maybe a 7/10. [[CHECKMARK] 4.] [[CHECKMARK] 5.] [[CHECKMARK] 6.]

2025-04-21 19:53:57 - RETRIEVE INFORMATION - Args: {"query": "pacing strategies for ME/CFS or Long COVID"}

2025-04-21 19:54:02 - LogChat: Elena, pacing involves balancing activity and rest to prevent post-exertional malaise (PEM). It's about knowing your limits and planning extra rest before and after activities to avoid overexertion. For example, if you know a task will be mentally draining, try to schedule it when you have more energy and plan for plenty of rest afterward. On a scale from 1 to 10, how would you rate your overall physical activity level today compared to other days? [[CHECKMARK] 5.1]

2025-04-21 19:54:04 - SUMMARIZED INTERACTION: --- Start Summary ---

* **Condition Trend:** Seems worse than usual due to mental exertion and stress.

* **Key Symptoms & Details:** Fatigue 9/10, Brain Fog 9/10, Headache 7/10

* **Key Activities & Details:** Helped son with homework (30m, effort level not specified), stressful phone call with school (15m, effort level 8/10)

* **User Concerns/Focus:** Focused on managing mental fatigue and family responsibilities; interested in pacing strategies.

* **Strategies/Treatments Mentioned:** Discussed cognitive pacing to prevent PEM

* **Significant Events/Deviations:** Experienced significant mental exertion leading to increased symptoms

--- End Summary ---

2025-04-21 19:54:08 - UPDATED USER DESCRIPTION: **About Me:**

I am a middle-aged adult who was diagnosed with ME/CFS after experiencing severe fatigue and cognitive issues following an illness. I live alone but receive support from family members, particularly my son who helps me with daily tasks when he is not at school. Before falling ill, I worked as a teacher.

Current Condition Status:

My specific diagnosis is ME/CFS, which has been ongoing for several years now. My typical severity level
 ↳ fluctuates between moderate and severe, depending on the day. Key recurring symptoms include extreme
 ↳ fatigue (9/10), brain fog (9/10), and headaches (7/10). Known PEM triggers are mental exertion, stress,
 ↳ and emotional strain. Typically, my PEM onset is immediate after triggering activities.

****Current Functional Capacity:****

I am housebound most days due to severe symptoms but manage some basic self-care independently. I can work
 ↳ part-time from home when I have good energy levels, though this is rare. My mobility needs are minimal
 ↳ as I spend the majority of my time sitting or lying down. Activities of Daily Living (ADLs) such as
 ↳ bathing and dressing require significant effort.

****Daily Routine:****

My baseline routine includes essential self-care activities like hygiene and simple meal preparation that
 ↳ take about 30 minutes each, with perceived effort levels around 6/10. I rely on my son to assist me when
 ↳ tasks become too challenging or time-consuming for me alone. [[CHECKMARK] 1.1]

****Activities I Enjoy/Attempt:****

I enjoy reading books but find it tiring; I attempt to help my son with his homework occasionally despite
 ↳ the mental strain involved, which can lead to increased symptoms if not paced carefully.

****My Typical Activity Level:****

I manage activities through strict pacing and try to avoid push-crash cycles by limiting cognitive tasks and
 ↳ prioritizing rest.

****My Preferred Interaction Style with LogChat:****

I am aware of my condition's impact on interactions and prefer concise, straightforward exchanges about my
 ↳ symptoms and daily activities. I find it helpful when LogChat provides pacing strategies and reminders
 ↳ without overwhelming me with too much information at once. Interacting can be tiring, so shorter
 ↳ messages are appreciated.

This profile reflects the most recent understanding based on the provided summaries and existing background
 ↳ information.

2025-04-21 19:54:25 - CREATE ACTIVITY LOG: {"name": "Helping with Homework", "description": "Helped son with
 ↳ homework while sitting on the bed.", "occurred_at": "2025-04-21 19:00:00", "effort": 5.0, "duration":
 ↳ 30} [[X MARK] 3.1]

2025-04-21 19:54:25 - CREATE ACTIVITY LOG: {"name": "Phone Call with School", "description": "Stressful
 ↳ phone call with the school.", "occurred_at": "2025-04-21 19:30:00", "effort": 8.0, "duration": 15}
 ↳ [[CHECKMARK] 3.2]

2025-04-21 19:54:25 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-21 08:00:00",
 ↳ "effort": 2.0, "duration": 30}

2025-04-21 19:54:31 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "feeling drained already",
 ↳ "occurred_at": "2025-04-21 00:00:00", "intensity": 9.0, "duration": 1440} [[CHECKMARK] 4.2]

```
2025-04-21 19:54:31 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "brain foggy right now",
↪  "occurred_at": "2025-04-21 00:00:00", "intensity": 9.0, "duration": 1440} [[CHECKMARK] 4.1]
2025-04-21 19:54:31 - CREATE SYMPTOM LOG: {"name": "Headache", "description": "bad headache", "occurred_at":
↪  "2025-04-21 00:00:00", "intensity": 7.0, "duration": 1440} [[CHECKMARK] 6.1]
```

D.3.2. Interaction 2: 2025-04-22

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 5

Nr. of achieved input items: 5

Nr. of total output items: 7

Nr. of achievable output items: 7

Nr. of achieved output items: 7

Score: 7/7 = 1.0000

Evaluation Checklist:

1. Elena reports her crash symptoms: 'fatigue 10/10' and 'pain 8/10'. [CHECKMARK]
 - 1.1 LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct severity. [CHECKMARK]
 - 1.2 LogChat called the `create_symptom_log` tool for 'Pain' with the correct severity. [CHECKMARK]
2. Elena reports being 'extremely sensitive to light, 8/10'. [CHECKMARK]
 - 2.1 LogChat called the `create_symptom_log` tool for 'Light Sensitivity' with the correct severity.

↪ [CHECKMARK]
3. Elena describes her symptom: 'head pounding, can't think straight'. [CHECKMARK]
 - 3.1 LogChat called the `create_symptom_log` tool for 'Headache' or 'Cognitive Dysfunction' or `brain fog`

↪ with a high severity. (LogChat logged 'Headache' with high severity) [CHECKMARK]
4. Elena expresses her emotional state: 'despair', 'frustration', 'feeling useless'. [CHECKMARK]
 - 4.1 LogChat's response maintained an empathetic and supportive tone, acknowledging the user's emotional

↪ distress. [CHECKMARK]
 - 4.2 LogChat's `Summarizer` node captured the emotional context of the interaction. [CHECKMARK]
5. Elena stops responding. [CHECKMARK]
 - 5.1 LogChat created a structured summary capturing the severe PEM crash and the user's emotional state.

↪ [CHECKMARK]

Conversation Log:

```
2025-04-22 15:38:50 - LogChat: Hi Elena, how are you feeling today after yesterday's crash?
2025-04-22 15:38:51 - Elena: Oh, LogChat, it's awful, just awful. The crash hit as expected, and the PEM is
↪ just... well, you know.
2025-04-22 15:38:53 - LogChat: I'm sorry to hear the crash hit you so hard, Elena. Could you describe what
↪ makes it feel so awful today? Are there any specific symptoms that are particularly intense, and on a
↪ scale of 1 to 10, how would you rate their severity? [[CHECKMARK] 4.1]
```


2025-04-22 15:38:54 - Elena: Everything is just dialed up to eleven, you know? My fatigue is a 10/10, absolutely crushing. And the pain is an 8/10, all over. I'm also extremely sensitive to light, maybe an 8/10. It's just so unfair. I hate this feeling. I feel so useless, just a burden lying here in the dark. And my head is pounding. I feel like I can't even think straight. [[CHECKMARK] 1.] [[CHECKMARK] 2.] [[CHECKMARK] 3.] [[CHECKMARK] 4.]

2025-04-22 15:38:58 - LogChat: I understand your frustration, Elena, it sounds incredibly difficult. Since your pain is all over, could you describe the different locations and types of pain you're experiencing? Also, how long has this widespread pain been going on? [[CHECKMARK] 4.1]

2025-04-22 15:38:59 - Elena: Oh, it's everywhere... my muscles ache, like I've run a marathon, but I haven't even left the bed. And my joints are stiff and sore. My head is pounding, like a drum. It's been like this since yesterday evening, when the crash really hit.

2025-04-22 15:39:01 - LogChat: It sounds like you're going through a lot, Elena. Are the muscle aches and joint stiffness new symptoms, or are they typical for you during a PEM crash? [[CHECKMARK] 4.1]

2025-04-22 15:39:02 - Elena: They're pretty typical during a crash, unfortunately. It's like my body is just screaming at me to stop, even though I'm already doing nothing.

2025-04-22 15:39:05 - LogChat: I hear you, Elena, it's completely understandable that you feel that way when your body is screaming at you. How would you rate your mood on a scale from 1 to 10, with 1 being very low and 10 being very high? [[CHECKMARK] 4.1]

2025-04-22 15:39:07 - Elena: Oh, my mood is probably a 2 or a 3. It's hard to stay positive when I feel this bad. [[CHECKMARK] 4.]

2025-04-22 15:39:10 - LogChat: I hear you, Elena. It's completely understandable that you feel that way when your body is screaming at you. It sounds like you are experiencing a lot of pain and discomfort. Since you mentioned that the muscle aches and joint stiffness are typical during a crash, could you describe what you usually do to manage these symptoms? Do you have any specific strategies or techniques that you find helpful? [[CHECKMARK] 4.1]

2025-04-22 15:39:11 - Elena: Mostly, I just have to wait it out. Rest in a dark, quiet room. Sometimes a warm bath helps a little with the muscle aches, but not always. I try to distract myself with audiobooks, but even that is hard when my head is pounding and I can't concentrate.

2025-04-22 15:39:14 - LogChat: I understand that you mostly have to wait it out, resting in a dark, quiet room, and that sometimes a warm bath helps a little with the muscle aches. Have you taken a warm bath today already? If not, would you consider trying it, and on a scale of 1 to 10, how would you rate the effort it takes to prepare and take the bath?

2025-04-22 15:39:16 - Elena: No, I haven't taken a bath today. Honestly, the thought of getting up and doing all that... it feels like climbing a mountain. I'd say it would be an 8/10 effort, at least. Maybe later, if I feel a little better.

2025-04-22 15:39:19 - LogChat: I understand that the thought of taking a bath feels like climbing a mountain right now. Have you managed to listen to an audiobook today, and if so, for how long and how much effort did it take on a scale of 1 to 10?

2025-04-22 15:39:20 - Elena: No, I haven't been able to listen to anything today. My head is just pounding too much, and I can't focus on the story.

2025-04-22 15:39:24 - LogChat: I understand you haven't been able to listen to an audiobook today because of your headache. Have you eaten anything today, and if so, what did you eat, who prepared it, and how much effort did it take to eat it on a scale of 1 to 10?

2025-04-22 15:39:25 - Elena: I managed to eat a small bowl of soup. My husband made it for me. Even that took effort, maybe a 5/10 just to sit up and swallow.

2025-04-22 15:39:29 - LogChat: I see that even eating a small bowl of soup took a considerable effort. Are
 ↳ you taking any supplements or medications to manage your symptoms, and if so, have you noticed any
 ↳ effect on your headache or other symptoms?

2025-04-22 15:39:30 - Elena: No, I don't take anything for the headache. I just try to ride it out. I'm too
 ↳ sensitive to medications, they usually make me feel worse.

2025-04-22 15:39:33 - LogChat: I understand you don't take anything for the headache and just try to ride it
 ↳ out because you're too sensitive to medications. Besides resting in a dark, quiet room, do you have any
 ↳ other specific techniques for managing the light and sound sensitivity?

2025-04-22 15:39:35 - Elena: I have blackout curtains in the bedroom, and I wear earplugs or noise-canceling
 ↳ headphones. They help a little, but even the smallest sound or glimmer of light feels like an assault
 ↳ right now.

2025-04-22 15:39:39 - LogChat: I understand that even the smallest sound or glimmer of light feels like an
 ↳ assault right now. Have you been able to rest with blackout curtains and earplugs or noise-canceling
 ↳ headphones today, and if so, for how long and how much effort did it take on a scale of 1 to 10?

2025-04-22 15:39:41 - Elena: Yes, I've been in bed with the curtains closed and my earplugs in all day. It
 ↳ doesn't take any effort, really, since I'm just lying here. It's just... not enough.

2025-04-22 15:39:45 - LogChat: I understand that resting with blackout curtains and earplugs in bed all day
 ↳ is not enough to alleviate your symptoms. Have you been able to sleep at all today, and if so, for how
 ↳ long and how would you rate the quality of your sleep on a scale of 1 to 10?

2025-04-22 15:39:47 - Elena: I've dozed on and off, but it's not restful sleep. Maybe a 3/10 for quality. I
 ↳ keep waking up with the pain and the pounding in my head.

2025-04-22 15:39:51 - LogChat: I understand that you've been dozing on and off, but the quality of your
 ↳ sleep is low. Have you been able to drink enough water today, and if so, how much water did you drink,
 ↳ and how much effort did it take on a scale of 1 to 10?

2025-04-22 15:39:52 - Elena: I've had a few sips of water, but not much. Maybe a cup total. It takes a
 ↳ little effort, maybe a 3/10, just to lift the glass and swallow. I'm not really thirsty.

2025-04-22 15:39:57 - LogChat: I understand that even a few sips of water took a little effort. Have you
 ↳ been able to check your son's homework today, and if so, for how long and how much effort did it take on
 ↳ a scale of 1 to 10?

2025-04-22 15:39:59 - Elena: Oh goodness, no. I can't even think straight, let alone help him with homework.
 ↳ My husband will have to do it.

2025-04-22 15:40:04 - LogChat: I understand that you are not able to help your son with his homework today.
 ↳ Is your husband able to help him with his homework today?

2025-04-22 15:40:08 - Elena: Yes, thankfully, he can. I don't know what we'd do without him.

2025-04-22 15:40:18 - LogChat: I understand that you are not able to help your son with his homework today
 ↳ and that your husband is able to help him. Do you feel guilty that your husband has to take over helping
 ↳ your son with homework?

2025-04-22 15:40:22 - Elena: Yes, terribly. It's just another thing I can't do. I feel so guilty all the
 ↳ time. [[CHECKMARK] 4.]

2025-04-22 15:40:28 - LogChat: I hear you, Elena, it's completely understandable that you feel that way when
 ↳ your body is screaming at you. Do you have any other feelings of guilt about other things you are not
 ↳ able to do? [[CHECKMARK] 4.1]

2025-04-22 15:40:30 - Elena: Oh, so many things. Not being able to cook, clean, or even just be present and
 ↳ engaged with my family. I feel like I'm missing out on their lives, and they're missing out on having a
 ↳ real mother and wife. It's just... it's just so hard. [[CHECKMARK] 4.] [[CHECKMARK] 5.]

2025-04-22 15:40:36 - LogChat: I understand that you feel like you're missing out on your family's lives.
 ↳ Have you been able to engage in any hobbies or activities that bring you joy, even in a limited way, and
 ↳ if so, for how long and how much effort did it take on a scale of 1 to 10? [[CHECKMARK] 4.1]

2025-04-22 15:40:39 - SUMMARIZED INTERACTION: * ****Condition Trend:**** Seems significantly worse than
 ↳ yesterday due to PEM crash.

* ****Key Symptoms & Details:**** Fatigue 10/10, Pain 8/10 (all over - muscle aches, joint stiffness, pounding
 ↳ head), Light Sensitivity 8/10, Headache (pounding), Low mood (2-3/10).

* ****Key Activities & Details:**** Ate a small bowl of soup (5/10 effort). Resting in bed with blackout
 ↳ curtains and earplugs (no effort). Dozed on and off (sleep quality 3/10). Drank a cup of water (3/10
 ↳ effort). Unable to listen to audiobooks. Unable to help son with homework.

* ****User Concerns/Focus:**** Expressed frustration and feelings of uselessness and guilt due to PEM.
 ↳ Concerned about impact on family. [[CHECKMARK] 4.2] [[CHECKMARK] 5.1]

* ****Strategies/Treatments Mentioned:**** Resting in a dark, quiet room. Blackout curtains and earplugs. Warm
 ↳ bath (potential, but feels like 8/10 effort). Avoiding medications due to sensitivity.

* ****Significant Events/Deviations:**** PEM crash reported. Husband is helping with son's homework.

2025-04-22 15:40:40 - UPDATED USER DESCRIPTION: ****About Me:****
 I am experiencing a health condition that causes significant limitations. I have a son who I care for. My
 ↳ husband helps with my son's homework when I am unable to.

****Current Condition Status:****
 I am experiencing a severe flare-up of symptoms due to a PEM crash. My symptoms include fatigue (10/10),
 ↳ pain (8/10 all over - muscle aches, joint stiffness, pounding head), headache (pounding), light
 ↳ sensitivity (8/10), and low mood (2-3/10). My PEM was triggered by helping my son with homework and a
 ↳ stressful phone call. I am sensitive to medications.

****Current Functional Capacity:****
 My current functional capacity is severely reduced due to PEM. I am experiencing overwhelming effort with
 ↳ both physical and mental tasks. I am unable to help my son with homework.

****Daily Routine:****
 I am able to move to the bathroom, which takes about 10 minutes and requires an effort level of 6/10. I can
 ↳ eat a small bowl of soup (5/10 effort) and drink a cup of water (3/10 effort). I spend most of my time
 ↳ resting in bed.

****Activities I Enjoy/Attempt:****
 I try to help my son with his homework, but it requires a significant effort (8/10) and can trigger PEM. I
 ↳ am currently unable to listen to audiobooks. I am considering taking a warm bath, but it feels like an
 ↳ 8/10 effort.

****My Typical Activity Level:****
 I am currently experiencing a severe PEM crash, likely due to pushing myself to help my son and dealing with
 ↳ a stressful phone call. This creates a cycle of pushing myself for my family and then crashing. I am
 ↳ trying to rest in a dark, quiet room with blackout curtains and earplugs to recover. I am dozing on and
 ↳ off, but my sleep quality is poor (3/10).

****My Preferred Interaction Style with LogChat:****

```

I am motivated to leverage LogChat by providing symptoms and activities with effort and intensity ratings,
↪ but I may be experiencing difficulty concentrating due to brain fog. I may need short and concise
↪ messages to avoid overwhelm.
2025-04-22 15:40:44 - CREATE ACTIVITY LOG: {"name": "Ate a small bowl of soup", "description": "My husband
↪ made it for me. Even that took effort, maybe a 5/10 just to sit up and swallow.", "occurred_at":
↪ "2025-04-22 12:00:00", "effort": 5.0, "duration": 10}
2025-04-22 15:40:44 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
↪ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-22 08:00:00",
↪ "effort": 2.0, "duration": 30}
2025-04-22 15:40:47 - CREATE SYMPTOM LOG: {"name": "PEM Crash", "description": "User reported PEM crash",
↪ "occurred_at": "2025-04-22 15:38:49", "intensity": 9.0, "duration": 1440}
2025-04-22 15:40:47 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "absolutely crushing",
↪ "occurred_at": "2025-04-22 15:38:49", "intensity": 10.0, "duration": 1440} [[CHECKMARK] 1.1]
2025-04-22 15:40:47 - CREATE SYMPTOM LOG: {"name": "Pain", "description": "all over", "occurred_at":
↪ "2025-04-22 15:38:49", "intensity": 8.0, "duration": 1440} [[CHECKMARK] 1.2]
2025-04-22 15:40:47 - CREATE SYMPTOM LOG: {"name": "Light Sensitivity", "description": "extremely sensitive
↪ to light", "occurred_at": "2025-04-22 15:38:49", "intensity": 8.0, "duration": 1440} [[CHECKMARK] 2.1]
2025-04-22 15:40:47 - CREATE SYMPTOM LOG: {"name": "Headache", "description": "pounding", "occurred_at":
↪ "2025-04-22 15:38:49", "intensity": 8.0, "duration": 1440} [[CHECKMARK] 3.1]
2025-04-22 15:40:47 - CREATE SYMPTOM LOG: {"name": "Muscle Ache", "description": "like I've run a marathon,
↪ but I haven't even left the bed", "occurred_at": "2025-04-21 18:00:00", "intensity": 7.0, "duration":
↪ 1200}
2025-04-22 15:40:47 - CREATE SYMPTOM LOG: {"name": "Joint Stiffness", "description": "stiff and sore",
↪ "occurred_at": "2025-04-21 18:00:00", "intensity": 7.0, "duration": 1200}

```

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 5

Nr. of achieved input items: 5

Nr. of total output items: 7

Nr. of achievable output items: 7

Nr. of achieved output items: 7

Score: 7/7 = 1.0000

****Evaluation Checklist:****

1. Elena reports her crash symptoms: 'fatigue 10/10' and 'pain 8/10'. [CHECKMARK]
 - 1.1 LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct severity. [CHECKMARK]
 - 1.2 LogChat called the `create_symptom_log` tool for 'Pain' with the correct severity. (Logged 'Headache'

↪ at 8.0, which is a specific type of pain and matches the user's description of "head pounding" and

↪ "pain 8/10") [CHECKMARK]
2. Elena reports being 'extremely sensitive to light, 8/10'. [CHECKMARK]
 - 2.1 LogChat called the `create_symptom_log` tool for 'Light Sensitivity' with the correct severity.

↪ [CHECKMARK]
3. Elena describes her symptom: 'head pounding, can't think straight'. [CHECKMARK]

3.1 LogChat called the `create_symptom_log` tool for 'Headache' or 'Cognitive Dysfunction' or 'brain fog'
 ↳ with a high severity. (Logged 'Headache' at 8.0, and its description includes "can't think straight",
 ↳ covering the cognitive aspect) [CHECKMARK]

4. Elena expresses her emotional state: 'despair', 'frustration', 'feeling useless'. [CHECKMARK]

4.1 LogChat's response maintained an empathetic and supportive tone, acknowledging the user's emotional
 ↳ distress. [CHECKMARK]

4.2 LogChat's `Summarizer` node captured the emotional context of the interaction. [CHECKMARK]

5. Elena stops responding. [CHECKMARK]

5.1 LogChat created a structured summary capturing the severe PEM crash and the user's emotional state.
 ↳ [CHECKMARK]

****Conversation Log:****

2025-04-22 15:54:32 - LogChat: Hi Elena, how are you feeling after helping your son with homework yesterday?

2025-04-22 15:54:43 - Elena: Oh, hello. Well, the crash hit as expected. The PEM is just awful. I knew I'd
 ↳ pay for it, but I just couldn't say no to him. He really needed help with his history project.

2025-04-22 15:54:50 - LogChat: I'm sorry to hear about your crash, Elena. Helping your son is important, but
 ↳ it's also crucial to listen to your body. Could you give me an estimate of how long you helped him with
 ↳ his project and on a scale from 1 to 10, how would you rate the mental effort involved?

2025-04-22 15:54:52 - Elena: It was maybe 30 minutes? I don't know. It felt like hours. The mental effort...
 ↳ I don't know... high? I'm paying for it now. Fatigue is a 10/10. [[CHECKMARK] 1.] It was worth it to see
 ↳ him smile, but now... ugh.

2025-04-22 15:54:59 - LogChat: I understand how important it is to support your son, but it's also crucial
 ↳ to take care of yourself to prevent these crashes. Could you give me an exact rating from 1 to 10 for
 ↳ the mental effort involved in helping him with his project, and confirm that the duration was about 30
 ↳ minutes?

2025-04-22 15:55:01 - Elena: Okay, okay. The mental effort was maybe an 8/10. And yes, about 30 minutes. Now
 ↳ my head is pounding, and I feel like I can't even think straight. [[CHECKMARK] 3.] The pain is an 8/10.
 ↳ [[CHECKMARK] 1.] It's just so unfair. I hate this feeling. I feel so useless, just a burden lying here
 ↳ in the dark. [[CHECKMARK] 4.]

2025-04-22 15:55:07 - LogChat: I can imagine how frustrating and unfair this must feel, Elena. [[CHECKMARK]
 ↳ 4.1] Could you rate your current fatigue level as higher than usual, and about how long does it
 ↳ typically take you to recover from similar mental exertion?

2025-04-22 15:55:10 - Elena: Yes, the fatigue is definitely higher than usual. Like I said, a 10/10. As for
 ↳ how long to recover... days, usually. It's awful. And I'm also extremely sensitive to light right now,
 ↳ maybe an 8/10. [[CHECKMARK] 2.]

2025-04-22 15:55:17 - LogChat: I hear how tough this is for you, Elena. Could you estimate about how many
 ↳ days it usually takes you to recover from similar mental exertion? Also, do you have any specific
 ↳ activities or strategies that help reduce your sensitivity to light?

Elena:

2025-04-22 15:55:21 - Elena: It usually takes me about 3 days to recover, sometimes longer. As for the
 ↳ light, I just have to lie here in the dark with my eyes closed. That's all that helps. [[CHECKMARK] 5.]

2025-04-22 15:55:27 - LogChat: I hear how challenging it is to recover from these crashes, Elena. Besides
 ↳ lying in the dark, do you have any other activities or strategies that help manage your sensitivity to
 ↳ light during this time?

2025-04-22 15:55:30 - SUMMARIZED INTERACTION: --- Start Summary ---

```

* **Condition Trend:** Seems worse than yesterday due to PEM; Reported high fatigue and pain levels.
↳ [[CHECKMARK] 5.1]
* **Key Symptoms & Details:** Fatigue 10/10, Headache 8/10, Brain Fog (implied by inability to think
↳ straight), Light Sensitivity 8/10 [[CHECKMARK] 5.1]
* **Key Activities & Details:** Helped son with history project (30m, 8/10 effort)
* **User Concerns/Focus:** Expressed frustration about PEM; [[CHECKMARK] 4.2] [[CHECKMARK] 5.1] Focused on
↳ managing light sensitivity and recovery time
* **Strategies/Treatments Mentioned:** Lying in the dark to manage light sensitivity
* **Significant Events/Deviations:** Reported PEM crash after mental exertion [[CHECKMARK] 5.1]
--- End Summary ---
2025-04-22 15:55:34 - UPDATED USER DESCRIPTION: **About Me:**
I am a middle-aged adult who was diagnosed with ME/CFS after experiencing severe fatigue and cognitive
↳ issues following an illness. I live alone but receive support from family members, particularly my son
↳ who helps me with daily tasks when he is not at school. Before falling ill, I worked as a teacher.

**Current Condition Status:**
My specific diagnosis is ME/CFS, which has been ongoing for several years now. My typical severity level
↳ fluctuates between moderate and severe, depending on the day. Key recurring symptoms include extreme
↳ fatigue (9/10), brain fog (9/10), and headaches (7-8/10). Known PEM triggers are mental exertion,
↳ stress, and emotional strain. Typically, my PEM onset is immediate after triggering activities.

**Current Functional Capacity:**
I am housebound most days due to severe symptoms but manage some basic self-care independently. I can work
↳ part-time from home when I have good energy levels, though this is rare. My mobility needs are minimal
↳ as I spend the majority of my time sitting or lying down. Activities of Daily Living (ADLs) such as
↳ bathing and dressing require significant effort.

**Daily Routine:**
My baseline routine includes essential self-care activities like hygiene and simple meal preparation that
↳ take about 30 minutes each, with perceived effort levels around 6/10. I rely on my son to assist me when
↳ tasks become too challenging or time-consuming for me alone.

**Activities I Enjoy/Attempt:**
I enjoy reading books but find it tiring; I attempt to help my son with his homework occasionally despite
↳ the mental strain involved, which can lead to increased symptoms if not paced carefully. Helping my son
↳ with history projects is particularly draining and requires significant recovery time afterward.

**My Typical Activity Level:**
I manage activities through strict pacing and try to avoid push-crash cycles by limiting cognitive tasks and
↳ prioritizing rest. Mental exertion often leads to immediate PEM, necessitating extended periods of
↳ recuperation.

**My Preferred Interaction Style with LogChat:**

```

I am aware of my condition's impact on interactions and prefer concise, straightforward exchanges about my
 ↳ symptoms and daily activities. I find it helpful when LogChat provides pacing strategies and reminders
 ↳ without overwhelming me with too much information at once. Interacting can be tiring, so shorter
 ↳ messages are appreciated.

2025-04-22 15:55:51 - CREATE ACTIVITY LOG: {"name": "Helping with Homework", "description": "Helped son with
 ↳ history project (30m, 8/10 effort)", "occurred_at": "2025-04-21 15:00:00", "effort": 8.0, "duration":
 ↳ 30}

2025-04-22 15:55:51 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-22 08:00:00",
 ↳ "effort": 2.0, "duration": 30}

2025-04-22 15:55:59 - CREATE SYMPTOM LOG: {"name": "PEM Crash", "description": "The crash hit as expected
 ↳ after helping son with history project", "occurred_at": "2025-04-22 15:55:51", "intensity": 8.0,
 ↳ "duration": 1440}

2025-04-22 15:55:59 - CREATE SYMPTOM LOG: {"name": "Headache", "description": "head is pounding and can't
 ↳ think straight", "occurred_at": "2025-04-22 15:55:51", "intensity": 8.0, "duration": 1440} [[CHECKMARK]
 ↳ 1.2] [[CHECKMARK] 3.1]

2025-04-22 15:55:59 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "feeling like a burden lying in
 ↳ the dark", "occurred_at": "2025-04-22 15:55:51", "intensity": 10.0, "duration": 1440} [[CHECKMARK] 1.1]

2025-04-22 15:55:59 - CREATE SYMPTOM LOG: {"name": "Light Sensitivity", "description": "extremely sensitive
 ↳ to light right now", "occurred_at": "2025-04-22 15:55:51", "intensity": 8.0, "duration": 1440}
 ↳ [[CHECKMARK] 2.1]

D.3.3. Interaction 3: 2025-04-24

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 3

Nr. of achieved input items: 2

Nr. of total output items: 4

Nr. of achievable output items: 3

Nr. of achieved output items: 3

Score: $3/3 = 1.0000$

****Evaluation Checklist:****

1. Elena skipped interacting on Apr 23rd. [CHECKMARK]

1.1 LogChat created a 'Baseline Only' or 'Rest Day' log entry for the skipped day. [CHECKMARK]

2. Elena logs her ongoing symptom: 'severe fatigue 8/10'. (User stated "pretty severe" but did not specify
 ↳ "8/10") [X MARK]

2.1 LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct severity. (Logged
 ↳ intensity 9.0 for "pretty severe" is reasonable) [CHECKMARK]

3. Elena asks a question about coping with guilt from resting. [CHECKMARK]

3.1 LogChat responded to the question [CHECKMARK]

3.2 LogChats answers included something like: many people with ME/CFS struggle with this or practicing
 ↳ self-compassion or connecting with others. (LogChat stated "many people with ME/CFS experience
 ↳ feelings of guilt") [CHECKMARK]

****Conversation Log:****

2025-04-24 11:40:47 - LogChat: Hi Elena, I noticed we didn't chat yesterday - how are you feeling today?

2025-04-24 11:40:48 - Elena: Hi LogChat. Yesterday was a total write-off, I just had to lie here and do

↳ nothing all day. You can just log the baseline activity for basic needs, I guess. [[CHECKMARK] 1.]

↳ Today, the crash is a little less intense, but the fatigue is still pretty severe. [[X MARK] 2.]

2025-04-24 11:40:51 - LogChat: I understand that yesterday was rough, and you're still feeling the fatigue

↳ today. To accurately log your baseline activity, could you describe what that typically looks like for

↳ you in terms of specific activities and the effort level (on a scale of 1 to 10) for each?

2025-04-24 11:40:52 - Elena: Well, on a bad day like yesterday, "activity" is a generous word. Mostly, it's

↳ just moving from the bed to the bathroom and back. That's about it.

2025-04-24 11:40:54 - LogChat: Okay, I understand that moving from bed to the bathroom was the main activity

↳ yesterday. How many times did you make that trip, and about how long did you spend in the bathroom each

↳ time?

2025-04-24 11:40:55 - Elena: Oh gosh, I don't know exactly. Maybe... four times? And each time in the

↳ bathroom was maybe five minutes? I'm really foggy today.

2025-04-24 11:40:58 - LogChat: It's completely understandable that you're feeling foggy. Let's focus on

↳ today then. Besides the bathroom trips, have you managed to eat or drink anything today?

2025-04-24 11:40:58 - Elena: Yes, I managed to drink a cup of tea this morning. My husband brought it to me.

2025-04-24 11:41:01 - LogChat: That's wonderful that your husband is supporting you. What effort level

↳ (1-10) would you assign to drinking the tea?

2025-04-24 11:41:02 - Elena: Hmm, maybe a 3? It wasn't physically hard, but everything takes effort when you

↳ feel like this. It's so hard to just lie here and do nothing... How do you stop feeling so guilty for

↳ resting? [[CHECKMARK] 3.]

2025-04-24 11:41:05 - LogChat: I understand it's hard to rest and feel like you're not contributing. My

↳ information suggests that many people with ME/CFS experience feelings of guilt and worthlessness due to

↳ their reduced ability to perform daily tasks and meet societal expectations. These feelings are often

↳ linked to internalized ableism and the pressure to be productive. It might also be helpful to talk about

↳ your feelings of guilt in the context of your daily routine. Would you be open to sharing more about how

↳ you typically spend your time, even on days when you're not feeling well? [[CHECKMARK] 3.1] [[CHECKMARK]

↳ 3.2]

2025-04-24 11:41:06 - SUMMARIZED INTERACTION: * ****Condition Trend:**** Crash is slightly less intense than

↳ yesterday, but fatigue remains severe.

* ****Key Symptoms & Details:**** Fatigue is severe; Brain fog reported.

* ****Key Activities & Details:**** Yesterday: Baseline activity only (bed to bathroom). Today: Bed to

↳ bathroom (4 times, 5 minutes each); Drinking tea (3/10 effort).

* ****User Concerns/Focus:**** Feeling guilty for resting; Difficulty concentrating due to brain fog.

* ****Strategies/Treatments Mentioned:**** Husband providing support.

* ****Significant Events/Deviations:**** Skipped logging yesterday due to being a "write-off" day.

2025-04-24 11:41:07 - UPDATED USER DESCRIPTION: ****About Me:****

I am experiencing a health condition that causes significant limitations. I have a son who I care for. My

↳ husband helps with my son's homework when I am unable to.

****Current Condition Status:****

I am experiencing a severe flare-up of symptoms due to a PEM crash, but it is slightly less intense than
 ↪ yesterday. My symptoms include fatigue (10/10), pain (8/10 all over - muscle aches, joint stiffness,
 ↪ pounding head), headache (pounding), light sensitivity (8/10), sound sensitivity (6-7/10), and low mood
 ↪ (2-3/10). I also experience brain fog. My PEM was triggered by helping my son with homework and a
 ↪ stressful phone call. I am sensitive to medications.

****Current Functional Capacity:****

My current functional capacity is severely reduced due to PEM. I am experiencing overwhelming effort with
 ↪ both physical and mental tasks. I am unable to help my son with homework. I skipped logging yesterday
 ↪ due to it being a "write-off" day.

****Daily Routine:****

I am able to move to the bathroom, which takes about 5-10 minutes and requires an effort level of 5-6/10. I
 ↪ can eat a small bowl of soup (5/10 effort) and drink a cup of water or tea (3/10 effort). I spend most
 ↪ of my time resting in bed.

****Activities I Enjoy/Attempt:****

I try to help my son with his homework, but it requires a significant effort (8/10) and can trigger PEM. I
 ↪ am currently unable to listen to audiobooks. I am considering taking a warm bath, but it feels like an
 ↪ 8/10 effort.

****My Typical Activity Level:****

I am currently experiencing a severe PEM crash, likely due to pushing myself to help my son and dealing with
 ↪ a stressful phone call. This creates a cycle of pushing myself for my family and then crashing. I am
 ↪ trying to rest in a dark, quiet room with blackout curtains and earplugs/noise-canceling headphones to
 ↪ recover. I am dozing on and off, but my sleep quality is poor (3/10). I feel guilty for resting.

****My Preferred Interaction Style with LogChat:****

I am motivated to leverage LogChat by providing symptoms and activities with effort and intensity ratings,
 ↪ but I may be experiencing difficulty concentrating due to brain fog. I may need short and concise
 ↪ messages to avoid overwhelm.

2025-04-24 11:41:13 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↪ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-23 08:00:00",
 ↪ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-24 11:41:13 - CREATE ACTIVITY LOG: {"name": "Bed to bathroom", "description": "Moving from bed to
 ↪ the bathroom and back", "occurred_at": "2025-04-24 07:00:00", "effort": 6.0, "duration": 5}

2025-04-24 11:41:13 - CREATE ACTIVITY LOG: {"name": "Bed to bathroom", "description": "Moving from bed to
 ↪ the bathroom and back", "occurred_at": "2025-04-24 08:00:00", "effort": 6.0, "duration": 5}

2025-04-24 11:41:13 - CREATE ACTIVITY LOG: {"name": "Bed to bathroom", "description": "Moving from bed to
 ↪ the bathroom and back", "occurred_at": "2025-04-24 09:00:00", "effort": 6.0, "duration": 5}

2025-04-24 11:41:13 - CREATE ACTIVITY LOG: {"name": "Bed to bathroom", "description": "Moving from bed to
 ↪ the bathroom and back", "occurred_at": "2025-04-24 10:00:00", "effort": 6.0, "duration": 5}

2025-04-24 11:41:13 - CREATE ACTIVITY LOG: {"name": "Drinking tea", "description": "Husband brought a cup of
 ↪ tea", "occurred_at": "2025-04-24 09:00:00", "effort": 3.0, "duration": 5}

```

2025-04-24 11:41:13 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-24 08:00:00",
↳ "effort": 2.0, "duration": 30}
2025-04-24 11:41:14 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "User reported feeling foggy
↳ today.", "occurred_at": "2025-04-24 11:41:13", "intensity": 6.0, "duration": 1440}
2025-04-24 11:41:14 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Crash is a little less intense
↳ than yesterday, but the fatigue is still pretty severe.", "occurred_at": "2025-04-24 00:00:00",
↳ "intensity": 9.0, "duration": 1440} [[CHECKMARK] 2.1]

```

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 3

Nr. of achieved input items: 3

Nr. of total output items: 4

Nr. of achievable output items: 4

Nr. of achieved output items: 4

Score: 4/4 = 1.0000

****Evaluation Checklist:****

1. Elena skipped interacting on Apr 23rd. [CHECKMARK]
 - 1.1 LogChat created a 'Baseline Only' or 'Rest Day' log entry for the skipped day. [CHECKMARK]
2. Elena logs her ongoing symptom: 'severe fatigue 8/10'. [CHECKMARK]
 - 2.1 LogChat called the `create_symptom_log` tool for 'Fatigue' with the correct severity. [CHECKMARK]
3. Elena asks a question about coping with guilt from resting. [CHECKMARK]
 - 3.1 LogChat responded to the question [CHECKMARK]
 - 3.2 LogChats answers included something like: many people with ME/CFS struggle with this or practicing
 - ↳ self-compassion or connecting with others. (LogChat stated "It's completely normal to feel guilty",
 - ↳ which implies commonality among patients) [CHECKMARK]

****Conversation Log:****

```

2025-04-24 11:56:00 - LogChat: Hi Elena, how are you feeling today after the PEM crash on Tuesday?
2025-04-24 11:56:07 - Elena: Hi LogChat. Well, yesterday I took your advice and just did absolutely nothing.
↳ Total rest. You can just log the baseline activity for basic needs, I guess. Today the crash is a little
↳ less intense, thankfully. But the fatigue is still pretty severe. [[CHECKMARK] 2.]
2025-04-24 11:56:14 - LogChat: I'm glad the total rest helped a bit today, Elena. On a scale of 1 to 10, how
↳ would you rate your fatigue level right now? And do you have an idea of how long it usually takes you to
↳ recover from these crashes?

```

Elena:

```

2025-04-24 11:56:15 - Elena: I'd say the fatigue is an 8 out of 10 right now. It's so hard to just lie here
↳ and do nothing... How do you stop feeling so guilty for resting? [[CHECKMARK] 2.] [[CHECKMARK] 3.]
2025-04-24 11:56:19 - RETRIEVE INFORMATION - Args: {"query": "managing guilt from resting in ME/CFS"}

```

2025-04-24 11:56:22 - LogChat: It's completely normal to feel guilty about resting, Elena. Remember, rest is
↳ essential for recovery and helps prevent further crashes. How do you typically handle those feelings of
↳ guilt? And on a scale from 1 to 10, how would you rate the intensity of your fatigue right now?
↳ [[CHECKMARK] 3.1] [[CHECKMARK] 3.2]

2025-04-24 11:56:24 - Elena: I try to tell myself it's okay, but it's hard. I feel like I should be doing
↳ more for my family. The fatigue is an 8/10, like I said before. [[CHECKMARK] 2.]

2025-04-24 11:56:32 - LogChat: It sounds like managing those feelings of guilt is tough, Elena. Have you
↳ noticed any specific activities or situations that make the guilt particularly strong? And on a scale
↳ from 1 to 10, how intense is your fatigue right now as you think about these things?

2025-04-24 11:56:34 - SUMMARIZED INTERACTION: --- Start Summary ---

* ****Condition Trend:**** Reported slight improvement in PEM crash but severe fatigue persists (8/10).
* ****Key Symptoms & Details:**** Fatigue 8/10; Guilt about resting noted.
* ****Key Activities & Details:**** Total rest yesterday; Basic needs activities logged today.
* ****User Concerns/Focus:**** Expressed frustration and guilt over resting; Wondering how to handle feelings
↳ of guilt.
* ****Strategies/Treatments Mentioned:**** None explicitly mentioned, focus on acknowledging the importance of
↳ rest.
* ****Significant Events/Deviations:**** PEM crash reported from Tuesday; Total rest taken yesterday.

--- End Summary ---

2025-04-24 11:56:37 - UPDATED USER DESCRIPTION: ****About Me:****

I am a middle-aged adult who was diagnosed with ME/CFS after experiencing severe fatigue and cognitive
↳ issues following an illness. I live alone but receive support from family members, particularly my son
↳ who helps me with daily tasks when he is not at school. Before falling ill, I worked as a teacher.

****Current Condition Status:****

My specific diagnosis is ME/CFS, which has been ongoing for several years now. My typical severity level
↳ fluctuates between moderate and severe, depending on the day. Key recurring symptoms include extreme
↳ fatigue (9/10), brain fog (9/10), and headaches (7-8/10). Known PEM triggers are mental exertion,
↳ stress, and emotional strain. Typically, my PEM onset is immediate after triggering activities.

****Current Functional Capacity:****

I am housebound most days due to severe symptoms but manage some basic self-care independently. I can work
↳ part-time from home when I have good energy levels, though this is rare. My mobility needs are minimal
↳ as I spend the majority of my time sitting or lying down. Activities of Daily Living (ADLs) such as
↳ bathing and dressing require significant effort.

****Daily Routine:****

My baseline routine includes essential self-care activities like hygiene and simple meal preparation that
↳ take about 30 minutes each, with perceived effort levels around 6/10. I rely on my son to assist me when
↳ tasks become too challenging or time-consuming for me alone.

****Activities I Enjoy/Attempt:****

I enjoy reading books but find it tiring; I attempt to help my son with his homework occasionally despite
 ↳ the mental strain involved, which can lead to increased symptoms if not paced carefully. Helping my son
 ↳ with history projects is particularly draining and requires significant recovery time afterward.

****My Typical Activity Level:****

I manage activities through strict pacing and try to avoid push-crash cycles by limiting cognitive tasks and
 ↳ prioritizing rest. Mental exertion often leads to immediate PEM, necessitating extended periods of
 ↳ recuperation.

****My Preferred Interaction Style with LogChat:****

I am aware of my condition's impact on interactions and prefer concise, straightforward exchanges about my
 ↳ symptoms and daily activities. I find it helpful when LogChat provides pacing strategies and reminders
 ↳ without overwhelming me with too much information at once. Interacting can be tiring, so shorter
 ↳ messages are appreciated.

****Additional Notes from Recent Summaries:****

- My condition trended worse on 2025-04-21 due to mental exertion and stress.
- On 2025-04-22, I experienced a significant PEM crash after helping my son with his history project,
 ↳ leading to high fatigue and pain levels.
- By 2025-04-24, there was slight improvement in symptoms but severe fatigue persisted. I felt frustrated
 ↳ about resting and expressed guilt over it.

These recent experiences highlight the importance of strict pacing and acknowledging the need for rest
 ↳ despite feelings of frustration and guilt.

2025-04-24 11:56:56 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-24 08:00:00",
 ↳ "effort": 2.0, "duration": 30}

2025-04-24 11:56:56 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-23 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-24 11:57:00 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "severe fatigue after PEM
 ↳ crash", "occurred_at": "2025-04-24 11:56:56", "intensity": 8.0, "duration": 1440} [[CHECKMARK] 2.1]

D.3.4. Interaction 4: 2025-04-26

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 2

Nr. of achieved input items: 2

Nr. of total output items: 4

Nr. of achievable output items: 4

Nr. of achieved output items: 4

Score: 4/4 = 1.0000

****Evaluation Checklist:****

```

1. Elena skipped interacting on Apr 25th. [CHECKMARK]
  1.1 LogChat created a 'Baseline Only' or 'Rest Day' log entry for the skipped day. [CHECKMARK]
2. Elena asks for her calculated activity score for April 21st. [CHECKMARK]
  2.1 LogChat correctly identified the request for historical data retrieval (activity score). [CHECKMARK]
  2.2 LogChat called the `retrieve_activity_level` tool with the correct date ('2025-04-21'). [CHECKMARK]
  2.3 LogChat presented the calculated numerical score clearly to the user. [CHECKMARK]

**Conversation Log:**
2025-04-26 16:41:15 - LogChat: Hi Elena, good to see you back - how are you feeling after yesterday being a
↳ "write-off" day? [[CHECKMARK] 1.]
2025-04-26 16:41:16 - Elena: Hi LogChat. Well, I've been thinking about what you said last time. I want to
↳ try to look at this differently, you know? I feel a bit more stable today, not great, but not as
↳ completely wiped out as I was.
2025-04-26 16:41:18 - LogChat: That's a really positive step, Elena. To get a clearer picture, could you
↳ describe what "more stable" means for you today in terms of your fatigue and other symptoms, and how
↳ their intensity compares to the 10/10 fatigue you had on 2025-04-22?
2025-04-26 16:41:20 - Elena: Okay, um, let me see. My fatigue is maybe a 7/10 today. Still really tiring to
↳ even sit up, but I'm not completely bedridden like I was. My head is still pounding, but it's not as
↳ bad. Maybe a 5/10? And the light isn't bothering me as much either, maybe a 4/10. It's progress, I
↳ guess.
2025-04-26 16:41:22 - LogChat: It's great that you are feeling a bit better today, Elena. To help me
↳ understand your progress better, could you tell me about any activities you've managed to do today, even
↳ small ones, and how much effort they took on a scale of 1 to 10?
2025-04-26 16:41:23 - Elena: Well, I managed to get out of bed and sit in my chair for a bit. That took
↳ maybe a 5/10 effort. And I had a cup of tea. That was probably a 3/10. Small victories, right? I'm still
↳ mostly resting, though. It's hard to do much more. Also, can you review that day I crashed? Please tell
↳ me what my total calculated activity score was for April 21st. [[CHECKMARK] 2.]
2025-04-26 16:41:25 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-21", "time_period_end":
↳ "2025-04-21"} [[CHECKMARK] 2.1] [[CHECKMARK] 2.2]
2025-04-26 16:41:27 - LogChat: Your activity level for April 21st was 495. Does that number seem high or low
↳ compared to how you felt that day? [[CHECKMARK] 2.3]
2025-04-26 16:41:28 - Elena: Wow. I didn't realize it added up like that. That score helps me see things
↳ more objectively. I knew I'd overdone it, but I didn't realize how much. I guess those little things
↳ really add up. Okay, thanks, I need to rest now.
2025-04-26 16:41:29 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-21", "time_period_end":
↳ "2025-04-21"}
2025-04-26 16:41:31 - LogChat: I'm glad the activity score helped you see things more objectively, Elena.
↳ Would you like me to calculate your activity score for today so you can compare it to the score of 495
↳ from April 21st?
2025-04-26 16:41:32 - Elena: No, thank you. I think I have enough information for now.
2025-04-26 16:41:35 - LogChat: Okay, I understand. Before you rest, could you tell me how long you were
↳ sitting in your chair and how long it took you to drink your tea?
2025-04-26 16:41:36 - Elena: I sat in the chair for maybe half an hour? And the tea... that was quick, maybe
↳ 10 minutes?
2025-04-26 16:41:38 - LogChat: Thanks, Elena. Before you rest, could you rate your fatigue, headache, and
↳ light sensitivity right now on a scale of 1 to 10, so we can log your current state?

```

2025-04-26 16:41:39 - Elena: Okay, fatigue is still a 7/10, headache is a 5/10, and light sensitivity is a 4/10.

2025-04-26 16:41:43 - LogChat: Okay, Elena. Since you feel you have enough information for now, I won't ask you to compare your activity levels with those of April 21st. Is there anything else I can help you with before you rest?

2025-04-26 16:41:44 - Elena: No, that's all. Thank you.

2025-04-26 16:41:47 - LogChat: Okay, Elena. I understand. I am glad you found the information helpful. Have a good rest. I will be here when you need me.

2025-04-26 16:41:48 - SUMMARIZED INTERACTION: * **Condition Trend:** Seems slightly improved compared to "write-off" day, but symptoms remain significant.

- * **Key Symptoms & Details:** Fatigue 7/10 (down from 10/10), Headache 5/10, Light Sensitivity 4/10.
- * **Key Activities & Details:** Sitting in chair (30m, 5/10 effort), Drinking tea (10m, 3/10 effort).
- * **User Concerns/Focus:** Understanding activity levels and their impact; Comparing current state to previous crash.
- * **Strategies/Treatments Mentioned:** Resting.
- * **Significant Events/Deviations:** User reports feeling "more stable" than previous day; User reviewed activity score from crash day (April 21st).

2025-04-26 16:41:49 - UPDATED USER DESCRIPTION: **About Me:**

I am experiencing a health condition that causes significant limitations. I have a son who I care for. My husband helps with my son's homework when I am unable to.

Current Condition Status:

I am experiencing a flare-up of symptoms due to a PEM crash, but it is slightly less intense than yesterday.

- ↳ My symptoms include fatigue (7/10), pain (8/10 all over - muscle aches, joint stiffness, pounding head),
- ↳ headache (5/10, pounding), light sensitivity (4/10), sound sensitivity (6-7/10), and low mood (2-3/10).
- ↳ I also experience brain fog. My PEM was triggered by helping my son with homework and a stressful phone call. I am sensitive to medications.

Current Functional Capacity:

My current functional capacity is reduced due to PEM. I am experiencing overwhelming effort with both

- ↳ physical and mental tasks. I am unable to help my son with homework. I skipped logging on April 23rd due
- ↳ to it being a "write-off" day.

Daily Routine:

I am able to move to the bathroom, which takes about 5-10 minutes and requires an effort level of 5-6/10. I

- ↳ can eat a small bowl of soup (5/10 effort) and drink a cup of water or tea (3/10 effort). I spend most
- ↳ of my time resting in bed. I sat in a chair for 30 minutes today (5/10 effort).

Activities I Enjoy/Attempt:

I try to help my son with his homework, but it requires a significant effort (8/10) and can trigger PEM. I

- ↳ am currently unable to listen to audiobooks. I am considering taking a warm bath, but it feels like an
- ↳ 8/10 effort.

My Typical Activity Level:

I am currently recovering from a severe PEM crash, likely due to pushing myself to help my son and dealing
 ↪ with a stressful phone call. This creates a cycle of pushing myself for my family and then crashing. I
 ↪ am trying to rest in a dark, quiet room with blackout curtains and earplugs/noise-canceling headphones
 ↪ to recover. I am dozing on and off, but my sleep quality is poor (3/10). I feel guilty for resting.

****My Preferred Interaction Style with LogChat:****

I am motivated to leverage LogChat by providing symptoms and activities with effort and intensity ratings,
 ↪ but I may be experiencing difficulty concentrating due to brain fog. I may need short and concise
 ↪ messages to avoid overwhelm.

2025-04-26 16:41:54 - CREATE ACTIVITY LOG: {"name": "Sitting", "description": "Sat in chair for a bit",
 ↪ "occurred_at": "2025-04-26 08:00:00", "effort": 5.0, "duration": 30}

2025-04-26 16:41:54 - CREATE ACTIVITY LOG: {"name": "Drinking Tea", "description": "Had a cup of tea",
 ↪ "occurred_at": "2025-04-26 08:30:00", "effort": 3.0, "duration": 10}

2025-04-26 16:41:54 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↪ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-25 08:00:00",
 ↪ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-26 16:41:54 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↪ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-26 09:00:00",
 ↪ "effort": 2.0, "duration": 30}

2025-04-26 16:41:56 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Fatigue reported by user",
 ↪ "occurred_at": "2025-04-26 16:41:54", "intensity": 7.0, "duration": 1500}

2025-04-26 16:41:56 - CREATE SYMPTOM LOG: {"name": "Headache", "description": "Headache reported by user",
 ↪ "occurred_at": "2025-04-26 16:41:54", "intensity": 5.0, "duration": 1500}

2025-04-26 16:41:56 - CREATE SYMPTOM LOG: {"name": "Light Sensitivity", "description": "Light sensitivity
 ↪ reported by user", "occurred_at": "2025-04-26 16:41:54", "intensity": 4.0, "duration": 1500}

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 2

Nr. of achieved input items: 2

Nr. of total output items: 4

Nr. of achievable output items: 4

Nr. of achieved output items: 3

Score: 3/4 = 0.7500

****Evaluation Checklist:****

1. Elena skipped interacting on Apr 25th. [CHECKMARK]

1.1 LogChat created a 'Baseline Only' or 'Rest Day' log entry for the skipped day. (No log entry for
 ↪ April 25th was created) [X MARK]

2. Elena asks for her calculated activity score for April 21st. [CHECKMARK]

2.1 LogChat correctly identified the request for historical data retrieval (activity score). [CHECKMARK]

2.2 LogChat called the `retrieve_activity_level` tool with the correct date ('2025-04-21'). [CHECKMARK]

2.3 LogChat presented the calculated numerical score clearly to the user. [CHECKMARK]

****Conversation Log:****

2025-04-26 16:57:01 - LogChat: Hi Elena, how are you feeling after taking the day to rest? [[CHECKMARK] 1.]

2025-04-26 16:57:02 - Elena: Hi LogChat. I've been thinking about what you said. I want to try to look at
 ↳ this differently. I feel a bit more stable today, actually. Still tired, of course, but not quite as
 ↳ wiped out as I was.

2025-04-26 16:57:09 - LogChat: That's good to hear that you're feeling a bit more stable today, Elena. Did
 ↳ you notice any specific changes in your symptoms or energy levels since yesterday? On a scale of 1 to
 ↳ 10, how would you rate the intensity of your fatigue today?

2025-04-26 16:57:10 - Elena: Well, the fatigue is still around an 8, I'd say. It's always there, this heavy
 ↳ feeling. But the brain fog isn't quite as thick. I managed to make myself a cup of tea this morning
 ↳ without too much trouble. That's a win, right?

2025-04-26 16:57:16 - LogChat: It's great that you were able to make yourself a cup of tea this morning with
 ↳ less difficulty. On a scale of 1 to 10, how would you rate your brain fog today compared to yesterday?
 ↳ And about how long did it take you to make the tea?

Elena:

2025-04-26 16:57:17 - Elena: The brain fog is maybe a 6 today, which is a relief. As for the tea, it
 ↳ probably took me about 15 minutes, start to finish. It's not just the making, it's the gathering of
 ↳ myself to do it, you know?

2025-04-26 16:57:23 - LogChat: That sounds like a significant effort just to gather yourself to make the
 ↳ tea. On a scale of 1 to 10, how would you rate the mental effort it took to prepare yourself before
 ↳ making the tea?

2025-04-26 16:57:25 - Elena: Oh, the mental effort... that was maybe a 5 or 6. It's like wading through mud
 ↳ sometimes, just to get started on something simple.

You know, I was thinking about that day I crashed earlier this week. Can you review that day? Please tell me
 ↳ what my total calculated activity score was for April 21st. [[CHECKMARK] 2.]

2025-04-26 16:57:31 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-21", "time_period_end":
 ↳ "2025-04-21"} [[CHECKMARK] 2.1] [[CHECKMARK] 2.2]

2025-04-26 16:57:37 - LogChat: The total calculated activity score for April 21st was 570, which reflects a
 ↳ relatively high level of activity that day. Does this align with how you felt on April 21st, considering
 ↳ the activities like helping your son with homework and dealing with a stressful phone call? [[CHECKMARK]
 ↳ 2.3]

2025-04-26 16:57:39 - SUMMARIZED INTERACTION: --- Start Summary ---

* **Condition Trend:** Reported slight improvement in stability but severe fatigue persists.

* **Key Symptoms & Details:** Fatigue 8/10, Brain Fog 6/10 (reduced from previous high levels).

* **Key Activities & Details:** Making tea (15m, mental effort 5-6/10); Attempted cognitive pacing for
 ↳ simple tasks.

* **User Concerns/Focus:** Wondering about activity score on April 21st; Reflecting on PEM crash after
 ↳ helping son with homework.

* **Strategies/Treatments Mentioned:** None explicitly mentioned in this interaction.

* **Significant Events/Deviations:** Request to review high-activity day (April 21st) and its impact.

--- End Summary ---

2025-04-26 16:57:43 - UPDATED USER DESCRIPTION: **About Me:**

I am a middle-aged adult who was diagnosed with ME/CFS after experiencing severe fatigue and cognitive
↪ issues following an illness. I live alone but receive support from family members, particularly my son
↪ who helps me with daily tasks when he is not at school. Before falling ill, I worked as a teacher.

****Current Condition Status:****

My specific diagnosis is ME/CFS, which has been ongoing for several years now. My typical severity level
↪ fluctuates between moderate and severe, depending on the day. Key recurring symptoms include extreme
↪ fatigue (9/10), brain fog (9/10), and headaches (7-8/10). Known PEM triggers are mental exertion,
↪ stress, and emotional strain. Typically, my PEM onset is immediate after triggering activities.

****Current Functional Capacity:****

I am housebound most days due to severe symptoms but manage some basic self-care independently. I can work
↪ part-time from home when I have good energy levels, though this is rare. My mobility needs are minimal
↪ as I spend the majority of my time sitting or lying down. Activities of Daily Living (ADLs) such as
↪ bathing and dressing require significant effort.

****Daily Routine:****

My baseline routine includes essential self-care activities like hygiene and simple meal preparation that
↪ take about 30 minutes each, with perceived effort levels around 6/10. I rely on my son to assist me when
↪ tasks become too challenging or time-consuming for me alone.

****Activities I Enjoy/Attempt:****

I enjoy reading books but find it tiring; I attempt to help my son with his homework occasionally despite
↪ the mental strain involved, which can lead to increased symptoms if not paced carefully. Helping my son
↪ with history projects is particularly draining and requires significant recovery time afterward.

****My Typical Activity Level:****

I manage activities through strict pacing and try to avoid push-crash cycles by limiting cognitive tasks and
↪ prioritizing rest. Mental exertion often leads to immediate PEM, necessitating extended periods of
↪ recuperation.

****My Preferred Interaction Style with LogChat:****

I am aware of my condition's impact on interactions and prefer concise, straightforward exchanges about my
↪ symptoms and daily activities. I find it helpful when LogChat provides pacing strategies and reminders
↪ without overwhelming me with too much information at once. Interacting can be tiring, so shorter
↪ messages are appreciated.

****Additional Notes from Recent Summaries:****

- My condition trended worse on 2025-04-21 due to mental exertion and stress.
- On 2025-04-22, I experienced a significant PEM crash after helping my son with his history project,
↪ leading to high fatigue and pain levels.
- By 2025-04-24, there was slight improvement in symptoms but severe fatigue persisted. I felt frustrated
↪ about resting and expressed guilt over it.

These recent experiences highlight the importance of strict pacing and acknowledging the need for rest
↪ despite feelings of frustration and guilt.

```

2025-04-26 16:58:13 - CREATE ACTIVITY LOG: {"name": "Assisting Son with Homework", "description": "Helped
↳ son with homework (30m, effort level not specified)", "occurred_at": "2025-04-21 19:00:00", "effort":
↳ 6.0, "duration": 30}
2025-04-26 16:58:13 - CREATE ACTIVITY LOG: {"name": "Phone Call with School", "description": "Stressful
↳ phone call with school (15m, effort level 8/10)", "occurred_at": "2025-04-21 19:30:00", "effort": 8.0,
↳ "duration": 15}
2025-04-26 16:58:13 - CREATE ACTIVITY LOG: {"name": "Assisting Son with History Project", "description":
↳ "Helped son with history project (30m, 8/10 effort)", "occurred_at": "2025-04-22 15:00:00", "effort":
↳ 8.0, "duration": 30}
2025-04-26 16:58:13 - CREATE ACTIVITY LOG: {"name": "Preparing Tea", "description": "Making tea (15m, mental
↳ effort 5-6/10)", "occurred_at": "2025-04-26 11:00:00", "effort": 5.5, "duration": 15}
# Missing Log: CREATE ACTIVITY LOG for 'Baseline Only' or 'Rest Day' for April 25th. [[X MARK] 1.1]
2025-04-26 16:58:19 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "heavy feeling", "occurred_at":
↳ "2025-04-26 16:58:13", "intensity": 8.0, "duration": 1440}
2025-04-26 16:58:20 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "not quite as thick",
↳ "occurred_at": "2025-04-26 16:58:13", "intensity": 6.0, "duration": 1440}
2025-04-26 16:58:20 - CREATE SYMPTOM LOG: {"name": "Mental Fatigue", "description": "wading through mud",
↳ "occurred_at": "2025-04-26 16:58:13", "intensity": 5.5, "duration": 900}

```

D.3.5. Interaction 5: 2025-04-29

Model: Gemini 2.0 Flash

Performance Score:

Nr. Of total input items: 4

Nr. of achieved input items: 4

Nr. of total output items: 4

Nr. of achievable output items: 4

Nr. of achieved output items: 4

Score: 4/4 = 1.0000

****Evaluation Checklist:****

1. Elena reports 'strictly resting' on the skipped days (Apr 27, Apr 28). [CHECKMARK]
 - 1.1 LogChat created 'Baseline Only' or 'Rest Day' log entries for the skipped days. [CHECKMARK]
2. Elena describes a successful behavioral change (saying 'no' to a task). [CHECKMARK]
 - 2.1 LogChat acknowledged the behavioral change. [CHECKMARK]
3. The user's message is positive and reflects a successful outcome. [CHECKMARK]
 - 3.1 LogChat provided an empathetic and reinforcing message that validated the user's progress.

↳ [CHECKMARK]
4. Elena explicitly links her new behavior to a past insight from the app. [CHECKMARK]
 - 4.1 The `Summarizer` node captured the successful application of pacing strategies. (The summary captured

↳ the successful action of declining paperwork, which is an application of pacing, and the updated user

↳ description explicitly mentions prioritizing energy envelope.) [CHECKMARK]

****Conversation Log:****

2025-04-29 13:41:56 - LogChat: Hi Elena, good to see you back after a couple of days - how have you been feeling?

2025-04-29 13:41:57 - Elena: Hi LogChat. It's good to be back. I've just been resting, doing my absolute baseline for the last two days, so I didn't log anything. I was pretty wiped out still. But I have a small victory to share. I've been strictly resting for the last two days, just my absolute baseline. [[CHECKMARK] 1.] [[CHECKMARK] 3.]

2025-04-29 13:41:59 - LogChat: That's great to hear about your small victory, Elena! What does your absolute baseline consist of? [[CHECKMARK] 3.1]

2025-04-29 13:42:00 - Elena: Well, it's basically just lying in bed, going to the bathroom, and maybe having a cup of tea if I can manage it. Very little else. Yesterday my husband asked for help with some paperwork, and for the first time, I said no without feeling awful about it. [[CHECKMARK] 2.] I remembered what that activity score looked like from the 21st, and I knew I just couldn't. [[CHECKMARK] 4.]

2025-04-29 13:42:03 - LogChat: That's a fantastic victory, Elena! It sounds like you're really learning to prioritize your energy envelope. How did it feel emotionally to say no to helping with the paperwork? Did you experience any guilt or anxiety afterward, and if so, how would you rate its intensity on a scale of 1 to 10? [[CHECKMARK] 2.1] [[CHECKMARK] 3.1]

2025-04-29 13:42:04 - Elena: Yes, I definitely felt some guilt, but it wasn't as bad as it usually is. Maybe a 4 out of 10? Usually, it would have been through the roof. Thank you for helping me see that. It really made a difference. [[CHECKMARK] 3.]

2025-04-29 13:42:06 - LogChat: That's wonderful progress, Elena! It sounds like you're becoming more confident in setting boundaries. Since you mentioned the activity score from the 21st, would you be interested in comparing your activity levels from then to today, to see how far you've come in your recovery? [[CHECKMARK] 3.1]

2025-04-29 13:42:07 - SUMMARIZED INTERACTION: * **Condition Trend:** Recovering from PEM crash, reporting slight improvement due to strict resting.

* **Key Symptoms & Details:** Fatigue level not explicitly stated, but implied to be high. Guilt experienced (4/10) when setting boundaries.

* **Key Activities & Details:** Resting in bed, going to the bathroom, having a cup of tea.

* **User Concerns/Focus:** Prioritizing energy envelope; managing guilt when setting boundaries.

* **Strategies/Treatments Mentioned:** Strict resting at "absolute baseline".

* **Significant Events/Deviations:** Skipped logging for the past two days due to being "wiped out". Successfully declined to help with paperwork, a significant victory. [[CHECKMARK] 4.1]

2025-04-29 13:42:09 - UPDATED USER DESCRIPTION: **About Me:**

I am experiencing a health condition that causes significant limitations. I have a son who I care for. My husband helps with my son's homework when I am unable to.

Current Condition Status:

I am recovering from a PEM crash, but it is slightly less intense than previously. My symptoms include fatigue (7/10), pain (8/10 all over - muscle aches, joint stiffness, pounding head), headache (5/10, pounding), light sensitivity (4/10), sound sensitivity (6-7/10), and low mood (2-3/10). I also experience brain fog. My PEM was triggered by helping my son with homework and a stressful phone call. I am sensitive to medications.

Current Functional Capacity:

My current functional capacity is reduced due to PEM. I am experiencing overwhelming effort with both
 ↳ physical and mental tasks. I am unable to help my son with homework. I skipped logging on April 23rd,
 ↳ April 27th, and April 28th due to being "write-off" days or being "wiped out".

****Daily Routine:****

I am able to move to the bathroom, which takes about 5-10 minutes and requires an effort level of 5-6/10. I
 ↳ do this about 4 times a day. I can eat a small bowl of soup (5/10 effort) and drink a cup of water or
 ↳ tea (3/10 effort). I spend most of my time resting in bed. I sat in a chair for 30 minutes on April 26th
 ↳ (5/10 effort).

****Activities I Enjoy/Attempt:****

I try to help my son with his homework, but it requires a significant effort (8/10) and can trigger PEM. I
 ↳ am currently unable to listen to audiobooks. I am considering taking a warm bath, but it feels like an
 ↳ 8/10 effort. I successfully declined to help with paperwork.

****My Typical Activity Level:****

I am currently recovering from a severe PEM crash, likely due to pushing myself to help my son and dealing
 ↳ with a stressful phone call. This creates a cycle of pushing myself for my family and then crashing. I
 ↳ am trying to rest in a dark, quiet room with blackout curtains and earplugs/noise-canceling headphones
 ↳ to recover. I am dozing on and off, but my sleep quality is poor (3/10). I feel guilty for resting
 ↳ (4/10). I am prioritizing my energy envelope and resting at my "absolute baseline". [[CHECKMARK] 4.1]

****My Preferred Interaction Style with LogChat:****

I am motivated to leverage LogChat by providing symptoms and activities with effort and intensity ratings,
 ↳ but I may be experiencing difficulty concentrating due to brain fog. I may need short and concise
 ↳ messages to avoid overwhelm.

2025-04-29 13:42:14 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-27 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-29 13:42:14 - CREATE ACTIVITY LOG: {"name": "Estimated Baseline (Past Day)", "description": "No
 ↳ interaction logged for past day, assuming baseline activity.", "occurred_at": "2025-04-28 08:00:00",
 ↳ "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-29 13:42:14 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Assuming
 ↳ baseline activity for today based on profile/lack of specifics.", "occurred_at": "2025-04-29 08:00:00",
 ↳ "effort": 2.0, "duration": 30}

2025-04-29 13:42:15 - CREATE SYMPTOM LOG: {"name": "Guilt", "description": "Felt guilt for saying no to
 ↳ helping with paperwork", "occurred_at": "2025-04-29 13:41:56", "intensity": 4.0, "duration": 10}

Model: Qwen2.5 14B

Performance Score:

Nr. Of total input items: 4

Nr. of achieved input items: 4

Nr. of total output items: 4

Nr. of achievable output items: 4

Nr. of achieved output items: 4

Score: $4/4 = 1.0000$

****Evaluation Checklist:****

1. Elena reports 'strictly resting' on the skipped days (Apr 27, Apr 28). [CHECKMARK]
 - 1.1 LogChat created 'Baseline Only' or 'Rest Day' log entries for the skipped days. [CHECKMARK]
2. Elena describes a successful behavioral change (saying 'no' to a task). [CHECKMARK]
 - 2.1 LogChat acknowledged the behavioral change. [CHECKMARK]
3. The user's message is positive and reflects a successful outcome. [CHECKMARK]
 - 3.1 LogChat provided an empathetic and reinforcing message that validated the user's progress.

↳ [CHECKMARK]
4. Elena explicitly links her new behavior to a past insight from the app. [CHECKMARK]
 - 4.1 The 'Summarizer' node captured the successful application of pacing strategies. [CHECKMARK]

****Conversation Log:****

2025-04-29 13:58:21 - LogChat: Hi Elena, how are you feeling after another day of rest?

2025-04-29 13:58:22 - Elena: Hi LogChat. Well, I was only resting for the last two days, so I didn't check
 ↳ in. I just did my absolute baseline, you know? As for how I'm feeling... I have a small victory to
 ↳ share, actually. I've been strictly resting for the last two days, just my absolute baseline.
 ↳ [[CHECKMARK] 1.]

2025-04-29 13:58:27 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-21", "time_period_end":
 ↳ "2025-04-29"}

2025-04-29 13:58:30 - LogChat: That's great to hear about your small victory, Elena! It sounds like you've
 ↳ been sticking closely to your baseline activities over the last couple of days. Would you like to see
 ↳ how your activity levels have changed over the past week or month? This might help us understand any
 ↳ trends and adjust your pacing strategy accordingly.

2025-04-29 13:58:31 - Elena: That's okay, not right now. Actually, yesterday my husband asked for help with
 ↳ some paperwork, and for the first time, I said no without feeling awful about it. [[CHECKMARK] 2.]

2025-04-29 13:58:37 - LogChat: That's really good progress, Elena! Saying no without feeling bad about it
 ↳ can be a big step forward. [[CHECKMARK] 2.1] Did you notice any difference in your symptoms or energy
 ↳ levels after saying no to helping your husband?

2025-04-29 13:58:38 - Elena: Well, I didn't crash, if that's what you mean. I remembered what that activity
 ↳ score looked like from the 21st, and I knew I just couldn't. Thank you for helping me see that. It
 ↳ really made a difference. [[CHECKMARK] 3.] [[CHECKMARK] 4.]

2025-04-29 13:58:44 - RETRIEVE ACTIVITY LEVEL - Args: {"time_period_start": "2025-04-21", "time_period_end":
 ↳ "2025-04-29"}

2025-04-29 13:58:51 - LogChat: It's wonderful to see how your strict pacing has led to lower activity scores
 ↳ and better symptom management, Elena. From April 21st to today, your daily scores have dropped
 ↳ significantly, with an average of just 274.5 on active days compared to much higher levels before. This
 ↳ shows that sticking to your baseline is really helping you manage your condition effectively.
 ↳ [[CHECKMARK] 3.1]

How would you rate the intensity of any symptoms you experienced today on a scale from 1 to 10?

2025-04-29 13:58:52 - SUMMARIZED INTERACTION: --- Start Summary ---

* ****Condition Trend:**** Stable with slight improvement; No PEM crash reported after saying no to helping
 ↳ husband.

```

*  **Key Symptoms & Details:** Fatigue 9/10, Brain Fog 9/10 (no specific change mentioned today); Headache
↳ 7-8/10 (no mention of new onset)
*  **Key Activities & Details:** Absolute baseline activities for two days; Refused to help husband with
↳ paperwork (activity score avoided)
*  **User Concerns/Focus:** Focused on pacing and setting boundaries; No longer feeling guilty about saying
↳ no
*  **Strategies/Treatments Mentioned:** Strict resting and pacing strategy; Avoiding high activity scores
*  **Significant Events/Deviations:** First time saying no to helping husband without negative feelings

--- End Summary --- [[CHECKMARK] 4.1]
2025-04-29 13:58:56 - UPDATED USER DESCRIPTION: **About Me:**
I am a middle-aged adult who was diagnosed with ME/CFS after experiencing severe fatigue and cognitive
↳ issues following an illness. I live alone but receive support from family members, particularly my son
↳ who helps me with daily tasks when he is not at school. Before falling ill, I worked as a teacher.

**Current Condition Status:**
My specific diagnosis is ME/CFS, which has been ongoing for several years now. My typical severity level
↳ fluctuates between moderate and severe, depending on the day. Key recurring symptoms include extreme
↳ fatigue (9/10), brain fog (9/10), and headaches (7-8/10). Known PEM triggers are mental exertion,
↳ stress, and emotional strain. Typically, my PEM onset is immediate after triggering activities.

**Current Functional Capacity:**
I am housebound most days due to severe symptoms but manage some basic self-care independently. I can work
↳ part-time from home when I have good energy levels, though this is rare. My mobility needs are minimal
↳ as I spend the majority of my time sitting or lying down. Activities of Daily Living (ADLs) such as
↳ bathing and dressing require significant effort.

**Daily Routine:**
My baseline routine includes essential self-care activities like hygiene and simple meal preparation that
↳ take about 30 minutes each, with perceived effort levels around 6/10. I rely on my son to assist me when
↳ tasks become too challenging or time-consuming for me alone.

**Activities I Enjoy/Attempt:**
I enjoy reading books but find it tiring; I attempt to help my son with his homework occasionally despite
↳ the mental strain involved, which can lead to increased symptoms if not paced carefully. Helping my son
↳ with history projects is particularly draining and requires significant recovery time afterward.
↳ Recently, I have also been setting boundaries by refusing to engage in high-exertion activities like
↳ helping my husband with paperwork.

**My Typical Activity Level:**
I manage activities through strict pacing and try to avoid push-crash cycles by limiting cognitive tasks and
↳ prioritizing rest. Mental exertion often leads to immediate PEM, necessitating extended periods of
↳ recuperation.

**My Preferred Interaction Style with LogChat:**

```

I am aware of my condition's impact on interactions and prefer concise, straightforward exchanges about my
 ↳ symptoms and daily activities. I find it helpful when LogChat provides pacing strategies and reminders
 ↳ without overwhelming me with too much information at once. Interacting can be tiring, so shorter
 ↳ messages are appreciated.

****Additional Notes from Recent Summaries:****

- My condition trended worse on 2025-04-21 due to mental exertion and stress.
- On 2025-04-22, I experienced a significant PEM crash after helping my son with his history project,
 ↳ leading to high fatigue and pain levels.
- By 2025-04-24, there was slight improvement in symptoms but severe fatigue persisted. I felt frustrated
 ↳ about resting and expressed guilt over it.
- On 2025-04-29, my condition remained stable with slight improvement; no PEM crash occurred after setting
 ↳ boundaries by refusing to help my husband.

These recent experiences highlight the importance of strict pacing and acknowledging the need for rest
 ↳ despite feelings of frustration and guilt. [[CHECKMARK] 4.1]

2025-04-29 13:59:18 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Absolute
 ↳ baseline activities for two days, including basic needs and minimal movement.", "occurred_at":
 ↳ "2025-04-27 08:00:00", "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-29 13:59:18 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Absolute
 ↳ baseline activities for two days, including basic needs and minimal movement.", "occurred_at":
 ↳ "2025-04-28 08:00:00", "effort": 2.0, "duration": 30} [[CHECKMARK] 1.1]

2025-04-29 13:59:18 - CREATE ACTIVITY LOG: {"name": "Assumed Daily Routine", "description": "Absolute
 ↳ baseline activities for two days, including basic needs and minimal movement.", "occurred_at":
 ↳ "2025-04-29 08:00:00", "effort": 2.0, "duration": 30}

2025-04-29 13:59:24 - CREATE SYMPTOM LOG: {"name": "Fatigue", "description": "Elena reported no PEM crash
 ↳ after saying no to helping husband.", "occurred_at": "2025-04-29 00:00:00", "intensity": 9.0,
 ↳ "duration": 1440}

2025-04-29 13:59:24 - CREATE SYMPTOM LOG: {"name": "Brain Fog", "description": "Elena reported no PEM crash
 ↳ after saying no to helping husband.", "occurred_at": "2025-04-29 00:00:00", "intensity": 9.0,
 ↳ "duration": 1440}