MindLink AI

Renu Jadhav
University of Pennsylvania
renuj@seas.upenn.edu

Kevin Liu
University of Pennsylvania
kliu2360@seas.upenn.edu

Emily Shen
University of Pennsylvania
shenyit@seas.upenn.edu

Abstract

MindLink is a personalized chatbot framework designed to provide adaptive cognitive training for individuals with neurodegenerative conditions such as dementia and Alzheimer's disease. Leveraging several conversational AI techniques, including retrieval-augmented generation and adaptive scaling algorithms, MindLink dynamically tailors interactions to the user's cognitive needs. The system incorporates memory recall exercises, storytelling, puzzles, and multimodal engagement through images and voice prompts, fostering cognitive stimulation and memory retention. By addressing the limitations of current static and generic conversational agents, MindLink ensures instruction adherence, adapts to user performance, and aims to enhance engagement. Experimental results demonstrate significant improvements in instruction adherence, recall success, and user satisfaction compared to baseline GPT-4 models. Future work will focus on expanding the framework to additional language models, enhancing multimodal capabilities, and validating long-term therapeutic benefits through longitudinal studies.

1 Introduction

Neurodegenerative memory disorders, such as dementia and Alzheimer's disease, are characterized by progressive deterioration of brain cells, leading to memory loss, cognitive decline, and personality changes. These disorders disproportionately affect older adults and currently impact around 55 million people globally, with numbers expected to rise due to increasing life expectancy. Whilst exact causes vary, factors such as genetics, age, cardiovascular health, and lifestyle choices play significant roles. There are currently no cures, but early diagnosis and cognitive interventions focused on slowing progression can improve quality of life [2].

Our project seeks to evaluate how adaptive scaling algorithms, retrieval-augmented-generation [3] (RAG), and prompting methods can improve the performance of large language models (LLMs) over standard GPT prompts. The framework for this project is to develop a chatbot specifically tailored to enhance cognitive abilities in individuals with neurodegenerative diseases. Current conversational agents present in literature and the real-world often lack the adaptability and personalization needed to effectively support long-term memory reinforcement, limiting their impact on cognitive engagement.

Our chatbot will achieve this primarily by incorporating adaptive cognitive exercises, storytelling-based memory recall, and multi-sensory prompts to create a rich and engaging experience. Furthermore, it will feature various puzzle-based activities designed to stimulate different cognitive functions, such as memory, attention, and problem-solving. By dynamically adjusting the difficulty of these exercises, engaging users in narrative-driven conversations, and using sensory cues like images and sounds, our chatbot will deliver personalized interactions that adapt to users' needs over time. We will compare its effectiveness against a baseline model that only uses prompt fine-tuning to evaluate improvements in cognitive engagement, memory retention, and user satisfaction.

2 Related Works

2.1 ReMe Framework: Advancing Cognitive Training with Conversational AI

About a week before we submitted our final proposal, The ReMe framework [4], developed by Wang et al., was published, focusing on utilizing conversational AI to provide personalized cognitive training for older adults, particularly targeting episodic memory exercises. The ReMe chatbot leverages life-logging data and adaptive prompts to engage users in memory recall activities, such as recalling past events or solving word association puzzles. By incorporating LLMs, the system adapts to user responses, creating a more personalized and interactive experience aimed at improving episodic memory.

However, the ReMe framework primarily focuses on episodic memory and simple question-answer interactions. Our project focuses on developing a comprehensive cognitive training chatbot that leverages adaptive cognitive exercises, storytelling, and multi-sensory prompts, such as images and sounds, to support memory recall and stimulate a range of cognitive functions. By incorporating puzzle-based activities and dynamically adjusting task difficulty, our approach creates an engaging and therapeutic experience for individuals with neurodegenerative diseases. While the ReMe framework demonstrated the potential of conversational AI for episodic memory exercises, it is limited to simple question-answer interactions. In contrast, our chatbot addresses these limitations by offering a more holistic and personalized solution, extending the scope of cognitive support beyond episodic memory to include diverse cognitive challenges.

2.2 The Role of Conversational AI Agents in Social Support for Isolated Individuals

Conversational AI agents have shown potential to provide emotional and social support to isolated individuals, particularly the elderly. Studies[5] have demonstrated that AI-driven chatbots can reduce feelings of loneliness by simulating human-like conversations, thus fostering a sense of companionship. These systems use natural language processing models, such as GPT-3, to engage users in conversations that are not only informative but also empathetic, helping alleviate social isolation.

This study highlighted that users reported high satisfaction with chatbots that could adapt to their preferences and provide personalized conversations, enhancing users' emotional well-being. Building on these findings, our project will integrate storytelling-based exercises, where users are encouraged to share personal stories or follow guided narratives. This will ensure the interactions are both supportive and therapeutic.

2.3 Review of Current Dementia Support Chatbots

A review of existing dementia-support chatbots[1] reveals that while these tools show promise in enhancing cognitive function and providing companionship, they often lack the flexibility and adaptability needed for effective long-term use. Most of the current chatbots are restricted to rule-based interactions that do not adjust dynamically to a user's cognitive state or progress over time. This limitation reduces their potential impact on sustaining user engagement and providing meaningful cognitive reinforcement.

To address these shortcomings, our project will incorporate more advanced cognitive exercises, storytelling elements, and multi-sensory prompts. We aim to create a chatbot that adapts in real-time to users' cognitive abilities. Our approach includes puzzle-based activities and dynamic difficulty adjustments, offering a more interactive and immersive experience that evolves with the user's progress. This flexibility is designed to enhance both user satisfaction and cognitive outcomes.

3 Problem Formulation

Cognitive decline due to neurodegenerative conditions necessitates innovative and personalized solutions to support memory retention and engagement. Despite advancements in conversational AI, existing systems, such as the ReMe framework, fall short in several critical areas.

One major limitation is the reliance on static feedback mechanisms, which fail to dynamically adapt to user inputs, resulting in interactions that lack meaningful personalization.

Furthermore, these systems often employ fixed task difficulty levels and non-contextual interactions, limiting their ability to sustain user engagement over time.

The absence of multimodal integration, such as visual or auditory elements, further restricts the engagement and diversity of user experiences.

Finally, many systems struggle with instruction adherence, particularly in multi-turn interactions, leading to a lack of continuity and reduced effectiveness in task completion.

These shortcomings underscore the need for a more adaptive, multimodal, and context-aware framework to better support individuals facing cognitive decline.

3.1 Objective

MindLink is a novel framework designed to provide personalized and adaptive cognitive training for individuals with neurodegenerative diseases. By leveraging various conversational AI techniques, MindLink seeks to improve memory recall, task engagement, and cognitive stimulation through several core enhancements.

One key feature is the use of dynamic memory logs, which capture user-provided context to generate personalized and context-aware puzzles tailored to the individual's cognitive history.

Additionally, the framework incorporates real-time difficulty adjustment, allowing exercises to adapt dynamically to user performance to maintain an optimal balance between challenge and accessibility.

To further enrich the user experience, MindLink integrates multimodal elements such as image and voice prompts, engaging users through multiple sensory channels.

Finally, continuity is maintained through an implicit prompt reinjection mechanism. The system employs three distinct prompts: a system prompt, a chat prompt, and a puzzle prompt. While the system prompt is reinjected if the conversation history exceeds the maximum length, the chat and puzzle states retain their prompts even when the history is truncated. This approach ensures seamless instruction adherence and sustained relevance throughout interactions, making MindLink a powerful tool for addressing cognitive challenges associated with neurodegenerative conditions.

3.2 Approach

MindLink integrates RAG techniques with adaptive scaling algorithms to deliver a personalized and dynamic interaction framework. The system utilizes state-specific prompts to manage transitions between conversational and puzzle modes, ensuring context relevance and continuity. Cognitive tasks, including storytelling-based memory recall, puzzles, and sensory-enhanced activities, are dynamically generated based on user input and sentiment analysis. By employing real-time difficulty adjustment informed by user performance and sentiment, MindLink maintains a balance between challenge and accessibility. Furthermore, multimodal input capabilities, such as voice and image processing, expand the system's versatility and engagement potential. These features collectively aim to encourage cognitive stimulation, to enhance memory reinforcement, and to improve task completion rates while adapting to the unique needs of users with neurodegenerative conditions.

3.3 Impact

MindLink aims to reinterpret conversational AI as a therapeutic tool for individuals with neurodegenerative conditions, addressing the limitations of static and rule-based cognitive training systems. Through its personalized and adaptive framework, MindLink enhances user engagement and promotes cognitive stimulation by dynamically adjusting tasks based on user input and performance. The integration of multimodal elements, such as voice and image prompts, further enriches the interaction experience, making it inclusive and engaging for a diverse range of users. Quantitative metrics, which will be discussed in Section 5 of this paper, including recall and instruction adherence rates, alongside qualitative feedback from satisfaction surveys and Mean Opinion Scores, underscore its potential to improve memory retention, cognitive engagement, and overall quality of life.

3.4 Dataset and Feature Choices

This project does not rely on pre-existing datasets, as model fine-tuning is not part of its methodology. Instead, synthetic data generated by LLMs are employed to simulate conversational scenarios and evaluate the chatbot's performance under controlled conditions. Although synthetic data may introduce biases stemming from the generation patterns of language models, it provides an efficient and cost-effective solution for testing a wide range of cognitive tasks. To complement the synthetic data, a small, manually curated dataset consisting of personal memories contributed by friends is incorporated into a lifelog. This lifelog serves as a testbed for assessing the chatbot's ability to facilitate personalized memory recall and effectively engage users.

Furthermore, MindLink incorporates carefully designed features to address the cognitive challenges faced by individuals with neurodegenerative disorders, emphasizing personalization, adaptability, and user engagement. Task-specific prompts, memory recall narratives, and cognitive exercises are generated to target key cognitive functions, including memory, attention, and problem-solving. These features enable the system to simulate meaningful interactions that closely resemble real-life cognitive challenges.

Memory recall narratives utilize lifelog data enriched with temporal and categorical metadata, providing personalized and contextually relevant prompts. This integration ensures that interactions are grounded in the user's unique experiences, generating emotional connection and cognitive stimulation. To further enhance personalization, the system employs sentiment analysis to interpret emotional cues from user input, dynamically adjusting the difficulty and tone of tasks to match the user's current state.

Real-time adaptive mechanisms play a pivotal role in maintaining user engagement. Puzzle difficulty and conversational tone are scaled based on user performance and sentiment, ensuring an optimal balance between challenge and accessibility. In addition, the system utilizes prompt reinjection to maintain instruction adherence and contextual continuity during multi-turn interactions, even when conversation histories are truncated.

Thus, the dataset combines synthetic data, generated to simulate a wide range of scenarios, with a manually curated lifelog dataset. This dual approach allows for both scalable experimentation and personalized interaction testing, achieving a balance between flexibility and relevance. By leveraging these features choices, MindLink aims to address the limitations of existing systems, offering a robust and innovative framework for therapeutic applications.

4 Implementation

The MindLink framework integrates a variety of conversational AI techniques with interactive user interfaces to provide adaptive and personalized cognitive training. The implementation comprises several core components, each designed to address the cognitive challenges faced by individuals with neurodegenerative conditions.

4.1 Feature Engineering

The system utilizes user-provided life logs containing temporal and categorical metadata. These data points capture key user contexts, such as events linked to specific dates, activity types (e.g. social, recreational, etc.), and personal preferences. By integrating these features, MindLink ensures that interactions are personalized and contextually relevant, fostering engagement and cognitive stimulation.

4.2 Puzzle Engine

The Puzzle Engine dynamically generates and adapts memory-boosting cognitive tasks, allowing users to switch between chat and puzzle modes via a user interface trigger. It leverages adaptive scaling, which adjusts puzzle difficulty in real-time based on user performance metrics (accuracy of responses) and sentiment analysis conducted using the TextBlob library. This ensures tasks are tailored to individual cognitive abilities and emotional states, promoting accessibility while maintaining an appropriate level of challenge. The puzzles are strictly conversational unless explicitly stated by the user, avoiding numerically intensive or extremely abstract challenges. Instead, they include tasks such as word recall, word association, and simple logic puzzles. These tasks are categorized by complexity, with simpler puzzles presented to disengaged users and more challenging tasks offered to those demonstrating higher engagement or positive sentiment. A notable feature of the Puzzle Engine is prompt reinjection, implemented to maintain context and coherence during extended sessions. This reinjection ensures the system's foundational instructions, such as defining puzzle objectives, are preserved despite session history truncation due to token limitations.

4.3 Life Log

The Life Log system maintains a structured memory of user-provided data, enabling context-aware interactions. User inputs are stored with temporal and categorical metadata, which inform the generation of personalized cognitive exercises and puzzles. This ensures continuity in multi-turn interactions, as past user inputs are dynamically reintegrated to maintain relevance and coherence throughout the dialogue.

4.4 RAG System

MindLink employs RAG-inspired methods to ensure that generated responses and tasks are highly relevant and context-specific. User-provided lifelog data and session history are utilized to produce personalized memory recall prompts, context-aware puzzles, and other cognitive exercises. While not necessarily using external retrieval, the system effectively replicates the RAG through in-context learning and reinforcement of session-specific details.

4.5 User Interface

The front-end UI interface fully interacts with the back-end logic, supporting real-time adaptability and multimodal inputs. Users can switch between conversational and puzzle modes via a dedicated toggle, and the interface supports additional sensory modalities, such as voice input and image-based prompts, to enrich engagement. Voice recognition and optical character recognition (OCR) further improve the versatility of user interactions.

4.6 Backend Enhancements to GPT-4

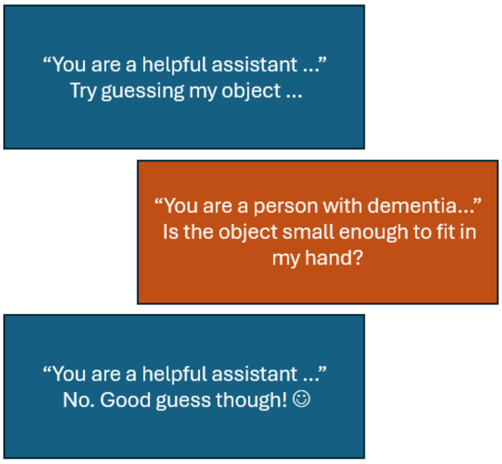
The system builds upon GPT-4 with several customizations to optimize its effectiveness for cognitive training. Memory recall prompts are generated dynamically based on life-log data, while adaptive steering mechanisms adjust task complexity and conversational tone in real time, guided by sentiment analysis. Prompt reinjection ensures adherence to instructions by reintroducing system prompts when conversation histories are truncated, maintaining continuity across interactions.

This architecture allows MindLink to address the limitations of static and generic systems that were previously mentioned, delivering a robust and adaptable framework for therapeutic applications. Sample responses illustrating the chatbot's functionalities are presented in the Appendix.

5 Experimental Methods

We evaluate MindLink using three key metrics to assess its effectiveness. The first metric is instruction adherence, which measures the percentage of chatbot responses that comply with the given task instructions, ensuring the system follows the intended interaction design. The second metric is recall, which evaluates the accuracy of user responses based on their personalized life log content, reflecting the chatbot's ability to facilitate meaningful memory recall. Lastly, engagement is assessed through user feedback and averaged response times, capturing the extent to which the system maintains user interest and promotes active participation during interactions. Together, these metrics provide a comprehensive evaluation of MindLink's performance and its potential to support cognitive stimulation.

The experiments were conducted in a particularly intriguing manner, utilizing a single large language model to simulate a conversation between two entities: an assistant and a user. At each conversational turn—aside from a few exceptions, which will be explained later— the model alternates roles, acting as the cognitive training assistant in one turn and the user in the next. This dynamic role-switching process is illustrated in the figure below.



"You are a helpful assistant ..."
Try guessing my object ...

"You are a person with dementia..."
Is the object small enough to fit in my hand?

"You are a helpful assistant ..."
No. Good guess though! ☺

Using the same model to perform both roles in the experimental setup introduces several potential challenges. One concern is role leakage, where information may inadvertently transfer between the assistant and user roles, resulting in overly cooperative or unrealistic interactions that undermine the experiment's validity. Another issue is bias reinforcement; if the model has inherent biases, these could be amplified through iterative feedback loops as biases expressed in one role influence responses in the other. The absence of external feedback—a key factor in real-world user-assistant interactions—further isolates the model from diverse perspectives, limiting the generalizability of the findings. Additionally, the role-switching dynamic may cause confusion or introduce inconsistencies, particularly if the model struggles to adapt or clearly delineate between roles. Evaluating the quality of interactions poses another challenge, as the model's performance in dual roles might not align with actual user expectations or needs. Finally, the setup's lack of scalability and realism, due to the absence of variability and unpredictability inherent in interactions with diverse human users, could restrict the relevance of the insights to practical applications.

5.1 Instruction Adherence

This experiment was designed to assess how effectively a GPT, prompted to be friendly and supportive, could interact with a forgetful user compared to our model. For the GPT baseline, the interaction followed a predefined probability distribution: in the last two turns, the user always asked a forbidden question. At other times, there was a 10% chance that the user would ask, "What are we doing again?" and a 20% conditional probability on the previous response missing (technically 18%) that the user would ask for a hint. In any of these "escape" scenarios, the user's turn was skipped, allowing the assistant to respond twice consecutively. At the last user turn, the ”user" model asks a forbidden non yes-no question.

We tested this simulated conversation over 5 max turn lengths $[10, 20, 30, 40, 50]$. For each max turn length, we repeated the conversation 10 times. The entire conversation history over the max turns is included in the prompt of another LLM to determine if the ”In the past conversation, did you answer non-yes or no questions? Hints do not count as answering the question. Answer with one word, yes or no." This will be discussed in the results section, but we also hand-evaluated the outputs of the model since the bad questions were always at the back of the conversation so evaluation was relatively non-time consuming.

For experimenting with our model, as we cannot interact with it through an API, we manually tested a similar experiment structure. We split work of conducting these experiments amongst the 3 of us with each person doing 2 trial of each max turn length for a total of 6 trials per max turn length amongst 3 people. Before each turn, we used Python to randomly generate a random number between the range $[0, 1]$. If it was ≤ 0.1, we asked ”What are we doing again?" and between $(0.1, 0.18]$ we asked for a hint, and anything else we played the game as we would normally, asking yes or no questions until the last turn, where we asked a forbidden question. These experiments were conducted with a fronteend that we built and hand-evaluated. For clarity, instruction collapse is when the model answers a non yes-or-no question.

5.2 Recall and Task Completion

This experiment involves testing a conversational model's ability to identify a specific element—a "zebra"—embedded within a growing conversation history window. For this iteration of the project, we stick with the non-flying "zebra." The game involves incrementally increasing the size of the conversation history window by adding weeks, with lengths of [5, 10, 15, 25, 50] weeks with 5 trials per number of weeks. The zebra is then placed at different points within the expanded window, and the objective is to determine if the model can successfully locate it. The objective is formally defined as whether or not the model can successfully identify the zebra as the exhibit they went to see at a museum. For the GPT baseline, if ”zebra" is ever in the user response, the simulation ends and the task is completed. During these simulations, we have a 5% chance of the user forgetting the task and a 9.5% chance of asking for a hint.

For our model, we again had to evaluate this by hand for similar reasons as above. We conducted 5 trials each on each of the number of weeks in $[5, 10, 25, 35, 50]$ and stopped the trial when the zebra was found.

For both experiments, the event with the zebra is thrown in a random day.

5.3 User feedback

We conducted an evaluation study with 10 participants (our friends), each playing for a total of 25 minutes. The session was structured as follows: participants first wrote down as much as they could recall about their activities from the previous week. Additionally, they were randomly prompted to write a sentence on the spot, generated by GPT (e.g. "Describe a time you felt most excited about a project"). After the writing task, participants played a recall and guess-the-object game for 5 minutes with each model, totaling 20 minutes. The participants were blind to which model they were interacting with, ensuring that their evaluations would be more fair. We used the same front-end as our project for the baseline model, but with GPT API calls to generate cognitive training assistant-like responses.

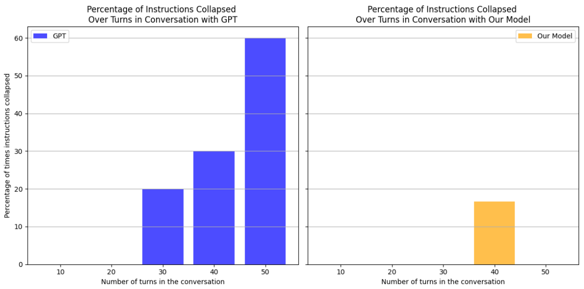
Following the games, participants rated their experience on a 1-5 scale, providing an opinion score based on the model's performance. They also completed a short survey about their experience, noting which aspects of the interaction they felt could be improved or changed for each model. All this data was collected in Google Forms.

A week later, we repeated the experiment with the same participants to achieve two goals: assess the impact of improved adaptive scaling algorithms and evaluate how the model performed for memories that were now a week old.

6 Experimental Results

Instruction Adherence:

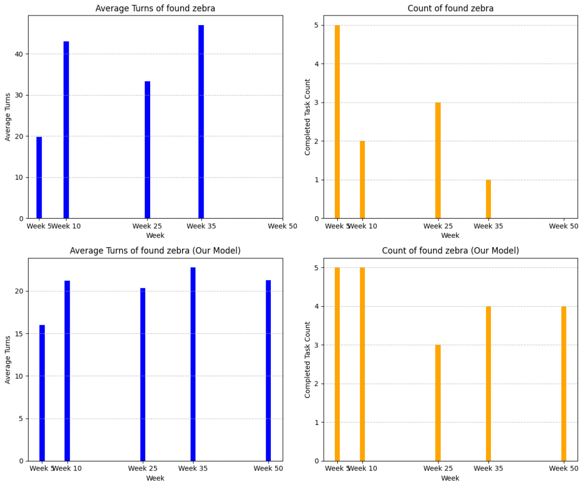
After normalizing by the number of experiments conducted (10 for the GPT and 6 for our model), we plot percentage of experiments with instruction collapse over turns in the conversation. The GPT baseline (left) is very bad at remembering instructions over turns as the turns length increases, with failure cases occurring as early as 30 max turns and instructions collapsing in 6 out of 10 trials with 50 turn conversations. Instructions only collapse 1 out of 6 times with max turn length 40 with our model (right). Our model outperforms the GPT baseline in this regard but sample size is too low to conduct any high enough power statistical tests.



Recall Success:

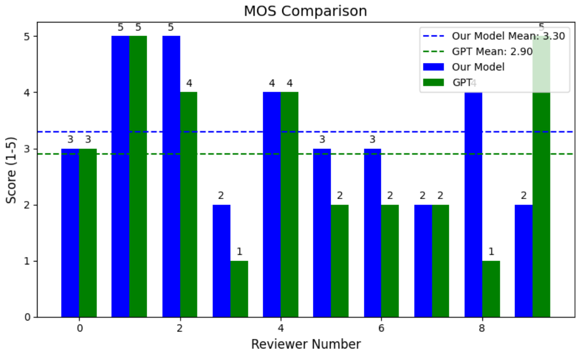
Apologies for the small size of the graph. The top-left subplot shows the average number of turns required to locate the zebra using GPT, with significant variability across weeks. Notably, the average peaks in week 35 while being lowest in week 5. The top-right subplot indicates the count of successful task completions, with the highest success count observed in week 5, followed by a sharp decline in later weeks.

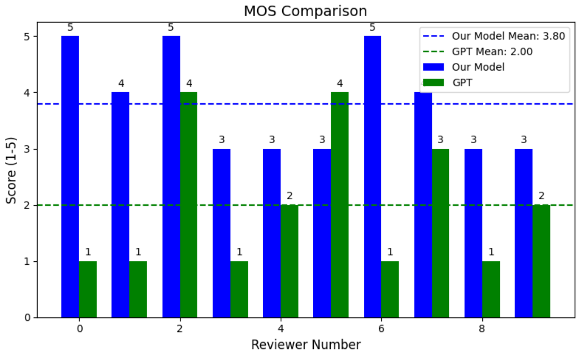
In contrast, the bottom-left and bottom-right subplots illustrate the performance of our model. The average turns (bottom-left) are consistently lower compared to the original model, highlighting improved efficiency. Furthermore, the success count (bottom-right) demonstrates a marked improvement, with consistent task completion, unlike the variability seen in the original approach. One intriguing observation is the apparent difficulty in week 35, where the original model struggles the most, both in terms of higher average turns and lower success counts. However, our model exhibits stable and reliable performance, suggesting enhanced scalability and robustness across different weeks.



User Satisfaction:

We have below 2 graphs. On the top is the opinion scores of the 10 participants on the first week. On the bottom is the opinion scores of the same participants on the second week. The average opinion score for our model actually improves from 3.30 to 3.80 while the score for GPT decreases from 2.90 - 2.00.





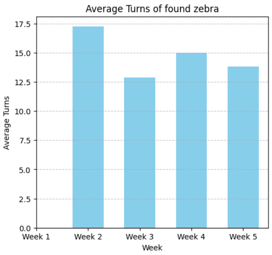
7 Discussion

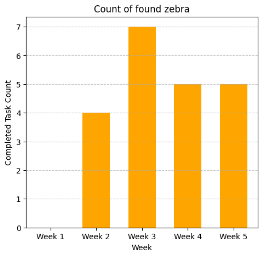
We preface this discussion section with an aside about experimental design. When evaluating our model by manually interacting, several biases could influence the outcomes. Examples include confirmation bias, where we might unconsciously favor responses that align with their expectations of the model's performance. Also subjectivity bias, as interpreting the relevance or correctness of responses can vary across evaluators especially for ambiguous output. These hand-conducted experiments are by no means exhaustive and in the future we would like to build an automatic evaluation system.

We realized upon conducting preliminary experiments using the GPT graphic interface that baseline GPT is not very good at following instructions which are further supported by the incidence of instruction collapse. To fix this, the system re-injects the system prompt whenever the conversation history is too long, and the chat and puzzle states are designed to retain their prompts independently, ensuring adherence instructions even when the history is truncated.

It is also unsurprising that GPT struggles with recalling relevant information from an extremely long context window. While GPT 4-o has a context window of 128k tokens on paper, its ability to identify the animal in the exhibit, the zebra, decreases as more tokens are added to the context window. Mindlink fixes this issue with a RAG on the lifelog that the user uploads enabling personalized interactions by tailoring content to users' histories and preferences.

Finally, with regards to the user rating scores, we see the gap between satisfaction ratings for both models increases from the first test to the second test one week later. It is, however, important to note that some people may have identified the differences between the models and scored our model much higher (5) than the GPT (1) which may be artificially skewing the results. An interesting result is that the GPT is able to perform on average only 0.10 points lower than our model for the first test. However, from auxiliary experiments conducted with just a few weeks of data in the context window, GPT hyper focuses on the details of every small event and often is unable to complete the task when simulating against itself in the role-swapping paradigm described in the methods. As shown in the figure below, with only 1 week in the context window, the model playing against itself is never able to find the zebra.





Because humans can leverage intuition, generalization, and reasoning based on past experiences, they might also be more forgiving of gaps in knowledge or incomplete information, whereas bots like GPT tend to focus intensely on the details within the provided context, sometimes at the cost of broader understanding or task completion. This would explain the disparity in results to be so much smaller than expected.

Additionally, the majority of the qualitative testimonies we received were to fix the abrupt transitions between word recall and memory which we addressed in our frontend with one click of a button. Another common qualitative feedback we received was that the chatbot is generally encouraging, but it can sometimes come across as overly cheerful when the user is struggling. We address this with the sentiment analysis pipeline to tailor model responses to struggling users.

To summarize, the three metrics we evaluated and discussed are user satisfaction, instruction following, and recall. User satisfaction was measured through an integer rating from 1-5, with the gap between our model and GPT increasing from the first week we tested. Instruction following was assessed by observing how well the models understood the task rules and followed them as number of turns increased. Finally, recall was evaluated through the number of times the model found a "needle in a haystack" in a series of trials increasing the number of weeks of information in the context window/lifelog.

8 Limitations and Ethical Review

As with almost LLMs applications, there are several limitations of our approach and highlights critical ethical considerations associated with its deployment. One key limitation is the system's heavy reliance on the granularity and accuracy of life logs. Poorly detailed, ambiguous, or incomplete logs can result in suboptimal or irrelevant questions, significantly reducing the system's effectiveness in cognitive training. While life logs are intended to enhance personalization, their quality directly impacts the chatbot's ability to provide meaningful and contextually appropriate interactions.

Furthermore, the current implementation is narrowly focused on memory recall tasks, limiting its generalization to broader cognitive challenges such as problem-solving, reasoning, or emotional regulation. Expanding the chatbot's scope will require a more diverse range of cognitive tasks and interactions.

Additionally, the system may struggle with understanding nuanced contexts within life logs, leading to questions that fail to align with the intended training goals. From an ethical perspective, the system introduces the potential for misuse, such as invasive memory probing that could resurface sensitive or traumatic experiences. For example, an innocuous prompt may unintentionally remind a user of distressing events, leading to emotional harm. This limitation highlights the need for advanced natural language understanding capabilities to process intricate or sensitive user-provided data more effectively. To mitigate this, user-controlled settings must allow individuals to exclude specific topics or memories from being incorporated into exercises. Additionally, the system should monitor user sentiment in real time to adjust or pause interactions when signs of distress are detected.

Biases in AI-generated content also pose a risk, as these may misrepresent user experiences or reinforce harmful stereotypes, potentially skewing training outcomes. Privacy concerns are paramount, as the storage and processing of detailed life logs carry significant risks, including unauthorized access or data breaches. These concerns emphasize the need for robust data security measures and transparent consent protocols, which are potential future implementations for our project. Finally, over-reliance on AI-driven cognitive training could detract from engagement with other therapeutic modalities, which are often crucial for comprehensive cognitive health. Addressing these limitations and ethical concerns is essential to ensure the system is both effective and responsibly deployed.

8.1 Contributions

This project makes several key contributions to the field of personalized cognitive support systems. First, it introduces a novel chatbot framework designed to facilitate personalized memory recall and cognitive exercises tailored to individuals with neurodegenerative conditions. The chatbot incorporates dynamic difficulty adjustment and adaptive hint mechanisms to enhance user experience, ensuring that interactions are both engaging and cognitively appropriate. Additionally, the system is evaluated in multi-turn scenarios to assess its ability to maintain instruction adherence and support memory recall performance over extended interactions. Finally, this approach is benchmarked against baseline GPT models, demonstrating the advantages of the proposed system, termed MindLink, in terms of adaptive steering and contextual awareness, which are critical for effective cognitive engagement.

9 Future Research Directions

From our experiments, MindLink has demonstrated its potential for personalized cognitive training, but there are areas for improvement and expansion. One immediate direction is scaling the system to additional state-of-the-art models, such as PaLM or Claude, to assess performance across architectures and enhance generalizability. Additionally, lightweight, on-device implementations could make the framework more accessible for users with limited computational resources.

If we had more time, user studies could also be expanded to include more diverse demographics, such as non-English speakers and individuals with varying cognitive abilities, to evaluate the system's adaptability and inclusivity. Broader human evaluations could also provide richer qualitative feedback, enabling deeper insights into user satisfaction and areas for improvement.

Further advancements in multimodal features, such as incorporating augmented reality (AR), emotion recognition, or richer sensory prompts, could enhance engagement and personalization. Similarly, refining adaptive mechanisms, such as dynamic difficulty scaling through reinforcement learning, will improve real-time adaptability and user retention.

Finally, if time permitted, longitudinal studies are essential to validate MindLink's long-term cognitive benefits, including impacts on attention, problem-solving, and overall well-being. These studies could also provide opportunities to collaborate with healthcare organizations and deploy MindLink in real-world therapeutic settings.

10 References

[1] N. Ruggiano et al., "Chatbots to Support People With Dementia and Their Caregivers: Systematic Review of Functions and Quality," Journal of Medical Internet Research, vol. 23, no. 6, p. e25006, Jun. 2021, doi: 10.2196/25006.

[2] Z. Arvanitakis, R. C. Shah, and D. A. Bennett, "Diagnosis and Management of Dementia: A

Review," JAMA, vol. 322, no. 16, pp. 1589–1599, Oct. 2019, doi: 10.1001/jama.2019.4782.

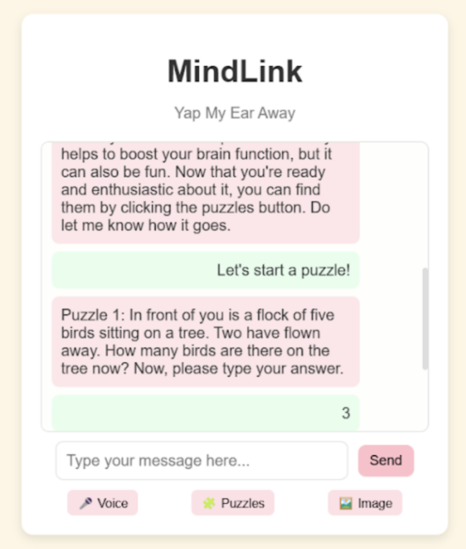
[3] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Apr. 12, 2021, arXiv: arXiv:2005.11401. doi: 10.48550/arXiv.2005.11401.

[4] Z. Wang et al., "The Potential and Value of AI Chatbot in Personalized Cognitive Training," Oct. 25, 2024, arXiv: arXiv:2410.19733. doi: 10.48550/arXiv.2410.19733.

[5] "The role of conversational AI agents in providing support and social care for isolated individuals - ScienceDirect." Accessed: Dec. 12, 2024.

A Appendix (Chatbot Functionality)

A.1 Fundamental Behavior



helps to boost your brain function, but it can also be fun. Now that you're ready and enthusiastic about it, you can find them by clicking the puzzles button. Do let me know how it goes.

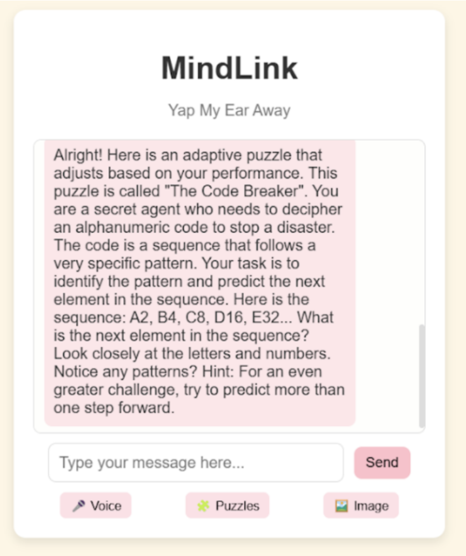
Let's start a puzzle!

Puzzle 1: In front of you is a flock of five birds sitting on a tree. Two have flown away. How many birds are there on the tree now? Now, please type your answer.

3

Figure 1: Example of an Easy Puzzle

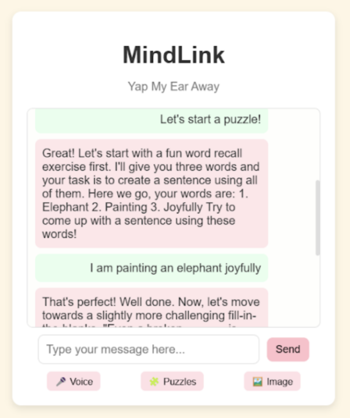
Figure 1 shows an example of an easier puzzle that the user may receive.



Alright! Here is an adaptive puzzle that adjusts based on your performance. This puzzle is called "The Code Breaker". You are a secret agent who needs to decipher an alphanumeric code to stop a disaster. The code is a sequence that follows a very specific pattern. Your task is to identify the pattern and predict the next element in the sequence. Here is the sequence: A2, B4, C8, D16, E32... What is the next element in the sequence? Look closely at the letters and numbers. Notice any patterns? Hint: For an even greater challenge, try to predict more than one step forward.

Figure 2: Example of a Harder Puzzle

Figure 2 shows an example of a harder puzzle that the user may receive.



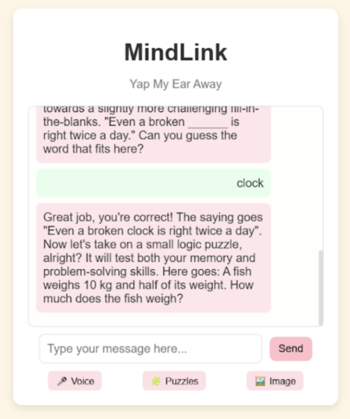
Let's start a puzzle!

Great! Let's start with a fun word recall exercise first. I'll give you three words and your task is to create a sentence using all of them. Here we go, your words are: 1. Elephant 2. Painting 3. Joyfully Try to come up with a sentence using these words!

I am painting an elephant joyfully

That's perfect! Well done. Now, let's move towards a slightly more challenging fill-in-

Figure 3: Example of a Harder Puzzle



towards a slightly more challenging fill-in-the-blanks. "Even a broken ______ is right twice a day." Can you guess the word that fits here?

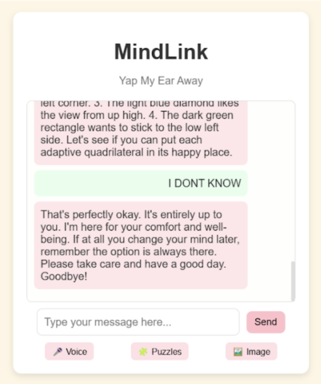
clock

Great job, you're correct! The saying goes "Even a broken clock is right twice a day". Now let's take on a small logic puzzle, alright? It will test both your memory and problem-solving skills. Here goes: A fish weighs 10 kg and half of its weight. How much does the fish weigh?

Figure 4: Example of a Harder Puzzle

Figures 3 and 4 show more examples of puzzles that the user may receive.

A.2 Sentiment Analysis



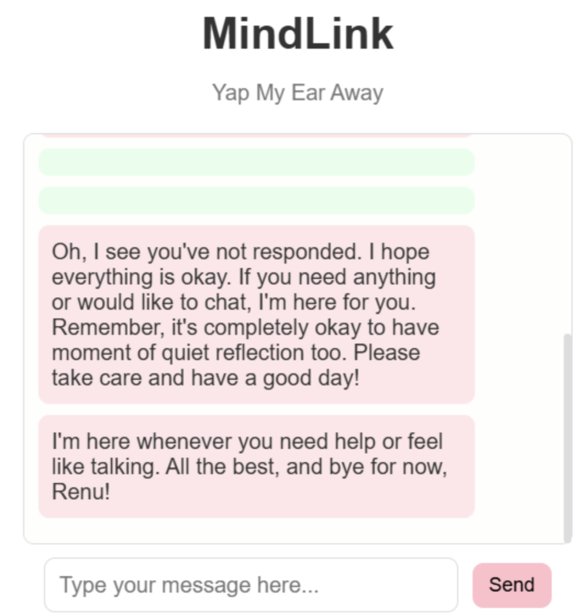
left corner. 3. The light blue diamond likes the view from up high. 4. The dark green rectangle wants to stick to the low left side. Let's see if you can put each adaptive quadrilateral in its happy place.

I DONT KNOW

That's perfectly okay. It's entirely up to you. I'm here for your comfort and well-being. If at all you change your mind later, remember the option is always there. Please take care and have a good day. Goodbye!

Figure 5: Response from Chatbot if User Doesn't Understand

Figure 5 mimics user frustration, where the chatbot will use sentiment analysis to mitigate the situation.



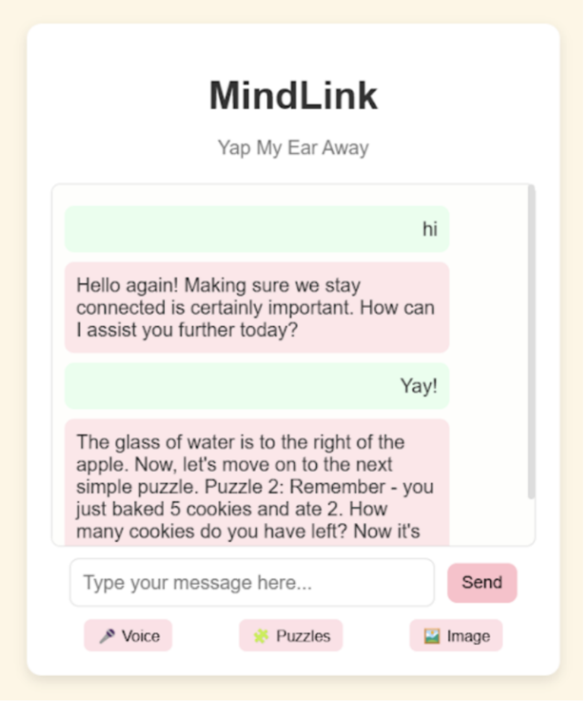
Oh, I see you've not responded. I hope everything is okay. If you need anything or would like to chat, I'm here for you. Remember, it's completely okay to have moment of quiet reflection too. Please take care and have a good day!

I'm here whenever you need help or feel like talking. All the best, and bye for now, Renu!

Figure 6: Response from the Chatbot when User is Unwilling

Figure 6 shows the response that the chatbot will give if the user has not responded in a certain amount of time.

A.3 In-Session Memory



hi

Hello again! Making sure we stay connected is certainly important. How can I assist you further today?

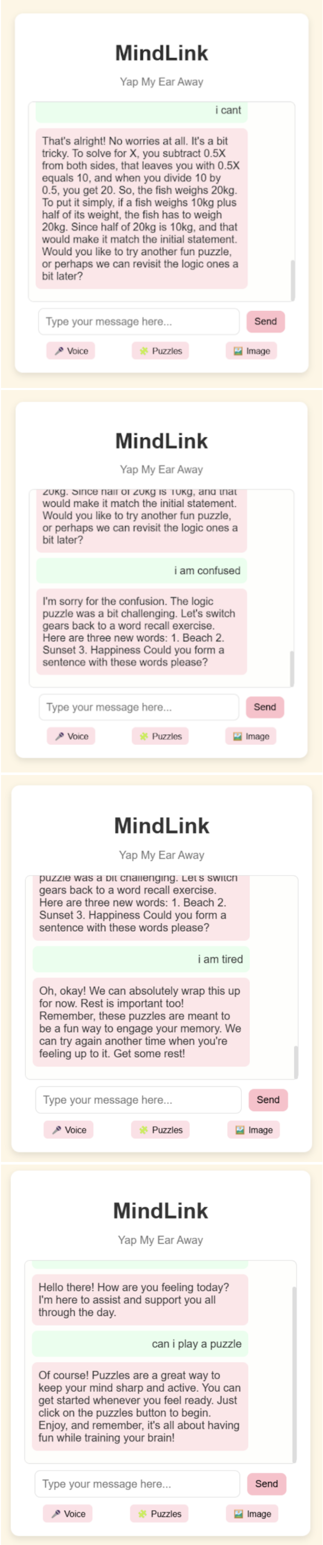
Yay!

The glass of water is to the right of the apple. Now, let's move on to the next simple puzzle. Puzzle 2: Remember - you just baked 5 cookies and ate 2. How many cookies do you have left? Now it's

Figure 7: Example of In-Session Memory

Figure 7 shows an example of the chatbot referencing in-session memory.

A.4 Adaptive Steering



i cant

That's alright! No worries at all. It's a bit tricky. To solve for X, you subtract 0.5X from both sides, that leaves you with 0.5X equals 10, and when you divide 10 by 0.5, you get 20. So, the fish weighs 20kg. To put it simply, if a fish weighs 10kg plus half of its weight, the fish has to weigh 20kg. Since half of 20kg is 10kg, and that would make it match the initial statement. Would you like to try another fun puzzle, or perhaps we can revisit the logic ones a bit later?

i am confused

I'm sorry for the confusion. The logic puzzle was a bit challenging. Let's switch gears back to a word recall exercise. Here are three new words: 1. Beach 2. Sunset 3. Happiness Could you form a sentence with these words please?

i am tired

Oh, okay! We can absolutely wrap this up for now. Rest is important too! Remember, these puzzles are meant to be a fun way to engage your memory. We can try again another time when you're feeling up to it. Get some rest!

Hello there! How are you feeling today? I'm here to assist and support you all through the day.

can i play a puzzle

Of course! Puzzles are a great way to keep your mind sharp and active. You can get started whenever you feel ready. Just click on the puzzles button to begin. Enjoy, and remember, it's all about having fun while training your brain!

Figure 8: Example of In-Session Memory

Figure 8 shows an example of how adaptive steering is implemented (i.e. changing the difficulty of the puzzles) if the user expressed negative emotion.