

# AI INTERFACES FOR AUGMENTING EPISODIC MEMORY

by

Wazeer Deen Zulfikar

B.E (Hons), Birla Institute of Technology and Sciences (2019)

Submitted to the Program in Media Arts and Sciences, School of Architecture  
and Planning, in partial fulfillment of the requirements for the degree of

Master of Science in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

© Wazeer Deen Zulfikar 2024. All rights reserved.

The author hereby grants to MIT a nonexclusive, worldwide, irrevocable,  
royalty-free license to exercise any and all rights under copyright, including to  
reproduce, preserve, distribute and publicly display copies of the thesis, or  
release the thesis under an open-access license.

## **AUTHORED BY**

---

Wazeer Deen Zulfikar  
Program in Media Arts and Sciences  
May 17, 2024

## **CERTIFIED BY**

---

Dr. Pattie Maes  
Professor of Media Arts and Sciences  
Thesis Advisor

## **ACCEPTED BY**

---

Dr. Joe Paradiso  
Academic Head  
Program in Media Arts and Sciences

# AI INTERFACES FOR AUGMENTING EPISODIC MEMORY

by

Wazeer Deen Zulfikar

Submitted to the Program in Media Arts and Sciences, School of Architecture  
and Planning

on May 17, 2024, in partial fulfillment of the  
requirements for the degree of

Master of Science in Media Arts and Sciences

## Abstract

Episodic memory, the memory of personal experiences, is a core component of human cognition. It functions within the neural substrate to store progress towards personal goals. Thus, it influences human behavior by enriching social interactions, forming a personal narrative, and facilitating personal growth. With the rise of challenges such as poor sleep, aging and dementia, and fragmented attention, people experience difficulties with episodic memory retrieval. These difficulties range from momentary lapses such as forgetting previous interactions during conversations, to recalling multiple events during reminiscing and decision-making.

In this work, we explore artificially intelligent (AI) systems that augment episodic memory by enabling people to interact with their memories effectively. We design, develop, and evaluate two systems: (i) MEMORO, a wearable audio-based memory assistant that presents concise suggestions in real-time while minimizing disruption to the user's primary task, and (ii) RESONANCE, a web-based reflective memory assistant that offers actionable suggestions to help users savor their past, present, and future experiences for mental health benefits. By conducting an in-person user study for MEMORO and a longitudinal online user study for RESONANCE, we investigate the effects of these systems on users, measure their technical efficacy, and gather feedback on user experiences. Recent advances in artificial intelligence offer novel opportunities to enhance episodic memory. Therefore, exploring interfaces that seamlessly integrate with human behavior is crucial to ensure that AI-based systems enrich everyday experiences.

Thesis advisor: Pattie Maes

# AI INTERFACES FOR AUGMENTING EPISODIC MEMORY

by

Wazeer Deen Zulfikar

This thesis has been reviewed and approved by the following committee members:

## **THESIS ADVISOR**

---

Dr. Pattie Maes  
Germeshausen Professor of Media Arts and Sciences  
Massachusetts Institute of Technology

## **THESIS READER**

---

Dr. Rosalind Picard  
Professor of Media Arts and Sciences  
Massachusetts Institute of Technology

## **THESIS READER**

---

Dr. Satrajit Ghosh  
Director of Open Data in Neuroscience Initiative and Principal Research  
Scientist  
Massachusetts Institute of Technology

## ACKNOWLEDGMENTS

I would like to express my deepest gratitude to a wide variety of people ranging from colleagues, friends, and family who were part of the journey. My gratitude is towards their invaluable insights, support, guidance and sometimes just being there.

Firstly, I would like to thank my advisor, Dr. Pattie Maes, for the immense level of support given to me throughout this work. Providing this opportunity, and then giving me the room to learn, build, and grow as a researcher was a powerful catalyst for me to be able to express myself through my work and contribute to the scientific community. In particular, being flexible and allowing me to embrace my unique timeline and experience at the lab was crucial.

I would like to especially appreciate the collaborators of my work in this thesis, Samantha Chan, Treyden Chiaravalloti, Jocelyn Shen and Cayden Pierce for being wonderful people to work with, ideate, and build. The projects would not have been possible without them and they helped me further understand and cherish the value of collaboration over competition among peers in the research community.

The members of Fluid Interfaces were the backbone upon which I felt comfortable and inspired to be creative and experiment with ideas. Thank you to Valdemar Danry and Joanne Leong for sharing office space with me, having interesting and insightful conversations on and off-topic, and supporting me in the difficult trenches of a project timeline. It made the office space feel much more friendly and encouraging. The help from Nathan Whitmore on statistics and neuroscience, Pat Pataranutaporn on inspirational storytelling and beautiful design, Angela Vujic on innovative ideas, Cathy Fang on brainstorming and support was instrumental! I would also like to thank Nataliya Kosmyna, Ruby Liu, Arnav Kapur, Natasha Maniar, Dunya Baradari, Nishat Protyasha, Camilo Rojas and Abhinandan Jain, and I was lucky to have these amazing lab members.

I extend my appreciation to my great friends in Cambridge who supported me in every way possible both emotionally and intellectually, and further allowed me to support them, which made me feel more connected. These very important people are Ana Trapero, Kushagra Tiwary, Aastha Shah, Jason Hou, Charles Boury, Meghana Potta, Lancelot Blanchard, Benedetta D'alia, Kevin Dunnell, Akshay Dharmavaram and Nikhil Lal. Finally, I would like to thank my family for their love and support – Amma, Atha, and Shafiqah – without you I would not be the person who I am today.

# CONTENTS

1	Introduction	10
1.1	Contributions	13
2	Background and Related Work	14
2.1	Role of Episodic Memory in Human Wellbeing	14
2.2	Wearable memory augmentation systems	14
2.3	Digital Journaling with AI	15
3	Memoro: Real-time Memory Augmentation	17
3.1	Introduction	17
3.1.1	Minimally Disruptive Interfaces	17
3.2	System Design	19
3.2.1	Memory Encoder	20
3.2.2	Retrieval Agent	21
3.2.3	Query Agent	23
3.3	User Study	24
3.3.1	Participants	24
3.3.2	Conditions	24
3.3.3	Procedure	25
3.4	Results	26
3.4.1	Technical Evaluation	26
3.4.2	Conversation Quality, Task Performance, and Task Load between Conditions	28
3.4.3	User Perceptions and Experience with Memoro	31
4	Resonance: Reflective Memory Augmentation	36
4.1	Introduction	36
4.1.1	Holistic Journal Interfaces	36
4.2	System Design	37
4.2.1	Inputting Memories and Imaginations	37
4.2.2	Suggestion Generator	38
4.3	User Study	40
4.3.1	Participants	40
4.3.2	Conditions	40
4.3.3	Procedure	40
4.4	Results	42

4.4.1	Changes in Mental Health . . . . .	42
4.4.2	Perceptions of Interaction with the AI System . . . . .	44
4.4.3	General Feedback . . . . .	47
5	Discussion . . . . .	51
5.1	Minimally Disruptive Interfaces for Real-Time Memory Augmentation . . . . .	51
5.1.1	Integrating LLMs in Wearable Memory Augmentation . . . . .	51
5.1.2	Impact of Using Memoro in a Primary Task . . . . .	52
5.1.3	Usability, Preferences, and Experiences . . . . .	52
5.2	Holistic Interfaces for Reflective Memory Augmentation . . . . .	54
5.2.1	Reduction in Current Depression by Using Resonance Daily . . . . .	54
5.2.2	Perceptions about the AI Suggestions and Imaginations . . . . .	54
5.2.3	Preferences and Experiences . . . . .	55
6	Limitations . . . . .	57
6.0.1	Technical Aspects . . . . .	57
6.0.2	Study Design and Population . . . . .	58
6.0.3	Social Implications: Privacy and Social Acceptability . . . . .	59
7	Ethical Considerations . . . . .	61
8	Conclusion . . . . .	63
A	Appendix . . . . .	64
A.1	Example Queries . . . . .	64
A.1.1	Baseline . . . . .	64
A.1.2	Query Mode . . . . .	64
A.1.3	Queryless Mode . . . . .	65
A.2	Prompt Templates . . . . .	66
A.2.1	Contextual Query . . . . .	66
A.2.2	Concise Suggestions . . . . .	66
A.2.3	Queryless Search . . . . .	66
A.3	Questionnaires . . . . .	67
A.3.1	User experiences and Perception . . . . .	67
A.3.2	Conversation Quality . . . . .	67
A.3.3	Task Performance/Recall Ability . . . . .	68
A.4	Questions for Seed Memories . . . . .	68
A.5	Prompt to Users for Episodic Memory Journalling . . . . .	69
A.5.1	Memory Prompt . . . . .	69
A.5.2	Imagination Prompt . . . . .	69

# LIST OF FIGURES

Figure 1	Minimally disruptive interfaces with wearable AI assistants	12
Figure 2	Holistic interfaces for being more connected during journaling with AI . . . . .	12
Figure 3	Interactions using Memoro’s Query and Queryless Modes	18
Figure 4	Closed loop system architecture has a memory encoder that is continuously updated using text-to-speech. . . . .	20
Figure 5	Workflow of the components of Memoro: memory encoder, retrieval agent, and query agent. . . . .	21
Figure 6	Example interactions by P3 and P19 show the Query Mode and the Queryless mode for the same memory respectively. . . . .	29
Figure 7	Task Performance (a) Confidence in recalling, (b) Relevance of recalled information. ***: $p < .001$ . . . . .	30
Figure 8	Task load: (a) Perceived difficulty in recalling, and (b) Raw NASA TLX scores. ***: $p < .001$ . . . . .	30
Figure 9	System Usability Scale (SUS) scores for the different conditions. *: $p < .05$ . . . . .	31
Figure 10	User perceptions and experience of the different assistants: (a) Appropriateness of response length, (b) Adaptability . . . . .	33
Figure 11	User perceptions and experience of the different assistants: (a) Device interruption and (b) Helpfulness. **: $p < .01$ . . . . .	33
Figure 12	User preferences between conditions. The plot shows the number of participants who preferred which condition the most, the second-most, the third-most, and the least.	34
Figure 13	Adding a new memory to Resonance . . . . .	38
Figure 14	Adding a new imagination to Resonance . . . . .	39
Figure 15	Reduction in PHQ8 scores for people using Resonance during the study period . . . . .	43
Figure 16	Changes in mood when entering daily memory and interacting with AI . . . . .	44
Figure 17	Likelihood to act on AI suggestions . . . . .	45

Figure 18	Perceptions of participants about the AI suggestions . . .	46
Figure 19	Perceptions of participants about the AI-guided imagina- tions . . . . .	48
Figure 20	Correlations between mean likeliness to act of each par- ticipant and self-report perceptions of receiving the AI suggestions . . . . .	48
Figure 21	Correlations between mean likeliness to act of each par- ticipant and self-report perceptions of imagining the AI suggestions . . . . .	49



## LIST OF TABLES

Table 1	Average response length from the system in different conditions . . . . .	27
Table 2	Average query time by the participant in different conditions . . . . .	27
Table 3	Average process time from the system in different conditions . . . . .	28
Table 4	Correct and Incorrect responses generated from the system in different conditions . . . . .	28
Table 5	Don't Know and Speech Recognition Error percentages in different conditions . . . . .	29

# 1 | INTRODUCTION

Episodic memory is a neurocognitive system that supports the memory of personal experiences. It was first proposed as a concept by Endel Tulving in 1972, and has evolved since then with increasing neurological evidence supporting its existence, properties, and functions [93, 94, 21, 62]. Episodic memories are described as summary records of sensory-perceptual-conceptual-affective processing [21], thereby consisting of emotional responses, insights, sensory perceptions, and temporal sequences of events. The episodic memory system lets people consciously re-experience these records by subjectively traveling through time within the reality of their mind [94]. This conscious time travel is enabled by auto-noetic awareness, the innate ability to differentiate between re-experiencing the past and experiencing the present [30], debated to be uniquely human. Moreover, episodic memory is suggested to be an evolutionary fringe to allow the encoding and recollection of short-term experiences to measure progress towards a goal [1, 99]. Recent advances have shown episodic memory to be beyond solely a re-activation of stored experiences, but a process of scene construction, and thereby being instrumental in simulating possible futures [36].

People in modern society are facing a rise in memory issues through failed encoding or retrieval of memories when needed. This arises due to poor sleep [77], the mental health epidemic, aging [51], and information overload [52]. Further, episodic memory is suggested to be more vulnerable to neuronal dysfunction than other memory systems [62]. When the conscious recollection of past experiences is detrimentally affected or biased, the quality of life of a person is significantly reduced. These effects are pronounced in people living with neurological disorders such as dementia which affects the episodic memory system [73], or people with depression with negatively biased recall of information from memory [97]. Further, it can have downstream effects as episodic memory is found to play a role in processes of creativity [4], divergent thinking [64], problem-solving [88], and building a personal narrative [26].

Over the previous decades, technologists have built several digital systems to assist and augment people's episodic memory ranging from lifelogging systems [33], digital journals [41], and just-in-time information retrieval systems [83]. While most of the widespread work is on capturing episodic memory into a

digital medium through lifelogging and journaling, there has been limited exploration of being able to interact with them effectively for the current goal of the user. The recent advances of artificial intelligence (AI) through the advent of large language models [18] (LLMs) are well-suited to mediate people’s interaction with their episodic memories effectively. This arises from the abilities of the LLMs to semantically search [59] and parse and learn from context [58].

To develop suitable AI systems for interacting with episodic memories effectively, and thereby augmenting episodic memory for people, it is critical to pay attention to the role of episodic memory in completing personal goals, in line with the role it plays in humans. In this work, we explore AI assistants that augment people’s ability to access their episodic memories to serve different goals. Taking the goals of better performance at a current task, or learning about oneself and preparing for the future, the interfaces are designed to suit the context in which AI is helpful to the user.

Episodic memory is used by people for better performance at a current task. Tasks could range from referencing past social interactions with other people to remembering what needed to be bought at a grocery store based on what was seen in the fridge at home. Challenges in episodic memory retrieval can sometimes result in misunderstandings or missed social cues. While a real-time memory augmentation system could accurately bring back information as needed [83] to aid with these situations, the interface is not optimally designed to help with the current task seamlessly. An optimal interface needs to be minimally disruptive to the current task of the user to reduce distraction. It needs to adapt to the user by picking up on associated memory cues such as time or place. Here, we develop Memoro, a real-time and wearable memory assistant designed to be minimally disruptive to enable people to optimally perform in the current task. The concept of minimal interfaces is depicted in figure 1.

Episodic memory is also accessed in a reflective manner and contributes to learning about oneself and to personal growth. A well-studied method to do this systematically is journaling, a process of externalizing experiences that have occurred and reflecting on them. When retrieving episodic memories of these experiences through reminiscence, and then further appreciating and anticipating future experiences similar to them, the benefits of gratitude and anticipation are realized for the person [61]. This capacity of finding benefits in one’s past, present, and future positive experiences is called savoring in positive psychology [9] and is correlated with measures of happiness and reducing depression[47]. Artificially intelligent systems that use LLMs are well-suited to

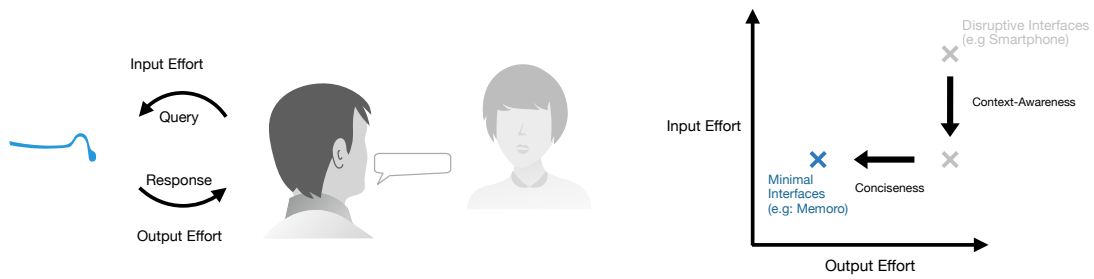


Figure 1: Minimally disruptive interfaces with wearable AI assistants

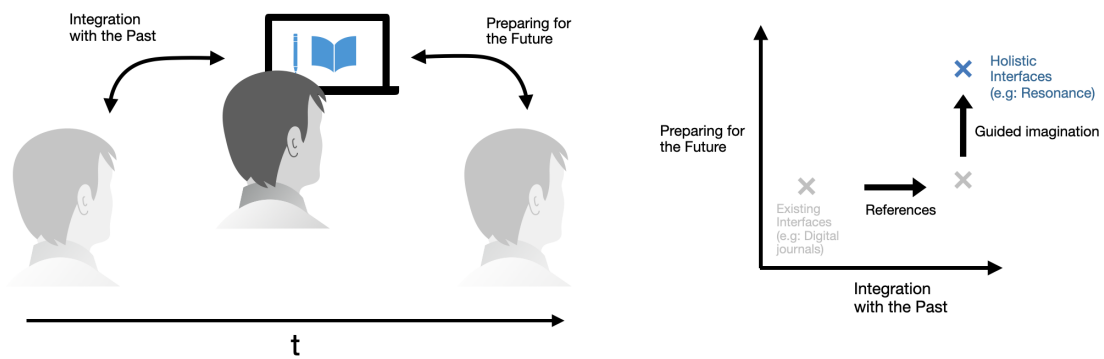


Figure 2: Holistic interfaces for being more connected during journaling with AI

work with semantically relevant memories and assist in the act of savoring. To realize an apt interface that enhances the connection of a person with their past and future self to derive mental health benefits, we built Resonance. Resonance is a web-based application that provides a holistic interface while offering actionable suggestions to people to engage in savoring when they journal daily. The suggestion uses a combination of quotes and references from the user’s past memories, an idea for feeling positive emotions in the future and anticipation of the suggestion through imagination. The concept of holistic interfaces is depicted in figure 2.

To evaluate Memoro and Resonance, we had (N=20) and (N=55) participants respectively in a user study with specific experimental conditions designed to elicit and study the efficacy of the assistants, their effects on the participants, and the general feedback.

## 1.1 CONTRIBUTIONS

In summary, the key contributions of this thesis are as follows:

- Development of a framework for minimal interfaces for real-time memory assistance
- *Memoro*, a real-time wearable memory assistant to interact with episodic memory with minimal disruption to the primary task
- Development of a framework for holistic interfaces for reflective memory assistance
- *Resonance*, a reflective web-based memory assistant to interact with episodic memory to savor one's past, present and future experiences for improved mental wellbeing
- Quantitative and qualitative evaluation of the two interfaces with user studies

## 2 | BACKGROUND AND RELATED WORK

### 2.1 ROLE OF EPISODIC MEMORY IN HUMAN WELL-BEING

The existence of episodic memory offers a fundamental role for people to subjectively re-experience the past and imagine the future to serve goals such as progress measurement and a source for developing conceptual knowledge about the world and the self[21, 20]. Episodic memory plays a key role in determining the current and future well-being of people [75] as measured through satisfaction with life.

There is growing neurological evidence for the episodic memory subsystem to be recruited for second-order processes such as creative thinking [63], divergent thinking [64], and imagination [85]. This led to the constructive memory hypothesis that shows that episodic memory is more constructive than just a reactivation of past events. It is an adaptive process combining pieces of past experiences to form a scenario. This supports the methods of how people recruit episodic memory to imagine future events. An intervention-based approach called episodic specificity induction (ESI) by Madore et al [63] further showed how the increased recruitment of the episodic memory system through conscious recollection of episodic elements of memories led to more creative thinking[63], better problem solving and increased divergent thinking [64]. ESI led to increased overall wellbeing and decreased rates of depression [46]. Episodic memory further contributes to learning and knowledge acquisition by playing a critical role in the process where information is gradually abstracted from specific episodes and integrated into long-term conceptual knowledge.[22]. These lines of evidence suggest the rich role that episodic memory plays in the lives of humans and their wellbeing, supporting reasons to augment it.

### 2.2 WEARABLE MEMORY AUGMENTATION SYSTEMS

Wearable memory augmentation has been a well-researched area since the 1990s when Mik Lamming coined the term "memory prosthesis" [57]. Since

then, there have been various forms of wearable memory augmentation systems, including reminder systems and lifelogging systems [57, 82, 95, 65, 37, 35, 45, 14]. Lifelogging devices continuously capture signals such as audio, video, and biosignals resulting in a vast store of data. In the audio domain, Vemuri et al. [95] introduced a personal audio memory aid that can record information and allow the user to search it using keywords. Hayes et al. [37] showed the personal audio loop (PAL) as a ubiquitous service to recover audio content. Yamano and Itou [102] and Shah et al. [87] recorded audio lifelogs using wearable microphones and experimented with different ways of browsing these lifelogs through a smartphone application. However, such types of browsing and keyword querying of audio data require a screen and, hence, use the users' visual focus and time to read the information provided. Gelonch et al found that an important factor in the acceptance of wearables in memory augmentation was the ease of use [31]. Furthermore, they were not designed to have quick and seamless interactions where disruption time during usage is critical, such as in conversations or driving.

Enabling voice-based interfaces for the users helps in memory retrieval from their lifelogs [76]. Furthermore, voice interfaces can enable users to maintain high face focus and eye contact during conversations [12]. Therefore, we present a voice-based retrieval approach for an audio-based wearable memory assistant that can handle natural language queries with a focus on minimizing disruption to the primary task of the user. With concise responses from the assistant serving as memory suggestions, we aim to reduce device interaction time and preserve the quality of the primary task while using the system. Additionally, when the user is attempting to retrieve specific details from a lifelog, we explore a method to allow users to skip having to form an explicit query by having the assistant infer their memory retrieval query based on the current conversational context.

## 2.3 DIGITAL JOURNALING WITH AI

As people have used external memory aids such as notebooks to document their lives for several decades, the concept of digitally storing and recalling life events has taken ground since Vannevar Bush's concept of the Memex in 1945. Journaling, through recording personal thoughts, daily experiences, and evolving insights, has aided people to remember experiences more profoundly [27], obtain mental health benefits through reduced anxiety and stress [29, 90,

68], improve critical and reflective thinking [67, 38, 28] and enhanced learning [78, 69].

A strong benefit of digital journaling is that past memories are stored and are available for easy access. Large language models have given digital systems abilities to search through semantic documents, parse context, and learn about users. The most common interface for such are conversational agents through chatbot interfaces that have been explored in interactive journaling [3]. While such systems may allow people to talk to an AI chatbot, they are primarily designed to be an information retrieval system for the user. Being able to interact with them by looking into one's past can encourage self-reflection [72] and practicing gratitude [44]. However, there can be further beneficial downstream effects such as looking into the future for anticipation and taking agency in one's new experiences. In positive psychology, this is captured succinctly by the phenomenon of Savoring[9], which is shown to drive depression levels lower and increase happiness. Digital journaling with AI can therefore be leveraged to meaningfully allowing a person to connect with their past and future when they sit down to journal. To design for this temporal connection, we develop a holistic interface that is based on neuroscientific evidence of the close connection of memory and imagination[85]. This is aimed to help people to derive mental health benefits through enhanced savoring. We develop a web application, Resonance, that embodies this reflective memory augmentation AI interface during daily journaling.



# 3 | MEMORO: REAL-TIME MEMORY AUGMENTATION

## 3.1 INTRODUCTION

Memory augmentation and information retrieval systems have been of key interest to the HCI community over the past several decades as tools to address these growing challenges. Since Vannevar Bush's conception of the Memex in 1945 [11], there has been extensive work on systems and devices to extend our memory [81, 16, 17, 55] such as lifelogging systems that continuously record the user's media and signals [65, 95, 45], and just-in-time information retrieval systems [50, 5, 81, 66, 25] that provide relevant information based on the user's context. While these wearable systems demonstrate the capabilities of users to retrieve vast amounts of information, limited research exists on designing interfaces that enable the retrieval of information in a minimally disruptive way when the user is already engaged in a primary task, which is often the case with wearables.

### 3.1.1 Minimally Disruptive Interfaces

Minimal disruption for a memory augmentation interface can be defined as (1) requiring minimal input from the user to request information, i.e., the input the user gives is short, and (2) providing minimal output, namely the suggestion or response provided by the augmentation system is the smallest amount of information that will give the user the information they need. The minimal disruption design consideration is critical for the usability of wearable memory augmentation systems [31], especially in social settings that are attention-demanding and where incidentally the highest number of memory lapses occur [74], such as conversations.

Therefore, an important challenge for the design of wearable memory augmentation systems is that of a seamless, user-friendly, and concise search interface [31] to keep disruption to the user's primary task minimal. Incorporating context awareness can reduce or, as we show in this paper, even completely eliminate the query input, allowing users to skip posing an explicit, comprehensive retrieval query, as the system can directly infer the user's specific memory

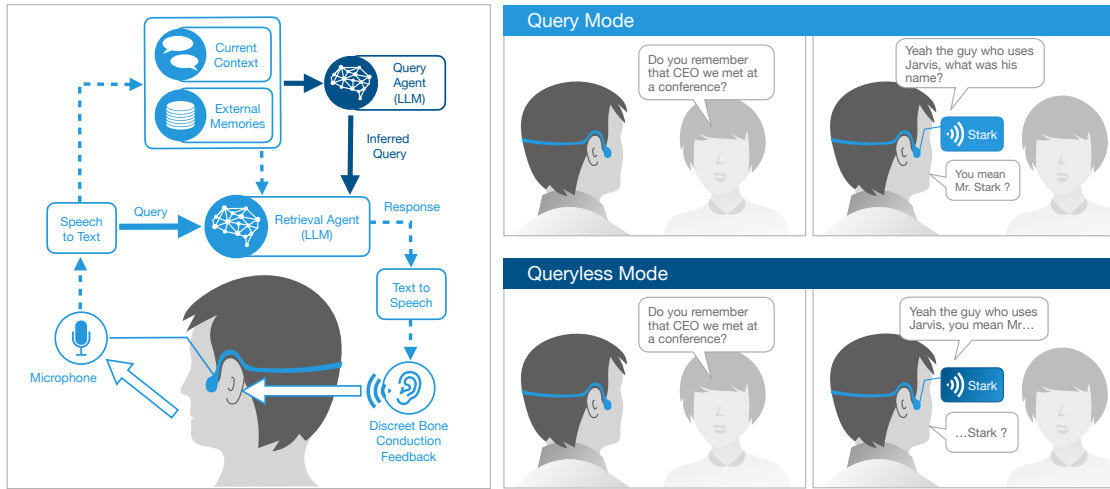


Figure 3: Interactions using Memoro’s Query and Queryless Modes

needs. Recent developments in large language models (LLMs) have improved capabilities in parsing conversational context in natural settings [8, 98] and enable more flexible search queries using alternative phrases [59]. They also enable the shortening of answers [32] for succinct suggestions. This highlights the opportunity to leverage LLMs to design easy-to-use and minimally disruptive interfaces. The minimal interface design is shown in 1.

Here, we aim to answer the following research questions

- **RQ1.** How can we design a seamless wearable memory assistant using LLMs to reduce disruption to the primary task with minimal and effective input and output?
- **RQ2.** What are the effects of using the memory augmentation system during the primary task of a real-time conversation across metrics such as quality of conversation, performance, and task load?
- **RQ3.** How do context awareness and conciseness affect the system’s usability, user perception, and experience?

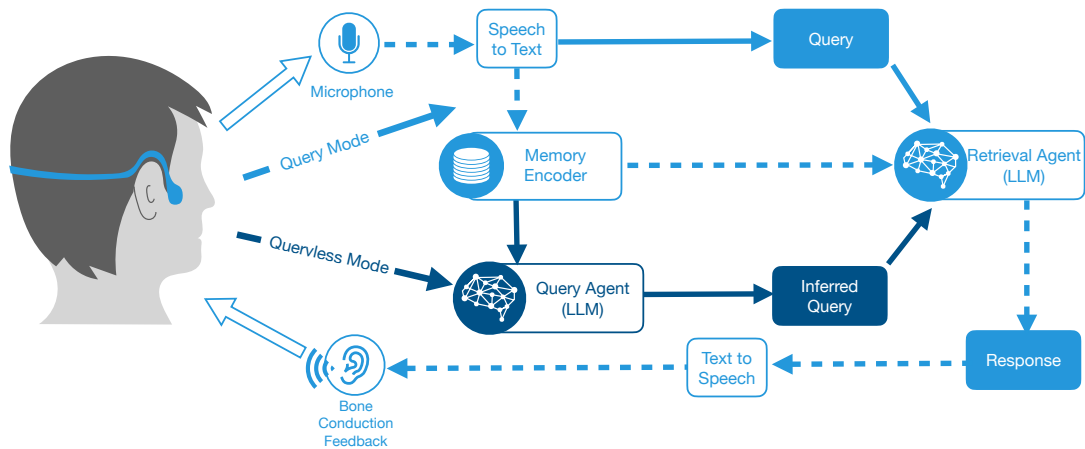
We developed a minimally disruptive audio-based wearable assistant, Memoro, that uses LLMs to aid the user in retrieving relevant information from previously recorded personal data through concise suggestions. Memoro continuously transcribes and encodes audio data from conversations the user engages in. The memory assistant has two modes of interaction for retrieval: Query Mode, where the user voices their natural language query, and Queryless Mode

where the user is presented with a suggestion relevant to the current conversational context without having to explicitly query the system. Both modes provide minimal memory responses to the user (see Figure 3). In terms of hardware form factor, Memoro uses a light-weight, bone-conduction headset for unobstructed and private responses.

## 3.2 SYSTEM DESIGN

Memoro, or "I remember" in Latin, is an audio-based memory assistant with a concise user interface. It continuously listens to the surrounding audio and encodes the raw speech transcriptions in memory, tagged by the timestamp at which it was transcribed and stored locally in the device, similar to previous works [95, 37]. Whenever the user is in a primary task and has a real-time need for retrieval of information, they can trigger the system by pressing a ring button. The button informs the system that the user has a memory need. The button push can trigger one of two interaction modes:

1. **Query Mode:** The user can explicitly query their Memoro system using natural language speech. If the user is in an ongoing conversation, the user can ask a brief question related to the conversation as the system is continuously listening, thus giving it conversational contextual awareness. For example, if the user is talking to a supermarket attendant and has said *"I have bought eggs and bread"* in the conversation and wishes to remember the third thing they intended to purchase, they can hold the trigger button for Query Mode while asking *"What was the third thing?"*. The system would then retrieve the answer, *"Bananas"*, from the previously recorded memories. The retrieved answer is converted to audio using text-to-speech and played to the user through a bone-conduction headset.
2. **Queryless Mode:** The user can also request predictive assistance, such that the system will infer the information that the user needs based on the current context and deliver the response without any explicit query from the user, similar to an autocomplete functionality. With the same example as above, after saying *"I have bought eggs and bread but need to buy .."*, the user could trigger the Queryless Mode for the system by pressing the button which will based on parsing of the conversational context, infer the query, and respond with the suggestion *"Bananas"* for the user to integrate into their incomplete sentence.



**Figure 4:** Closed loop system architecture has a memory encoder that is continuously updated using text-to-speech.

Memoro has three components: the memory encoder, the retrieval agent, and the query agent. Figure 4 shows an overview of the complete system architecture.

The wearable platform consists of a commercial bone conduction headset that communicates with a smartphone or laptop. The bone conduction headset gives the user a parallel channel of audio [70, 40], allowing them to have conversations with people while being able to hear audio responses from Memoro without impeding their field of view. The headset has an in-built microphone. Speech recognition is implemented using Google’s Speech-to-Text API and speech synthesis of the text response from the memory assistant uses the Google Text-to-Speech API. The large language model used is OpenAI GPT<sub>3</sub> (davinci-003) [8] with a temperature of 0.

### 3.2.1 Memory Encoder

Auditory memories are stored using a two-step process. A continuous transcription is run on what the microphone picks up, including both the speech of the user and the conversation partner, under the assumption that privacy consent has been addressed. The transcription is first stored as the **Current Context** of the conversation. The current context is maintained in a fixed-sized

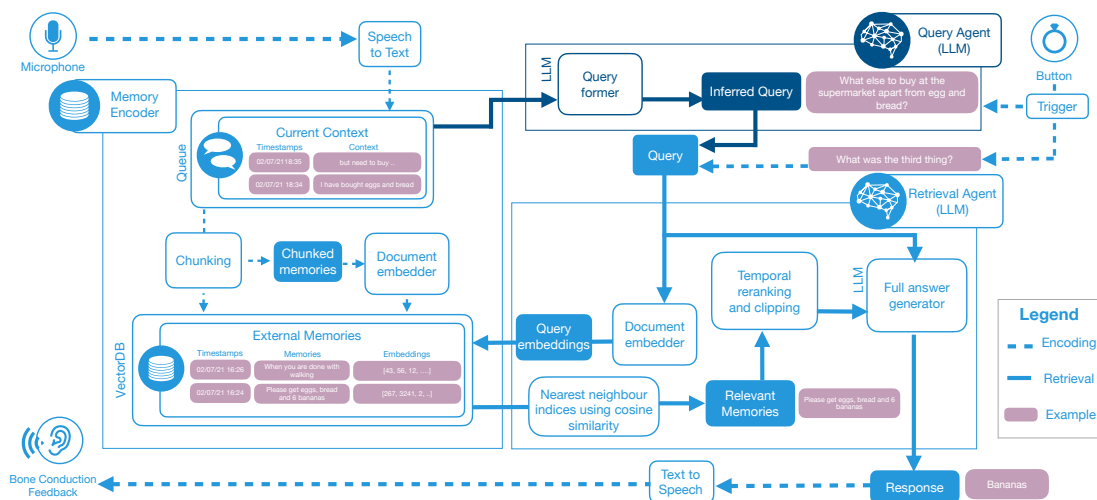


Figure 5: Workflow of the components of Memoro: memory encoder, retrieval agent, and query agent.

buffer of the last  $\alpha$  characters of data. We set  $\alpha$  to 75 characters for capturing the most recent couple of sentences in the prototype but can be set larger to capture more context. The buffer is continuously updated by adding new information and removing information that is beyond the  $\alpha$  threshold specified earlier. The set of information removed from the current context is chunked together into a single block and then encoded into the **External Memories** as a memory.

Encoding of the memory is done using sentence embedding vectors of the text transcription of the full block. The embeddings capture the meaning of the memory enabling semantic search beyond keyword matching. Embeddings are calculated using pre-trained all-MiniLM-L6-v2 sentence transformer model [79] which maps sentences and paragraphs to a 384-dimensional dense vector space. Through these embeddings, the most semantically relevant memories containing the answer to the user query can be selected during retrieval. The embeddings, the text transcription, and the start timestamp for each memory block are stored using a vector database for faster retrieval [56]. Figure 5 shows the encoding process of transcriptions into external memories.

### 3.2.2 Retrieval Agent

The aim of the retrieval agent is to take a query and respond with a concise answer from the user’s encoded external memories, enabling the *Query Mode*.

It uses a method called retrieval augmented generation developed by Lewis et al[59] and used in state-of-the-art question-answering systems [71, 89].

### *Contextual query*

To increase ease of use and reduce input to the memory assistant, queries from the user can be shortened using contextual awareness. As the device continuously tracks the context of the ongoing flow of the conversation, it enables the user to query the memory assistant with questions that build on this flow for a less disruptive interaction. For instance, if a user is saying the following sentence, *"John teaches science, math and..."*, and wishes to recall the third subject that John teaches, with context awareness of the assistant, the user could directly query *"What else?"* as opposed to having to formulate the full context-unaware query *"What is the subject that John teaches other than science and math?"*.

The contextual search is implemented using the following approach. When the user voices a natural language query to the memory assistant, the query and the **Current Context** containing the most recent conversation are combined to retrieve relevant external memories from the vector database. First, the vector embeddings for the query and current context, which are concatenated, are calculated using the same embeddings model used in the memory encoder. These vector embeddings are used to search for the most semantically similar external memories by comparing them to the stored embeddings of the **External Memories** which are pre-calculated during the encoding process. The comparison uses the established approximate K nearest neighbor search with cosine score as the similarity measure [56]. The text transcriptions of the 10 most similar external memories constitute the relevant memories for the contextual search. The relevant memories are reordered based on ascending timestamps to form temporally linear memories and then clipped to the token limit (4096 tokens) of the large language model that was used in this implementation. The query, current context, and retrieved relevant memories are then combined, as described in Figure 5, to form a prompt for the text generation language model. The prompt uses a combination of explicit and structured prompt engineering. Explicit prompts directly request the LLM to generate an answer to the user query from the relevant memories, while the structured aspect uses a template to guide the generation to a parseable form. The prompt is designed to be able to search through relevant memories and generate the answer. The prompt can be found in Appendix A.2.

### *Concise Suggestions*

Once the answer has been retrieved using the above method, it is further post-processed to be more concise to minimize response duration and reduce output from the assistant. Searching through External Memories, rather than sifting through new information, allows for further conciseness [23]. For instance, "Her name is Sarah" can be replaced with "Sarah". Therefore, the objective of this step is to eliminate any extraneous words such as connectives that do not address the question. Further, contextual compression could be used to remove any words that have already been retrieved by the user, either in the query or in the current conversational context. For instance, with the current context as "She is an engineer" the query "What was her name and what is her specialization?" and the generated answer "Her name is Emily and she works as a Software Engineer" gets compressed to "Emily, Software". Addressing the query from the user, the answers can be shortened to specifically what is needed to complete the user's need. This is critical as language models tend to be more verbose as they are optimized for informativeness [91]. The conciseness and redundancy removal are implemented by passing the query, current context, and the generated answer from the previous run to the retrieval agent with a template prompt that instructs the language model appropriately. The template prompt can be found in Appendix A.2.

### 3.2.3 Query Agent

In order to further streamline the interaction between the user and Memoro, we implemented an additional feature in the memory assistant that enables the user to receive on-demand predictive assistance without having to explicitly form a query, enabling the *Queryless Mode*. This is facilitated by the user requesting the memory assistant to parse the ongoing flow of the conversation and infer their precise memory need. For example, if the user is already saying "He likes to play Settlers of Catan, Pandemic and ...", and then triggers the assistant, the query agent can predict the user query "What is the third board game he likes?" allowing the user to skip query formation. To achieve this, we use a method that infers the query that the user is likely to ask based on a **Current Context** buffer, similar to the one implemented in the *Query Mode*. The question inference leverages another iteration of prompting the language model to produce the query. The prompt can be found in Appendix A.2. The inferred query is then passed to the retrieval agent and the resulting concise answer is then presented to the user using text-to-speech synthesis. By implementing this feature, we aim to

minimize the time spent in interactions during conversations, making Memoro more efficient and user-friendly.

## 3.3 USER STUDY

### 3.3.1 Participants

Participants were recruited through email lists as well as snowball sampling and word-of-mouth. 20 participants took part in the study (9 male, 9 female, 2 non-binary, age range = 18 to 32,  $age_{mean} = 23.4$ ,  $age_{SD} = 4.2$ ). Participants were fluent or native English speakers with normal or corrected-to-normal hearing. Participants rated their listening memory between 'Somewhat bad' (4), 'Neither good nor bad' (3), 'Somewhat good' (11), and 'Extremely good' (2). Additionally, the participants rated their frequency of experiencing tip-of-the-tongue moments in conversation as 'Never' (1), 'Sometimes' (13), 'About half the time' (4), and 'Most of the time' (2). The participants rated their frequency of using voice assistants as 'Not at all' (7), 'Once a month' (3), 'A few times' (4), 'Once a week' (3), 'More than once a week' (3). The study received ethics approval from the university ethics review board, and participants gave written consent to take part in the study.

### 3.3.2 Conditions

The conditions were designed to elicit differences to technically and subjectively evaluate Memoro during the conversation. To address **RQ2**, we had a No SYSTEM condition, where the participant engages in the task without the use of the system to compare and understand the effects on conversational quality and task load. In order to address **RQ3**, which was to determine the effect of contextual awareness and conciseness on the system's usability and user preferences, we set up a Baseline LLM system that is identical to the retrieval agent for question answering but does not use contextual awareness or conciseness. Therefore, participants needed to ask comprehensive questions in the Baseline condition and receive complete answers from the system. Overall, there were four conditions:

- No SYSTEM which was the control condition
- BASELINE LLM system with explicit query and raw, full-length answers



- QUERY MODE of Memoro with explicit contextual query and concise answers
- QUERYLESS MODE of Memoro with no query and concise answers.

In addition, technical evaluations were conducted to measure the system response accuracy and its conciseness. The interaction modes were separately analyzed for a detailed evaluation.

### 3.3.3 Procedure

#### *Introductions to Fictional People*

We created four fictional people who were introduced to the participants, one for each condition. The introductions consisted of information-dense details such as the fictional persons' occupations, families, hobbies, and interests. The introductions were played as audio with an image of the fictional person (generated using an online AI face generator<sup>1</sup>) displayed on screen and were around 2 minutes long per person. The experiment was designed to make it very difficult to remember all these details. The introductions formed the External Memories for subsequent interactions with the memory assistant. No additional information was encoded into the External Memories during the conversation for a careful study of the interaction modes of the system.

#### *Related Conversation*

To simulate scenarios where the participants would be in a real-time conversation and allow them to use the system, we engaged the users in an open-ended conversation consisting of scripted questions about the fictional people, with the researcher. For each fictional person, there were two general questions and four specific questions). The researcher made sure to use the question set during the conversation. The responses from the participants were not scripted and they could choose when and how to interact with the system in the given condition at their discretion.

---

<sup>1</sup> <https://thispersondoesnotexist.com/>

## 3.4 RESULTS

We show the analysis from the user study of the systems' usability, technical evaluation, user perceptions and experience, and preferences. The results are further discussed in relevance to the research questions in the Discussion chapter.

### 3.4.1 Technical Evaluation

A total of 392 interactions with the system were captured in the user study for all conditions: 102 for the Baseline, 150 for the Query Mode, and 140 for the Queryless Mode. Each interaction indicates a moment when the user requested memory assistance by using the button. We used these interactions for the technical evaluation.

#### *Conciseness and Processing Times*

The normality assumption for the response length data was not met according to the Shapiro-Wilk test ( $p < .05$ ). Friedman tests ( $k=3$ ) were conducted to determine if there were main effects in the conciseness. The test indicated significant differences between conditions ( $\chi^2 = 135, p < .001$ ) in the response length from the system. The Wilcoxon signed-rank tests with Bonferroni correction showed that the Query Mode resulted in significantly shorter responses for the queries asked by the user as compared to the Baseline ( $p < .001$ ), with an 85% reduction in the mean number of characters from 115.4 to 16.6. The Queryless Mode has a response length similar to the Query mode. The average query time was also reduced by 15% from 3.4 seconds for the Baseline to 2.9 seconds for the Query Mode ( $p = .03$ ). The query time is not applicable for the Queryless Mode. The average processing time of the system for the Baseline and Query Mode was 1.4 seconds and 2.3 seconds for the Queryless Mode. The processing time reflects the time from the end of the query to the start of the audio feedback of the answer. Tables 1, 2, 3 shows the detailed statistical results.

#### *Accuracy of responses generated by the System*

Overall, the accuracy of the Baseline and Query Mode was 80.3% and 84% respectively as shown in table 4. Notably, for 11.7% and 6.0% of interactions, the systems correctly determined that the question did not have an answer in the External Memories, as shown in table 5. Further, in the inaccurate responses, the

**Table 1:** Average response length from the system in different conditions

Condition	Average Response Length (n chars)
Baseline	115.4±82.9
Query Mode	16.6±11.0
Queryless Mode	21.1±11.8

**Table 2:** Average query time by the participant in different conditions

Condition	Average Query Time (s)
Baseline	3.4±2.8
Query Mode	2.9±3.9
Queryless Mode	-

participants could identify the inaccuracy and request the correct response with a different query. The Queryless Mode had an accuracy of 70.7% and the drop was due to the Query Agent misinterpreting the context. For instance, during an interaction of P17, the Current Context contained *"His favorite authors are Neil Gaiman and Ursula .."* and the inferred query was *"What are William Thompson's hobbies and interests?"* which was incorrect as the participant was looking for the last name of *Ursula*. However, we observed the response accuracy of the Queryless Mode was sufficient for a detailed evaluation. This was reflected by the final user preferences.

### ***Handling diverse queries from users***

Using a large language model (LLM) allows the system to understand the intent of a user and enables natural language search beyond keyword matching, such as semantics. With sufficient information, it can predict the query by understanding the user's intent. The performance of the retrieval and query agents using LLMs are illustrated with the following examples of the interactions by two of the participants (P3, P19) in Figure 6. In the first example (P3), the user opted to substitute the term *'gym'* with the phrase *'place for working out'*, and the retrieval agent comprehended the intention of the user. In the second example (P19), the query agent interpreted that the user was looking for the third activity and inferred a query for the retrieval agent, resulting in a successful interaction. More such examples can be found in Appendix A.1.

**Table 3:** Average process time from the system in different conditions

Condition	Average Process Time (s)
Baseline	1.4±0.7
Query Mode	1.3±0.6
Queryless Mode	2.3±0.8

**Table 4:** Correct and Incorrect responses generated from the system in different conditions

Condition	Correct (%)	Incorrect (%)
Baseline	80.3	2.9
Query Mode	84.0	6.7
Queryless Mode	70.7	23.5

### 3.4.2 Conversation Quality, Task Performance, and Task Load between Conditions

#### *Quality of Conversation*

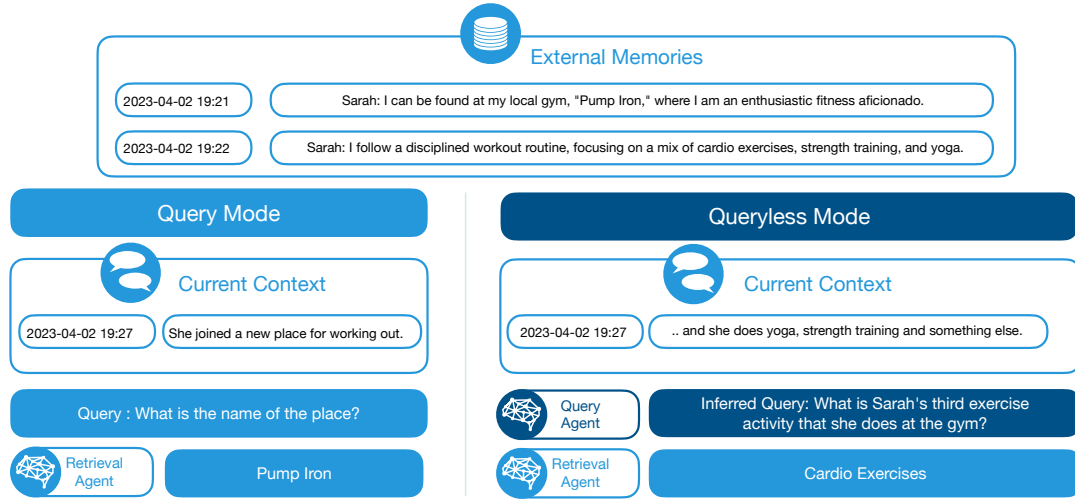
There were no significant differences in conversation quality between conditions for the measures of attention ( $\chi^2 = 3.63$ ,  $p=.303$ ), concentration ( $\chi^2 = 7.21$ ,  $p=.0655$ ), eye contact ( $\chi^2 = 7.00$ ,  $p=.0719$ ), and how relaxed they were during the conversation ( $\chi^2 = 3.85$ ,  $p=.278$ ). The quality of the conversation was preserved and not reduced in any of the conditions. We found a significant difference in the naturalness of conversation between the conditions ( $\chi^2 = 13.8$ ,  $p<.01$ ). There were significant differences between the No System condition and the system conditions: No System-Baseline  $p<.01$ , No System-Query  $p<.01$ , No System-Queryless  $p<.01$ , No System  $M=5.75$ ,  $SD=1.41$ , Baseline  $M=4.30$ ,  $SD=2.03$ , Query  $M=4.25$ ,  $SD=1.94$ , Queryless  $M=4.55$ ,  $SD=1.82$ .

#### *Task Performance and Task Load*

There was a significant difference in the confidence in recalling information between the conditions ( $\chi^2 = 19.9$ ,  $p<.001$ , Figure 7a). Confidence in recalling was significantly higher in the system conditions compared to the No System condition and: No System-Baseline  $p<.001$ , No System-Query  $p<.001$ , No System-Queryless  $p<.001$ , No System  $M=43.1$ ,  $SD=26.9$ , Baseline  $M=75.4$ ,  $SD=19.6$ , Query  $M=80.0$ ,  $SD=17.8$ , Queryless  $M=77.1$ ,  $SD=18.7$ .

**Table 5:** Don't Know and Speech Recognition Error percentages in different conditions

Condition	<i>Don't Know</i> (%)	Speech Recognition Error (%)
Baseline	11.7	3.9
Query Mode	6.0	3.3
Queryless Mode	0.7	2.8



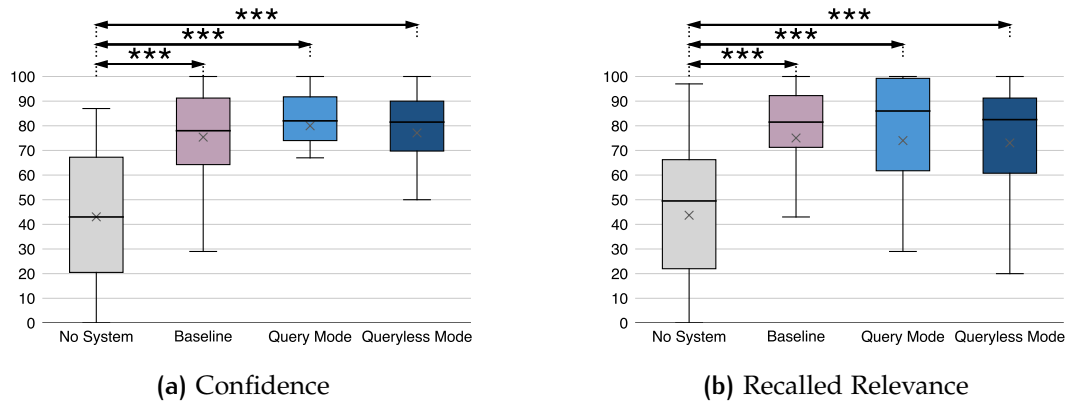
**Figure 6:** Example interactions by P3 and P19 show the Query Mode and the Queryless mode for the same memory respectively.

There was a significant difference in the relevance of recalled information between the conditions ( $\chi^2 = 18.5, p < .001$ , Figure 7b). There were significantly higher relevance ratings for the system conditions compared to the No System condition: No System-Baseline  $p < .001$ , No System-Query  $p < .001$ , No System-Queryless  $p < .001$ , No System  $M = 43.7, SD = 29.6$ , Baseline  $M = 75.1, SD = 26.3$ , Query  $M = 74.0, SD = 32.1$ , Queryless  $M = 73.1, SD = 24.6$ .

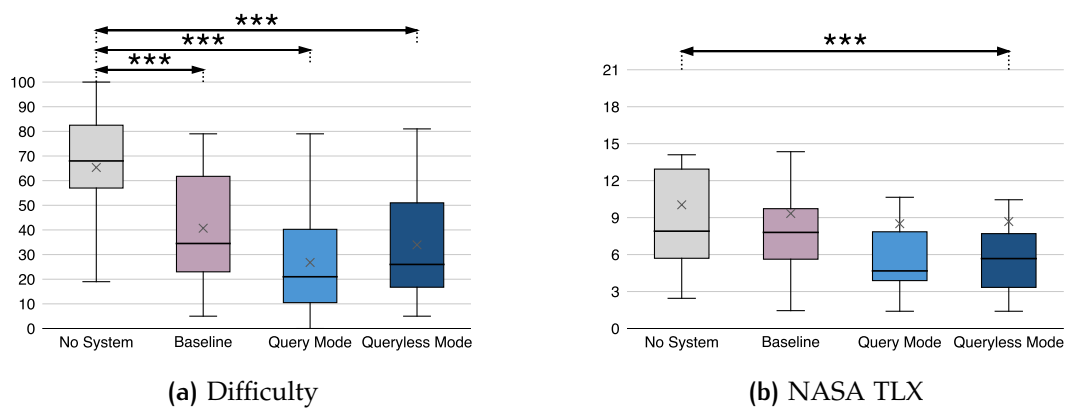
There was a significant difference in the difficulty in recalling information between the conditions ( $\chi^2 = 12.1, p < .001$ , Figure 8a). Participants found it significantly more difficult to recall information without the system compared to the system conditions: No System-Baseline  $p < .001$ , No System-Query  $p < .001$ , No System-Queryless  $p < .001$ , No System  $M = 65.4, SD = 26.0$ , Baseline  $M = 40.7, SD = 24.4$ , Query  $M = 26.8, SD = 21.0$ , Queryless  $M = 34.0, SD = 24.8$ .

We found significant differences in task load (RTLX) scores between conditions ( $\chi^2 = 12.0, p < .001$ , Figure 8b). Post-hoc analysis showed a significant difference in RTLX between No System ( $M = 10.0, SD = 7.06$ ) and the Queryless

Mode ( $M=8.68$ ,  $SD=11.4$ ). Overall, the RTLX scores were generally lower in the system conditions compared to the No System condition: Baseline  $M=9.34$ ,  $SD=7.19$ , Query  $M=8.51$ ,  $SD=9.93$ .



**Figure 7:** Task Performance (a) Confidence in recalling, (b) Relevance of recalled information. \*\*\*:  $p < .001$



**Figure 8:** Task load: (a) Perceived difficulty in recalling, and (b) Raw NASA TLX scores. \*\*\*:  $p < .001$

### 3.4.3 User Perceptions and Experience with Memoro

#### *System Usability*

The Query Mode of Memoro had the highest mean usability score of 80.0 ( $SD=11.8$ , Figure 9). The Queryless Mode had a usability score of 77.1 ( $SD=8.1$ ) and the Baseline had the lowest usability score of 68.75 ( $SD=15.15$ ). Since the data was normally distributed according to the Shapiro-Wilk test ( $p>.05$ ), a repeated measures ANOVA showed a main effect of the systems on the usability score ( $F_{(2,38)}=5.053$ ,  $p=.011$ ). A Tukey HSD post-hoc test showed a significant difference ( $p=.015$ ) between the usability of Baseline and Query Mode.

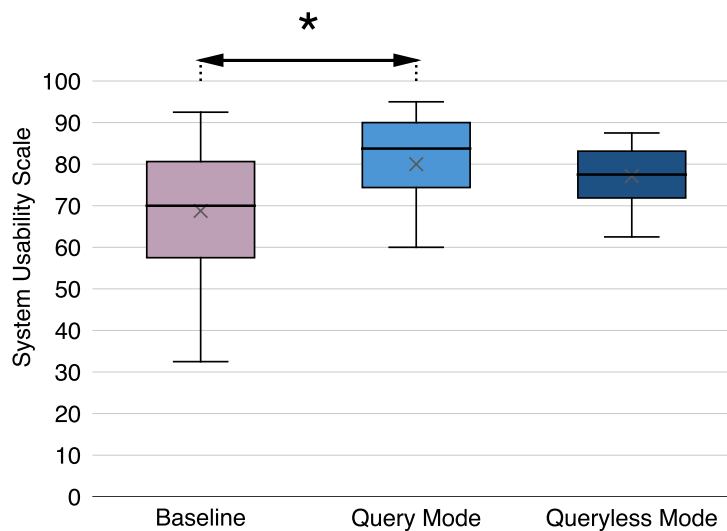


Figure 9: System Usability Scale (SUS) scores for the different conditions. \*:  $p<.05$

The normality assumption for the rating data was not met according to the Shapiro-Wilk test ( $p<.05$ ). Friedman tests ( $k=3$ ) were conducted to determine if there were main effects of the system conditions on the measures. Wilcoxon signed-rank tests with Bonferroni correction were used for post-hoc analysis when effects were found.

#### *Rated Length of Responses*

The Friedman test indicated significant differences between conditions ( $\chi^2 = 26$ ,  $p<.01$ ) in the rated appropriateness of the response lengths (Figure 10a). The Query ( $M=5.55$ ,  $SD=1.05$ ) and Queryless ( $M=5.45$ ,  $SD=1.19$ ) Modes had

significantly higher ratings in length appropriateness (Query-Baseline:  $p < .01$ , Queryless-Baseline:  $p < .01$ ) compared to the Baseline ( $M=2.80$ ,  $SD=1.47$ ). There were no significant differences between Query and Queryless Modes ( $p=.685$ ).

### ***Adaptiveness of System***

Adaptiveness is defined as how closely the system is able to monitor the current context of the conversation with respect to the user perception. There was a significant difference between conditions ( $\chi^2 = 11.7$ ,  $p < .01$ ) in the rated adaptiveness to the conversation (Figure 10b). The Query ( $M=5.35$ ,  $SD=1.31$ ) and Queryless ( $M=5.10$ ,  $SD=1.07$ ) Modes had significantly higher ratings in adaptiveness (Query-Baseline:  $p < .01$ , Queryless-Baseline:  $p < .01$ ) compared to the Baseline ( $M=3.40$ ,  $SD=1.85$ ). There were no significant differences between Query and Queryless Modes ( $p=.448$ ).

### ***Device Interruption***

The Friedman test showed a main effect of conditions on device interruption to the conversation ( $\chi^2 = 7.43$ ,  $p=.0243$ , Figure 11a). However, post-hoc analysis did not indicate any significant differences between the conditions: Baseline  $M=5.55$ ,  $SD=1.28$ , Query  $M=4.40$ ,  $SD=1.60$ , Queryless  $M=4.65$ ,  $SD=1.53$ , Query-Baseline:  $p=.0173$ , Queryless-Baseline:  $p=.0362$ , Queryless-Query:  $p=.498$ .

### ***Helpfulness and Usefulness***

There was no significant difference in the conditions in terms of helpfulness (Figure 11b):  $\chi^2 = 4.25$ ,  $p=.119$ , Baseline  $M=5.15$ ,  $SD=1.18$ , Query  $M=5.85$ ,  $SD=1.18$ , Queryless  $M=5.30$ ,  $SD=1.30$ . There was a significant difference in usefulness between the conditions ( $\chi^2 = 11.9$ ,  $p < .01$ ). Post-hoc analysis showed a significantly higher rated usefulness ( $p < .01$ ) for Query Mode ( $M=5.50$ ,  $SD=1.36$ ) compared to the Baseline ( $M=4.30$ ,  $SD=1.53$ ). No significant differences were found between Baseline and Queryless Mode ( $M=5.05$ ,  $SD=1.43$ ,  $p=.0358$ ), and Query and Queryless Modes ( $p=.233$ ).

### ***Politeness, Naturalness, Ease of Ignoring Device***

The Friedman test showed a significant difference in reported politeness of using the device in the conditions ( $\chi^2 = 8.10$ ,  $p=.0174$ ). Post-hoc analysis showed a significant difference in politeness ( $p=.0144$ ) between the Baseline ( $M=2.90$ ,  $SD=1.37$ ) and Query Mode ( $M=3.70$ ,  $SD=1.45$ ). No significant differences were



found between Baseline and Queryless Mode ( $M=3.65$ ,  $SD=1.35$ ,  $p=.0420$ ), and Query and Queryless Modes ( $p=.897$ ).

We found no significant difference in the conditions in how natural users acted (self-reported):  $\chi^2 = 3.30$ ,  $p=.192$ , Baseline  $M=3.35$ ,  $SD=1.57$ , Query Mode  $M=3.45$ ,  $SD=1.67$ , Queryless Mode  $M=4.05$ ,  $SD=1.47$ . There was also no significant difference in the conditions in how easy it was for the participant to ignore that they were wearing the device:  $\chi^2 = .128$ ,  $p=.938$ , Baseline  $M=4.00$ ,  $SD=2.03$ , Query Mode  $M=4.15$ ,  $SD=1.63$ , Queryless Mode  $M=4.10$ ,  $SD=1.59$ .

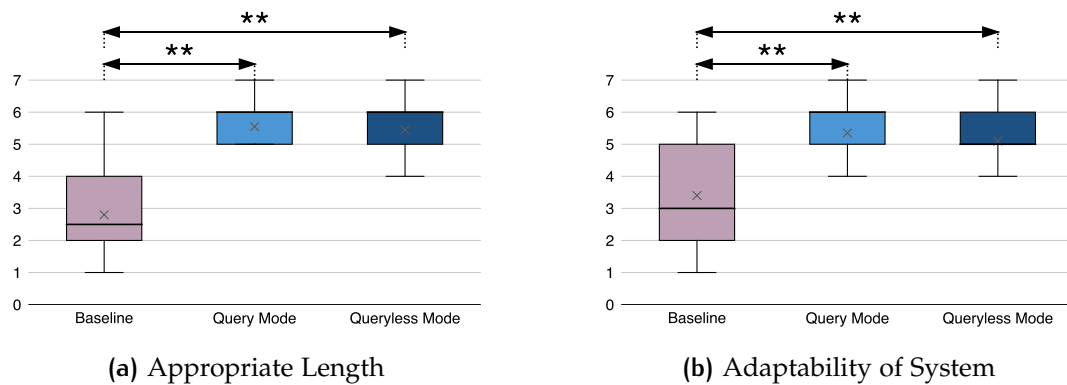


Figure 10: User perceptions and experience of the different assistants: (a) Appropriateness of response length, (b) Adaptability

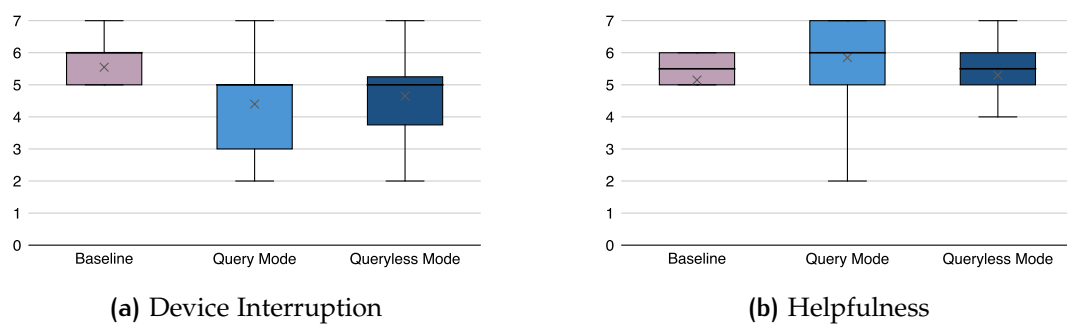


Figure 11: User perceptions and experience of the different assistants: (a) Device interruption and (b) Helpfulness. \*\*:  $p < .01$

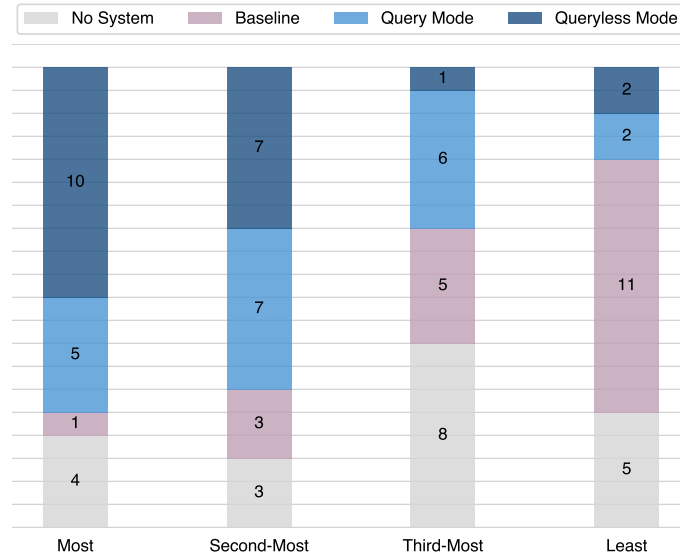


Figure 12: User preferences between conditions. The plot shows the number of participants who preferred which condition the most, the second-most, the third-most, and the least.

### User Preferences and Qualitative Feedback

The preference rankings are shown in Figure 12. 10 of 20 participants preferred the Queryless Mode the most, and 11 of 20 preferred the Baseline the least.

Participants felt that the Queryless Mode felt the “most seamless” (P19) and that it was “very nice and barely noticeable” (P14). They explained that they preferred it the most because it “preempts context” (P17), it “required the least amount of effort” and “anticipated” their needs (P12) and questions (P10). Participants also reasoned that it was the “best in terms of real-life usage, mainly because using it interrupted conversation the least” (P4) and it “made the conversation less awkward” (P7). P16 explained that Queryless mode was preferred to the Baseline and Query mode as “it seems a bit difficult and rude to ask question to the device, while I am still in conversation with the person”. Although it can be useful, P15 felt that more practice is needed to get used to using it: “...given some practice, I think the first questionless one has potential to be super useful with some practice. I just need to know when to hit the button for best results.”

A few participants preferred the Query Mode over the Queryless Mode. P20 explained that “[The Query Mode] is slightly higher [ranked] because I could ask a question and felt the other person knew that I was consulting someone else for the answer which made it more slightly OK than [the Queryless Mode]”. It “felt

*more appropriate/polite to use*" (P15) and it was the *"most easily integrated into the conversation"* (P6). In some cases, users felt that the Query Mode had higher accuracy (P1, P2, P3, P4, P7) and *"was better at answering"* (P5).

Most participants (16 out of 20) preferred at least one of the system conditions over the No System condition. The users who preferred having No System explained the systems as *"clunky"* (P17), or it depended on the task (P18); P5 explained *"I prefer natural conversation more which was easier without the assistant."*

Many participants felt that the Baseline was too lengthy (P2, P6, P7) and *"to the point it was a little distracting"* (P5). P8 mentioned *"...[it] went on for a long time and there wasn't a way to get it to stop or ask it to get to the point without waiting and stalling the conversation. I'd rather just move on and just leave it than have to wait unless it's REALLY important."* A few users preferred having No System over Baseline because it *"broke the conversation flow too much to be preferred over no system."* (P11) and *"[The Baseline] is ranked 4th because it provided redundant answers and didn't actually adapt to the conversation. I felt like it wasn't as useful as just having to remember information off the top of my head."* (P4). P10 liked Queryless Mode the most but ranked No System over Query Mode and Baseline: *"But if I have to explicitly ask it questions, I would prefer to just rely on my memory"*.

# 4

## RESONANCE: REFLECTIVE MEMORY AUGMENTATION

### 4.1 INTRODUCTION

Augmenting the "remembering self" to enhance overall life happiness is a concept grounded in cognitive psychology. Kahneman's theory of the "experiencing self" and the "remembering self" describes how happiness is significantly influenced by how people remember experiences in addition to the experiences themselves [48].

Journaling has provided a way for people to externalize experiences after they happened for various purposes such as to remember/forget them, process them more deeply, look back at them later, and to plan for future situations that may arise. Savoring, coined by Bryant and Veroff [10], is a practice in positive psychology reflecting the perceived capacity of a person to find benefit in past, present, and future experiences. To augment digital journaling and connect the user with their past and future selves requires the design of a more holistic interface.

#### 4.1.1 Holistic Journal Interfaces

To move journaling from a passive task to one that is more active by imbuing the concepts of savoring, there is a need for more holistic interfaces. A holistic journaling interface can be defined as (1) more integration with the past i.e. revisiting relevant memories to notice patterns, learning about oneself and practicing gratitude, and (2) more preparation for the future i.e. anticipating and visualizing new experiences to reap the benefits of looking forward to events. The holism consideration is critical to help people connect with their past and future selves.

Incorporating citations and references to past memories can give a window into the past of a user. Similarly, guided imagination through an action-based suggestion facilitates a user to be prepared and anticipate future positive experiences. It can help broaden their awareness of what can be done to feel more positive emotions, and in turn their own happiness. The holistic interface design is shown in 2. We design and develop Resonance, a web-based interface,

to have an artificially intelligent system engage the user in savoring during daily journaling.

The research questions here are:

- **RQ1.** How does interacting with Resonance daily over a period of time affect the mental wellbeing of a user?
- **RQ2.** How do users perceive reading and imagining the AI suggestions, and how does it influence mental wellbeing?
- **RQ3.** What themes emerge from user feedback on the strengths and weaknesses of Resonance?

## 4.2 SYSTEM DESIGN

Resonance is a web application through which a user can log in with unique credentials, enter memories, and interact with the AI suggestions. It consists of a primary dashboard displaying past memories and a button to input a new memory from the last 24 hours. The dashboard can be seen in the figure ?? and the new memory window can be seen in the figure 13. After a new memory is inputted, a suggestion related to that memory is shown to the user. The user is then asked to explicitly imagine the suggestion and input their imagination to the system.

### 4.2.1 Inputting Memories and Imaginations

Users log on to the application and can enter a memory from the previous 24 hours. The memory is inputted using voice for ease and for more intimate reflections. The speech is transcribed to text and is logged as a journal entry. For a quick reference to that entry, a three-word title is generated which is displayed in the dashboard along with the memory. The three-word title is generated using a large language model.

To encourage and guide the users to focus on the episodic elements while inputting their daily memories, a guideline adapted from a study on episodic memories and future thinking [103] is used. The guidelines request details such as people present, sensory perceptions, the time, and the place where the event in memory occurred. The guidelines used are shown in the figure ??.

As indicated in psychology studies that investigate auto-nocuous consciousness and similarities between episodic memories and imagined events, we use the


**Describe a meaningful event from the last 24 hours, emphasizing on how you felt.**

[Click here for less info](#)

Choose any memory you wish, but please make sure it is something you are comfortable with and is not distressing.

Provide as much detail as you can about the event: the time, the place, with whom the event happened with, and sensory perceptions.

Describe any thoughts or emotions you had while the event was happening.




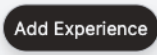
 

Figure 13: Adding a new memory to Resonance

same guidelines when the user is asked to input their imagination after the suggestions are displayed to them.

#### 4.2.2 Suggestion Generator

The purpose of the suggestion is to broaden the awareness of a person and how they can feel more positive emotions through an achievable action. To realize this purpose, given the memory for that day, the AI system generates a personalized, creative, and positively inclined suggestion for the user to act upon. Each of these principles is guided using prompt engineering and is explained below.

##### *Personalization of the suggestion*

The suggestions need to be personalized to make it achievable for the person while reminding them of past positive moments for gratitude. To make it personal, the system retrieves the five most semantically related memories of that particular user. The semantic similarity is calculated by comparing the cosine similarity of the Ada-2 embeddings of each memory. The five most similar memories are formatted into the prompt for generating the suggestion.

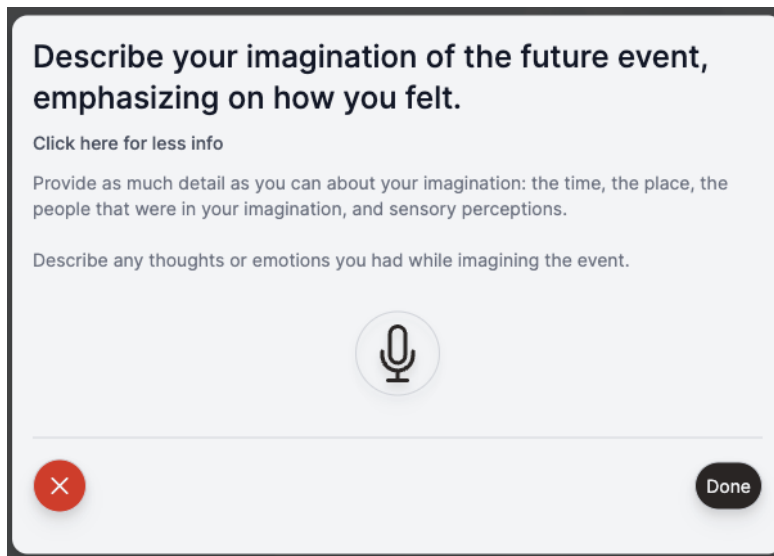


Figure 14: Adding a new imagination to Resonance

The prompt is also designed to have the suggestion refer to and explicitly cite the past memories if they are relevant to the suggestion.

### ***Creativity of the suggestion***

Creativity of suggestions was important to help broaden the awareness of person as to what actions can be done in daily life to feel more positive emotions and happiness. The creativity was induced by adding all the past suggestions to the prompt and designing the prompt to ensure the new suggestion has not already been suggested.

### ***Positive inclination of the suggestion***

A core aim of the suggestion is to indicate to the user an action that could be taken to increase the feeling of positive emotions. To do this effectively, a chaining of large language models was used. First, the new memory was used to generate a positive emotion that could be felt given that memory. This targeted positive emotion was then used to generate an explicit suggestion that would help feel that positive emotion if the action was taken. The separation of steps acts as a guardrail to reduce the likelihood of giving suggestions for negative emotions to the user.

## 4.3 USER STUDY

A two-week online study was conducted to evaluate the effects of Resonance on people. Participants were asked to use the tool every day for the two-week study duration.

### 4.3.1 Participants

Participants were recruited through the MIT Behavioural Research Lab which has a curated list of participants from across the United States. 71 participants started the study of which 55 participants successfully completed all the required components of the 2-week study. Completion was determined through submissions of pre and post-study surveys and at least 80% of daily entries. The demographic distribution of the 55 participants was: 22 male, 31 female, 2 non-binary, age range = 18 to 57,  $age_{mean} = 27.4$ ,  $age_{SD} = 6.5$ . Participants rated their journaling frequency between 'Never' (11), 'Rarely' (20), 'Once a week' (7), 'Several times a week' (7) and 'Daily' (10). Participants also rated how useful they perceive journaling as between 'Not at all' (2), 'Slightly' (8), 'Moderately' (19), 'Very' (13), and 'Extremely' (13). The study was reviewed and approved by the university ethics board.

### 4.3.2 Conditions

The conditions were designed to study the effects of Resonance during the two-week study. As journaling has been shown to help with a person's wellbeing, we had a control condition that acted as the frame of comparison. The control condition had the exact same interface and onboarding. The only difference was the absence of suggestions and imagination of the suggestions.

### 4.3.3 Procedure

Each participant signed up with their credentials to create individual accounts and used that throughout the study to interact with the web application.

#### *Start of Study Surveys and Onboarding*

Upon obtaining consent, a set of surveys was administered through an online platform to get demographic data and assess the baseline state of each par-



ticipant. The baseline included measurements of savoring beliefs (SBI), and depression scores (PHQ8) through the standard surveys [54, 9].

When a participant signs up they are required to go through an onboarding phase where they are prompted to input five seed memories covering a range of topics such as a travel experience to a family tradition. The seed memories act as a starting point for the AI suggestions for the first days of the study. In addition, the seeded memories give a window into non-recent memories for the participant which would not be available through the daily memories, which are intended to be from the past 24 hours. The questions for the seed memories are in the Appendix A.4.

### *Daily Memories*

On each day of the study, the participant is required to input a meaningful memory from the past 24 hours. A minimum of one was instructed while they were allowed to input more than one if they wanted to. To keep them as a singular memory of a short time slice, the participants were prompted to input a particular event along with episodic elements such as the people present, sensory perceptions, time and place of the event following the guidelines from [103], also available in the Appendix A.5. The memories were inputted through voice and participants could edit the transcriptions before submitting them. A minimum of 200 characters was needed to submit a memory.

### *Daily Surveys*

To get further insight into the daily mood of a participant, each time a participant started inputting a memory into the application, a two-question survey was administered to gauge the 5-point Likert scale levels of how positive and negative they were feeling at that moment. After the memory (and AI interaction if applicable) was done, the participants filled in the same two-question survey about their mood. Additionally, if in the experimental condition, the participant filled in a one-question 5-point Likert scale of how likely it is that they would act on that suggestion.

### *End of Study Surveys*

After two weeks of daily memories, and interacting with the AI in the experimental condition, each participant was asked to fill out online surveys to measure their current state. The surveys were the same as the baseline measurements at the start of the study including SBI and PHQ8. Additionally,

the participants of the experimental condition were asked to fill in surveys to gauge the efficacy of the suggestion and imagination features of the AI interface. Finally, participants from both conditions were asked to give open-ended feedback on the overall tool.

## 4.4 RESULTS

We show the analysis of data collected from the user study on the changes in mental health scores and user perceptions and feedback on the individual features of the interface. The participants were evenly divided into two conditions with  $N=28$  in the experimental condition and  $N=27$  in the control condition. A total of 275 onboarding seed memories and 754 journaled memories were collected from the 55 participants. 390 of these journal memories had corresponding AI suggestions and imagination entries as they came from participants in the experimental condition. The results are further discussed in relevance to the research questions in the Discussion chapter.

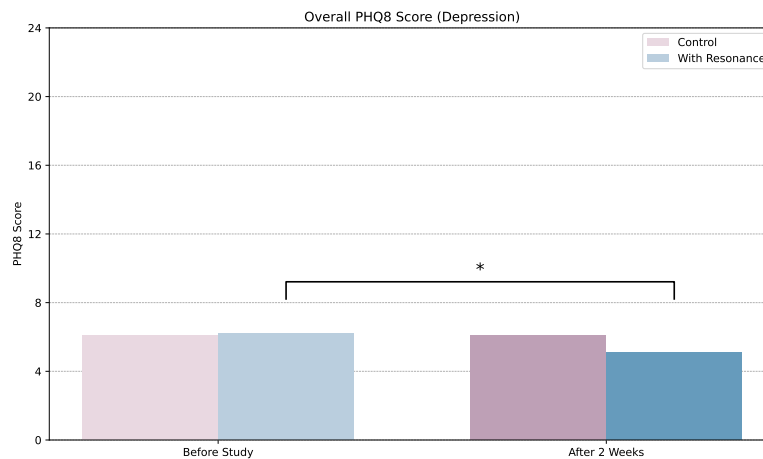
### 4.4.1 Changes in Mental Health

The survey reports corresponding to mental health were PHQ8 surveys [54] at the beginning and end of the study duration. A within-subject analysis was performed for the experimental and control conditions to evaluate if there were effects on these scores.

As the data was not normally distributed, as measured by the Shapiro-Wilk test, a Wilcoxon signed-rank test was conducted. The PHQ8 scores for the control condition did not change ( $M = 6.107$ ,  $SD = 5.094$  before the study and  $M = 6.107$ ,  $SD = 5.181$  after the study,  $p = 0.744$ ). Comparatively, the PHQ8 scores for the experimental condition changed from  $M = 6.214$ ,  $SD = 4.425$  before the study to  $M = 5.107$ ,  $SD = 3.92$  after the study. This difference was statistically significant with  $p = 0.011$ . The changes in PHQ8 scores are shown in figure 15.

The individual differences in the control and experimental conditions were tested with a between-subject analysis using the Mann-Whitney U test and were not statistically significant with  $p = 0.123$ .

The daily surveys measured the positive and negative mood of a participant using a 5-point Likert scale to the question "How positive/negative are you feeling right now?". They were collected before and after a memory was added



**Figure 15:** Reduction in PHQ8 scores for people using Resonance during the study period

for the control condition, and before the memory was added and after the imagination for the experimental condition. A within-subject analysis was performed for the experimental and control conditions to evaluate if there were effects on these scores.

As the data was not normally distributed, as measured by the Shapiro-Wilk test, a Wilcoxon signed-rank test was conducted. The positive scores for the control condition increased from  $M = 3.302$ ,  $SD = 1.074$  before inputting the memory to  $M = 3.440$ ,  $SD = 1.046$  after inputting the memory. The negative scores for the same decreased from  $M = 2.006$ ,  $SD = 0.930$  before inputting the memory to  $M = 1.937$ ,  $SD = 0.926$  after inputting the memory. These differences were statistically significant with  $p = 0.0001$  and  $p = 0.042$  respectively. Comparatively, the positive scores for the experimental condition increased from  $M = 3.117$ ,  $SD = 1.131$  before inputting the memory to  $M = 3.168$ ,  $SD = 1.101$  after inputting the imagination. The negative scores for the same decreased from  $M = 1.967$ ,  $SD = 1.048$  before inputting the memory to  $M = 1.874$ ,  $SD = 0.980$  after inputting the memory. The positive difference was not statistically significant with  $p = 0.235$ , while the negative difference was statistically significant with  $p = 0.008$ . The changes in PHQ8 scores are shown in figure 16.

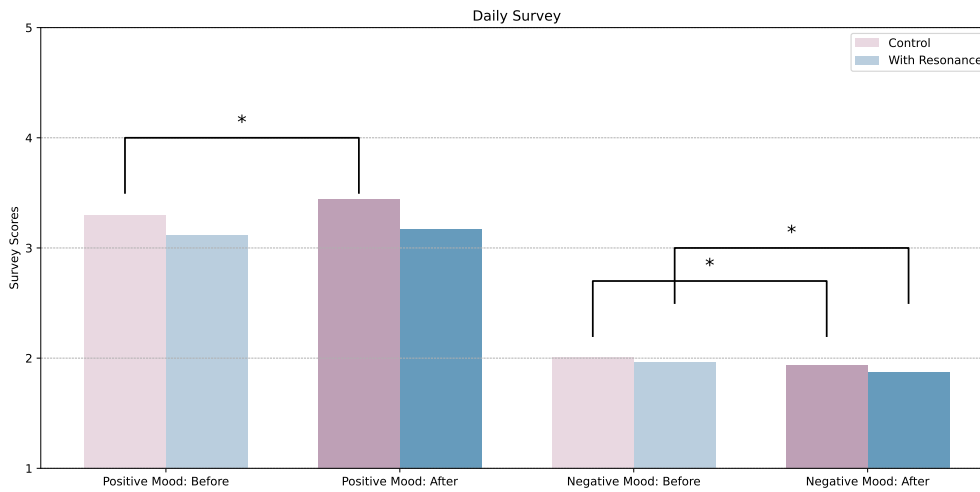


Figure 16: Changes in mood when entering daily memory and interacting with AI

#### 4.4.2 Perceptions of Interaction with the AI System

After every AI suggestion, the participants in the experimental condition rated a 5-point Likert scale of "How likely they are to act on the suggestion?" to measure the relevance of the suggestion to the participant.

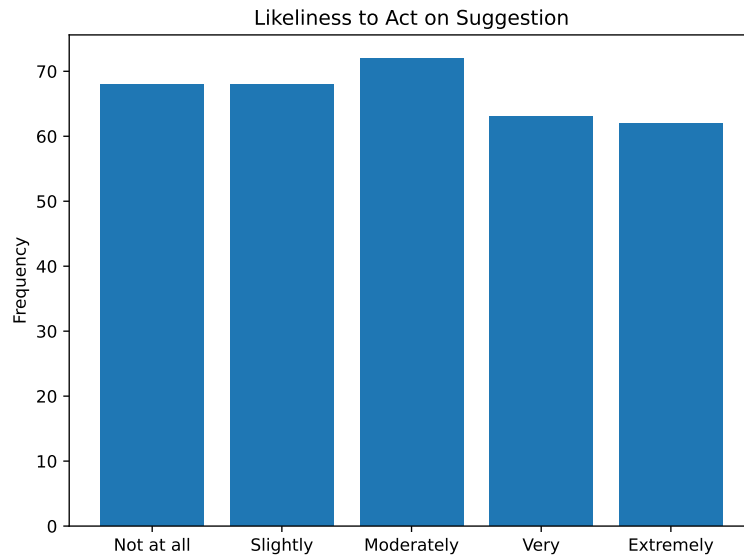
Additionally, at the end of the study, the participants in the experimental condition filled in surveys about their experiences with respect to the AI suggestions and their corresponding imaginations.

##### *AI suggestions*

There was a total of 333 responses to the likeliness to act on a suggestion which was uniformly divided across the 5 options of "Not at all" (68), "Slightly" (68), "Moderately" (72), "Very" (63) and "Extremely" (62). They are shown in figure 17.

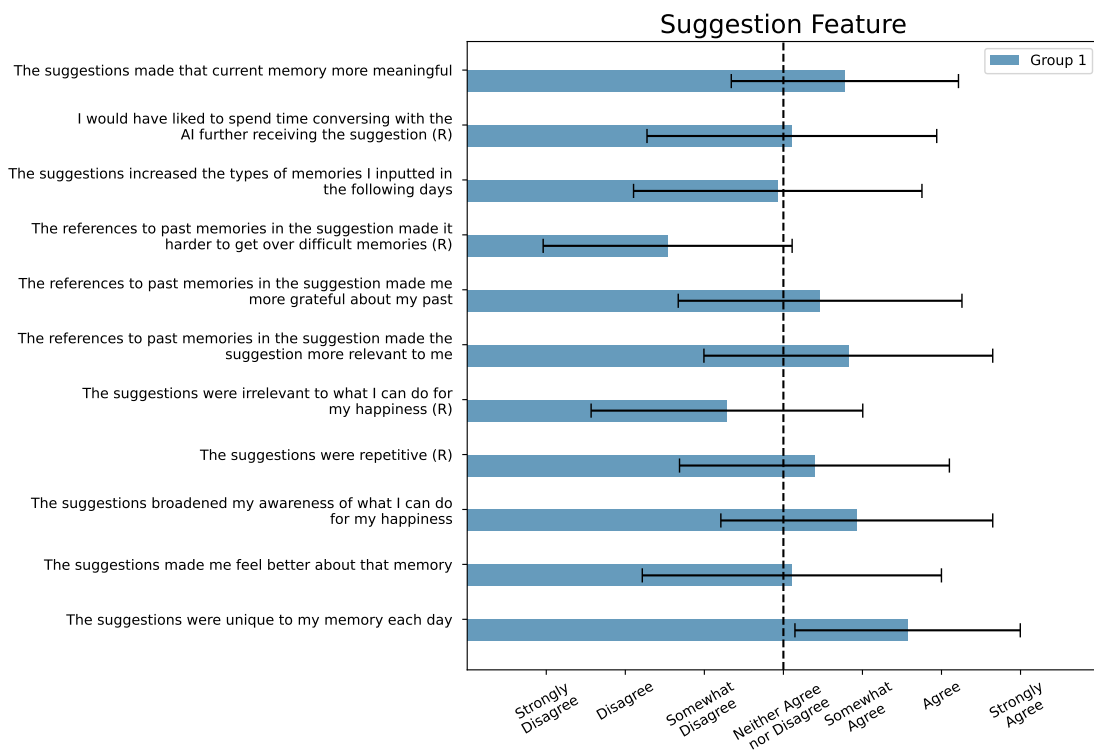
The following measures were taken at the end of the study on a 7-point Likert scale of the AI suggestions based on the journal memory and shown in figure 18.

- *The suggestions were unique to my memory each day*  $M = 5.57, SD = 1.43$
- *The suggestions made me feel better about that memory*  $M = 4.11, SD = 1.89$
- *The suggestions broadened my awareness of what I can do for my happiness*  $M = 4.93, SD = 1.72$



**Figure 17:** Likeliness to act on AI suggestions

- *The suggestions were repetitive (R)  $M = 4.39, SD = 1.71$*
- *The suggestions were irrelevant to what I can do for my happiness (R)  $M = 3.29, SD = 1.72$*
- *The references to past memories in the suggestion made the suggestion more relevant to me  $M = 4.82, SD = 1.83$*
- *The references to past memories in the suggestion made me more grateful about my past  $M = 4.46, SD = 1.79$*
- *The references to past memories in the suggestion made it harder to get over difficult memories (R)  $M = 2.54, SD = 1.57$*
- *The suggestions increased the types of memories I inputted in the following days  $M = 3.93, SD = 1.82$*
- *I would have liked to spend time conversing with the AI further receiving the suggestion (R)  $M = 4.11, SD = 1.83$*
- *The suggestions made that current memory more meaningful  $M = 4.78, SD = 1.44$*



**Figure 18:** Perceptions of participants about the AI suggestions

### ***Imagination of AI suggestions***

The following measures were on a 7-point Likert scale about the imagination of AI suggestions and shown in figure 19.

- *Imagining the suggestions made me feel better about that memory*  $M = 4.46$ ,  $SD = 1.86$
- *Imagining the suggestions made it more likely that I will act on it*  $M = 4.29$ ,  $SD = 2.05$
- *Imagining the suggestions was a waste of my time (R)*  $M = 3.57$ ,  $SD = 1.93$
- *I enjoyed imagining a suggestion even if I was not going to act on that suggestion*  $M = 4.32$ ,  $SD = 1.81$
- *If I wanted to act on the suggestion, then I would have preferred to not have to imagine it (R)*  $M = 3.43$ ,  $SD = 1.85$
- *The references to past memories in the suggestions helped me imagine the suggestion more vividly*  $M = 4.43$ ,  $SD = 1.95$
- *Imagining the suggestions increased the types of memories I inputted in the following days*  $M = 3.82$ ,  $SD = 1.87$
- *Imagining the suggestion made me look forward to it more, if I was to act on it*  $M = 4.18$ ,  $SD = 1.85$

### ***Correlation Analysis with Likelihood to Act***

A Spearman correlation analysis was conducted on the mean likelihood to act on suggestions for each user with their self-reported perceptions on the different aspects of reading and imagining the AI suggestions. The figures 20 and 21 show the correlation values.

#### **4.4.3 General Feedback**

At the end of the study period, participants in the experimental condition were asked the following open-ended questions to collect feedback on certain aspects of the tool, they were the following:

1. Do you have comments about getting the AI suggestions?

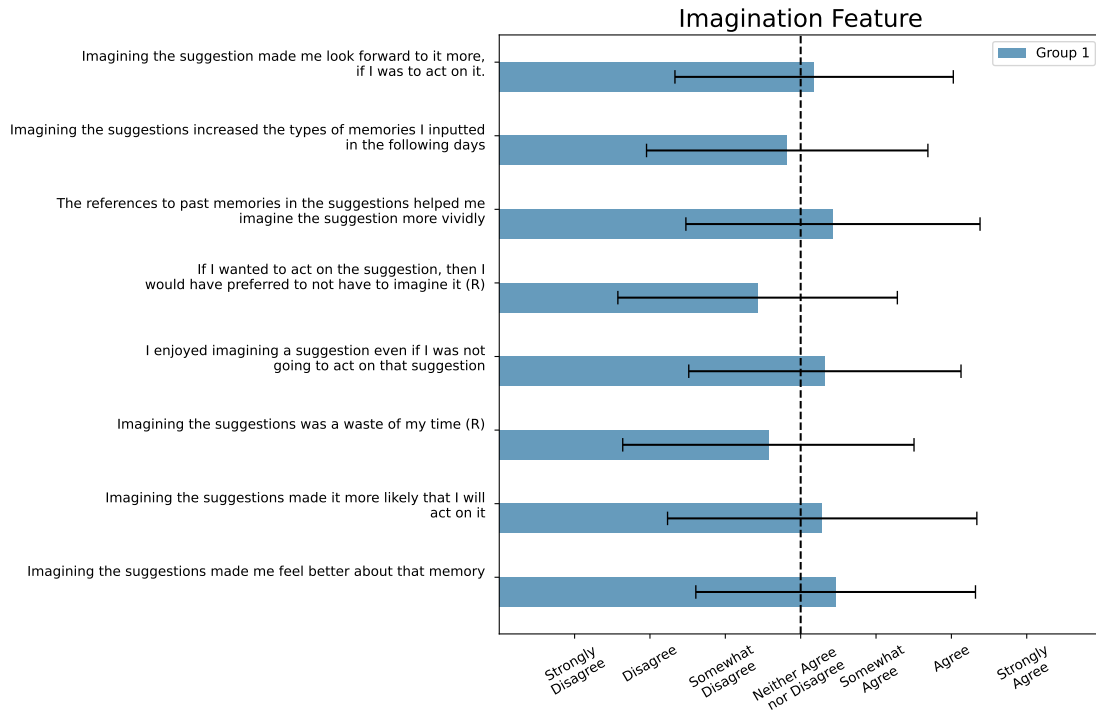


Figure 19: Perceptions of participants about the AI-guided imaginations

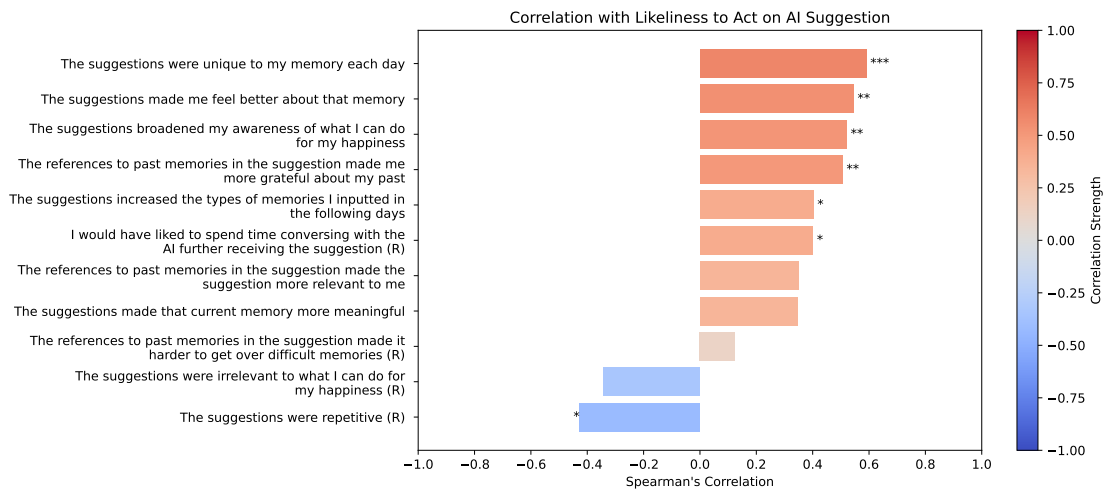


Figure 20: Correlations between mean likeliness to act of each participant and self-report perceptions of receiving the AI suggestions



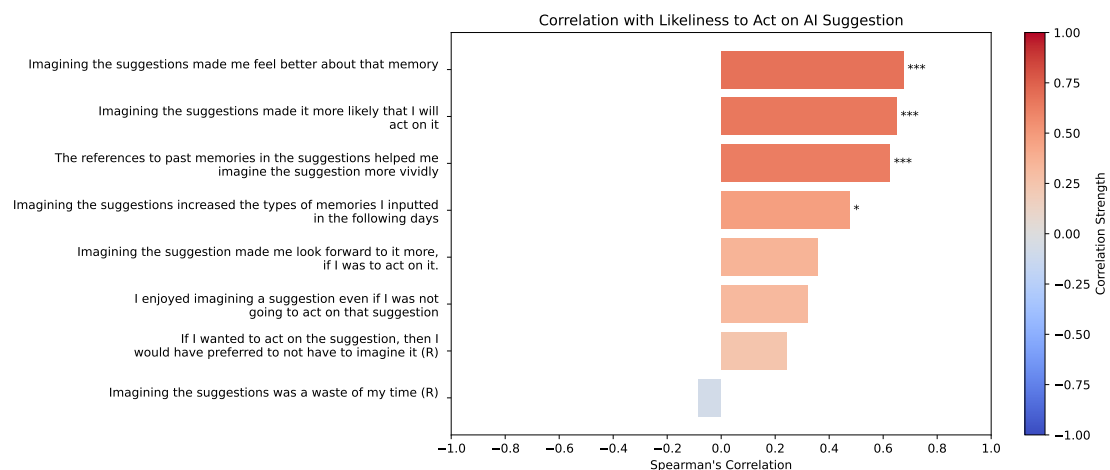


Figure 21: Correlations between mean likeliness to act of each participant and self-report perceptions of imagining the AI suggestions

2. Do you have comments about having to imagine the AI suggestions?
3. What did you like about the tool?
4. What were some of the concerns?

On organizing the feedback and coding them, the following themes emerged with specific quotes for each of them:

#### *Helpfulness of AI Suggestions*

- "AI suggestion are very helpful"
- "suggested healthy ways to connect with myself and others."
- "suggestions themselves sometimes make me feel better"
- "considerably better when responding to a positive memory as opposed to a negative memory"

#### *Unrealistic and Less Personal Suggestions*

- "Like "baking froyo"
- "too bossy-sounding or task-oriented"
- "For a small event/win, it suggested to celebrate way too much"

### ***Benefits of Imaginations***

- "imagined scenarios would make me laugh"
- "It strengthened the memory for the day"
- "may have shaped the areas in my life I want to be more aware of"

### ***Optional Imagination***

- "imagination section be optional should I choose to move forward with the suggestion"
- "imagination section subjects people to think about areas of their life which may be particularly difficult"
- "immediately start rationalizing why that situation would never happen, so it was hard for me to imagine the scenarios"
- "couldn't select what you wanted to imagine"

### ***Concerns Around Privacy***

- "speak out loud about these memories and events was really difficult because I live in shared spaces and have little privacy"
- "privacy is the largest concern"
- "hard to use when share living spaces"

### ***Preferences towards Voiced Entries***

- "I liked imagining while talking rather than imagining while writing"
- "audio transcription made things easier"
- "verbalize my reflection, it made it more meaningful than writing"

### ***Appreciation of Memory Titles***

- "summary titles were interesting."
- "short but relevant title"

# 5 | DISCUSSION

In this chapter, we discuss how the user studies for Memoro and Resonance addressed their respective research questions.

## 5.1 MINIMALLY DISRUPTIVE INTERFACES FOR REAL-TIME MEMORY AUGMENTATION

With the user study of 20 participants using Memoro in semi-spontaneous conversations, we discuss the results for the following research questions on the integrations of LLMs into wearable memory augmentation, the effect of using Memoro, and user experiences and perceptions of it.

### 5.1.1 Integrating LLMs in Wearable Memory Augmentation

Discussing RQ1: “How can we design a seamless wearable memory assistant using LLMs to reduce disruption to the primary task with minimal and effective input and output?” As recent advances in LLMs lead to improved capabilities in natural language processing tasks such as question answering and summarization, we found that using them in a wearable can facilitate a concise and seamless interface. It can be helpful to users for memory retrieval as all system conditions including baseline had a “helpfulness” rating above 5.15 of 7. We found that our approach of introducing minimal output from Memoro using LLMs reduced perceived disruption/interruption (Baseline  $M=5.55$ , Query  $M=4.40$ , and Queryless  $M=4.65$ , out of 7) while preserving their helpfulness. The use of LLMs in the semantic search of memories also showed that they could improve flexibility in querying by allowing users to use synonyms or alternate phrasings. This contributed to the highly rated adaptiveness of Memoro for both modes (Query  $M=5.35$  and Queryless  $M=5.10$ , out of 7) to the conversation and significantly higher ratings compared to Baseline. Through the Queryless Mode, we also demonstrate that LLMs can handle understanding user intentions in memory retrieval tasks during a conversation for minimal

input. The conciseness of output was significantly improved methodologically with an 85% reduction in answer length, and users rated them as having improved appropriateness of response length as compared to the Baseline condition. Overall, through the two modes of interaction of Memoro, we show a method of using LLMs for a concise interface in memory retrieval by providing flexibility in queries, parsing conversational context, and improving conciseness in responses.

### 5.1.2 Impact of Using Memoro in a Primary Task

Discussing RQ2: “What are the effects of using the memory augmentation system during the primary task of a real-time conversation across metrics such as quality of conversation, performance, and task load?” The emphasis on minimal disruption as being a core design principle for Memoro was to enable seamless interactions by users with their external memories while being preoccupied with a primary task, such as a conversation. Further discussing RQ2, in our study with social interactions, we validated that using Memoro did not affect the conversational quality in terms of attention, concentration, eye contact, or how relaxed they were as compared to when they used no system. The only aspect that was affected was that the conversations felt more natural with the No System condition compared to the system conditions. Along with this, participants showed a significant increase in recall confidence, a significant decrease in difficulty in recalling answers, and a significant increase in the amount of relevant information recalled during both modes of Memoro. The use of Queryless mode also resulted in a significant decrease in task load compared to the No System condition, making the conversation task cognitively easier for the user.

### 5.1.3 Usability, Preferences, and Experiences

Discussing RQ3: “How do context awareness and conciseness affect the system’s usability, user perceptions, and experience?” Overall, on evaluating the usability of Memoro, we find that the highest mean SUS score is for the Query Mode (80.0), followed by Queryless Mode (77.1). By adding contextual awareness and conciseness to the responses, there was a significant improvement in the usability from the Baseline LLM (68.8) condition. The SUS score of 80.0 lies in between the good and excellent range and is considered acceptable as it is well above the average score of 68 [6]. This was further reflected in user

preferences where 19 out of 20 participants rated a mode of Memoro over the Baseline and participants also mentioned that they would rather have no system and rely on their own memory over Baseline mode (Section 5.5). When analyzing the SUS scores for Memoro, previous work on comparing interfaces for Internet-of-Things (IoT) device manipulation during conversations showed that voice interfaces only achieved the mean SUS score of 70.88 [12] compared to a visual head-mounted display with a score of 83. One of the reasons for the longstanding issues with voice interfaces [16] is the accuracy of speech-to-text recognition. Although the recognition tool for Memoro and the previous study's tested voice interface was the same: Google Speech-to-Text API (Google Assistant), Memoro received higher usability scores and this might have been due to the use of LLMs to "offset" the inaccuracy of the speech-to-text. These findings indicate that an important consideration in designing wearable memory retrieval assistants is to enable the users to ask brief questions and get concise and to-the-point answers. Our findings can inform further work on integrating LLMs into the wearable context.

While Query mode was the most usable and the most accurate (Section 5.2.2), Queryless Mode was the most preferred condition among the participants (10 out of 20). From the NASA RTLX scores, an explanation for these preferences could be the significant decrease in task load when using the Queryless Mode compared to using no system in the task. In addition to its good usability and accuracy, we argue that there is value in the Query mode too as it had significantly higher rated usefulness (Section 5.3.4) and felt more polite to use compared to the Baseline (Section 5.3.5). Further, on examining the participants who preferred 'No System' over any of the other conditions, hence preferring no memory assistance (P7, P16, P18, P19), we found that two of them (P7, P16) rated their auditory memory as 'Extremely Good'. They were the only two participants with that rating in the study. The other two (P18, P19) indicated that they have never used voice assistants in their daily life. This aligns with previous studies [92] that people perceived increased benefits of voice assistants if they had used them before. These preferences indicate the need for more research into the influence of these factors in the design of wearable memory assistance.

## 5.2 HOLISTIC INTERFACES FOR REFLECTIVE MEMORY AUGMENTATION

With the user study of 55 participants over a 2 week period of daily usage, there emerged the following results related to the research questions of Resonance around the mental health effects, the perceptions of the interface features such as the AI suggestions and imaginations, and finally the experiences and open-ended feedback from the participants

### 5.2.1 Reduction in Current Depression by Using Resonance Daily

Discussing RQ1: “How does interacting with Resonance daily over a period of time affect the mental wellbeing of a user?”

As engaging in savoring has been shown to improve mental health and reduce depression, Resonance uses an AI interface to help people be reminded of relevant past positive experiences, savor the current experience (the journaled memory), and imagine future experiences of a positive suggestion from the AI system. While a daily session with Resonance lasts for less than five minutes, the repeated use of such a feature everyday can have compounding effects on the mental wellbeing of a person. Through the two-week user study period, we observe a significant reduction in the PHQ8 scores [54], a standard measure of current depression, from  $M = 6.214, SD = 4.425$  before the study to  $M = 5.107, SD = 3.92$  after the study. To test if the change was due to the AI interface, there was an identical condition that served as the control where all the features except the AI suggestions and imagination were present in the web application. The control condition had no significant difference in the PHQ8 scores of the participants. This highlights the effect of having Resonance’s AI features for the improvement in the mental wellbeing of the participants.

### 5.2.2 Perceptions about the AI Suggestions and Imaginations

Discussing RQ2: “How do users perceive reading and imagining the AI suggestions, and how does it influence mental wellbeing?”

Overall, the participants had a positive outlook toward the AI suggestions and imagination features. There was a uniform distribution of likeliness to act on the suggestion across the 5-point Likert scale with  $std =$ , and the scores on likeliness to act were significantly positively correlated with increase in positive

mood ( $\rho = 0.26, p = 0.001$ ) and decrease in negative mood for individual sessions ( $\rho = -0.11, p = 0.03$ ). On further analysis to investigate what makes suggestions more likely to act, there were significant positive correlations found between mean likeliness to act across the study period and if participants felt that the suggestions were unique to that memory ( $\rho = 0.59, p = 0.001$ ), if they broadened their awareness about what they can do for their happiness ( $\rho = 0.52, p = 0.004$ ), if they made them more grateful about the past ( $\rho = 0.51, p = 0.006$ ).

The imagination feature was included to increase the possibility of the likeliness to act, and to increase benefits from anticipating future positive experiences. This was supported by the significant positive correlation between mean likeliness to act and if imagining helped them to act on it ( $\rho = 0.65, p = 0.002$ ) and if the past memories referenced helped them imagine more vividly ( $\rho = 0.62, p = 0.004$ ).

### 5.2.3 Preferences and Experiences

Discussing RQ3: “What themes emerge from user feedback on the strengths and weaknesses of Resonance?”

Based on the open-ended feedback collected on the subjective experiences of participants using Resonance, a thematic analysis was conducted to understand what worked well and what can be improved in such tools. Broadly, participants felt that the AI suggestions were helpful but also could be irrelevant and unrealistic. While the helpfulness was conditional it performed better for positive memories than negative memories where more emotional validation and a lack of a human touch were experienced. Unrealism occurs in situations that likely stem from stochasticity and the insufficient understanding of the world model in our chosen LLM. Imagining the suggestions reinforced the memory, broadened awareness, and was enjoyable for a few participants. Some noted there was some difficult situations or unrealistic situations where imagination was not feasible. In a similar angle, a few participants requested for more flexibility such as optionally imagining the suggestions.

Moreover, participants seemed to prefer to verbalize their memories and imaginations as opposed to writing as it was more meaningful, easier, and faster. The participants who preferred also being able to type their memories cited confidentiality and shared living spaces as reasons. The generated titles were also largely preferred by participants who found them interesting, personalized, and relevant. Finally, a few participants expressed their concern about the

privacy of digital journaling, especially with AI involved which hindered the sharing of personal details.



# 6 | LIMITATIONS

We discuss the limitations of developing memory augmentation systems using AI in terms of the technical aspects, study designs used to evaluate the systems, and the social implications.

## 6.0.1 Technical Aspects

Firstly, the encoding of external memory in Memoro is based on timestamps and direct transcription of the recording of audio, inspired by existing lifelogging tools [95, 37], and as the focus of the study was to explore minimally disruptive memory retrieval during a primary task. Integrating more information such as location, non-verbal gestures, facial expressions, and recognition of the conversation partner during memory encoding, can significantly advance the memory assistant by understanding more of the user’s context [84, 15]. The location (from the GPS sensor) and conversation partner information can assist in filtering older memories for accurate retrieval. Non-verbal gestures can give insightful information on body language such as low engagement or heightened nervousness which can increase the importance of the memories encoded during that period. The importance could be further modulated by users explicitly. These features can enable diverse queries of the form *“Who did I meet in the cafeteria yesterday?”* or *“What was the name of the person Ann spoke to me about 2 days ago?”*. Further, implicit prompting based on disfluencies in speech, and accelerometer-based gestures can reduce input effort and time by having users perform subtle hand gestures instead of clicking the trigger button. Implicit prompting can lead to studies understanding how short the query needs to be for a conversation to seem “uninterrupted” from an external perspective.

Secondly, the use of LLMs in Memoro and Resonance for information retrieval can lead to erroneous answers that do not exist in the database. The memories can also contain conflicting information which can lead to incorrectly generated suggestions. While tackling erroneous answers in LLMs is an ongoing challenge, future work can address these concerns with a more sophisticated knowledge graph of the user’s memories.

Thirdly, while we look at discreet audio feedback from the system to maintain eye gaze and reduce distraction during conversations, we acknowledge that there is a chance of the masking of the conversation with sound coming from the voice interface and voicing queries (for Query Mode) might disrupt the conversation. The timing for receiving the audio feedback is determined by the user, as such, users can choose to trigger retrieval during breaks between sentences (for Query and Queryless modes) or potentially mask queries within the conversation such as by rephrasing the conversational partners' questions (for Query mode). Some users may prefer an Optical head-mounted display (OHMD) for visual feedback. For users who prefer OHMD, a similar assistant with visual answers could be given where the text-to-speech response can be skipped. A study evaluating the pros and cons of audio-based versus heads-up display-based interaction in memory assistance would be an interesting next step.

Finally, for Resonance, we asked the users to journal through their voice and had a few participants who dropped out citing the reason as preferring to type their journal entries. The choice of voice-based input was to encourage users to go deeper into the episodic elements of their memories and input longer entries, which was validated by a few participants in their feedback. Future work can allow the option of typing along with voice for the memories to give flexibility.

### 6.0.2 Study Design and Population

For the user study for Memoro, the participants were from a group from the local community who may be more accustomed to such technology as voice assistants. The experiment also was situated in a lab setting for a controlled study. Longitudinal and in-the-wild studies situated in natural settings with a geographically diverse user group while enabling both retrieval modes simultaneously are needed to understand the usefulness and applicability of Memoro in daily life outside of laboratory-based social interactions. Relatedly, longitudinal studies can employ text similarity algorithms to aid in the objective measurement of the recall ability of users. Similarly, future directions include field studies with a specific subpopulation with a higher frequency of memory assistance needs, such as the elderly, where such a system could be more useful. An example is the tip-of-the-tongue (TOT) scenario [7], forgetting of certain words, which commonly occurs in older adults and increases with Aphasia. There can be an exploration of other forms of information presentation where, instead of giving direct answers, the system would give users episodic or se-

semantic clues and optional answers [34], or answers in voices of people you admire or are familiar with [13].

With respect to the user study of Resonance, follow-up interviews with the participants could elicit deeper feedback on the preferences and concerns of such a tool. Further, while the post-study surveys were filled after 24 hours of using the tool, another follow-up survey after a week or two could reveal how sustained the effects of Resonance are on mental wellbeing. Finally, the novelty of such tools tends to have a positive effect in the initial days. A longer study spanning over months could have day one of the study after a week of familiarizing with the tool. Additionally, the longer study could also investigate the long-term effects of using Resonance. For the evaluation of the system, the choice of only two conditions with one of them being the control to explore if such a system could be used easily by participants and has some effect. As the effects on wellbeing were validated in the study, future work could include more conditions such as only receiving AI suggestions, without imaginations and having the choice to imagine the suggestion, to reveal the nuanced effects of these design features of the interface.

### 6.0.3 Social Implications: Privacy and Social Acceptability

Finally, it is important to consider legal, ethical, privacy and social acceptability issues in deploying memory assistants that record audio from everyday conversations and intimate memories and events from daily life. Ensuring data security for pervasive memory augmentation systems is critical beyond ensuring encrypted data storage [24]. As research in psychology [2] shows how we are prone to the simultaneous reinforcement of recovered memories and attenuation of unrecovered memories, memory augmentation interfaces can contribute to unintended altering and manipulation of captured memories through its process of retrieval. With the increase in the subtleness of wearables with recording capabilities [43], future memory augmentation systems need to implement concrete and transparent methods, such as speaker verification [80], to manage permissions of recording. As this system is geared for daily use, the privacy of bystanders in the vicinity needs to also be accounted for. Further, in some states and countries, recording other people without their knowledge is illegal. While this work assumes consent for recording from all parties involved, possible methods to address privacy controls in natural settings may be to record synthesized notes, rather than direct transcriptions, to require opt-in or enable opt-out, and be able to selectively erase data on request.

Social acceptability of lifelogging devices can be situational [24], where certain contexts such as during sports and meetings can be more permissive to it as compared to intimate conversations and in healthcare settings. Cultural beliefs and user stereotypes could also shape the social perceptions of wearables and user-worn recording devices [86, 49, 39]. Bystander considerations also play a role in social acceptability where interactions that provide an explanation [100] are likely to be better acceptable than fully hidden interactions. Future research efforts should focus on designing strategies to improve social acceptability, possibly following guidelines in social acceptability research in HCI [53].

# 7

## ETHICAL CONSIDERATIONS

Episodic memory is an integral part of human cognition and underpins an array of higher-level cognitive processes such as decision-making, learning, and maintaining relationships. Engaging with such a core aspect of cognition can lead to unintended and unforeseen impacts on the user. Loftus et al [60] have shown how malleable and unreliable human memory can be with leading questions and emotional bias. Further, [42] demonstrates that the construction of memory is weakest when it is in the process of being retrieved, making it the most suitable time for manipulation. This research explores different methods of retrieval aimed at improving well-being, but the consequences could also be to constantly remind people of traumatic experiences.

Further, logging and storage of daily auditory interactions could encompass sensitive content such as intimate conversations and critical information such as passwords. As with any form of digital data storage, the data repository is vulnerable to unauthorized access and cybersecurity breaches. Such risks must be communicated clearly to users when developing and distributing lifelogging tools.

Additionally, people repeatedly become over-reliant on technology for cognitive processes leading to negative consequences. Focusing on navigation tools like Google Maps, research has shown that using GPS can lead to poorer recall of routes, suggesting that reliance on such tools might be degrading certain aspects of our memory [96]. A study has indicated that reliance on smartphones can negatively impact our ability to think, remember, pay attention, and regulate emotions. [101] A memory augmentation tool could have a similar effect in reduced attention to conversations if used excessively.

Lastly, it is imperative to consider the potential for misuse of such personal archives. When intentions are misaligned, the memories stored could be exploited, with ramifications ranging from legal implications to interpersonal conflicts. The utilization of personal records is inherently tied to the objectives of the individual accessing them. This duality of lifelogging technologies is eloquently captured in Ted Chiang's narrative, "The Truth of Fact, the Truth of Feeling," [19] where the use of memory augmentation devices leads to increased self-reflection and positive behavioral change.

Our research aims to facilitate the constructive applications of episodic memory augmentation while mitigating its possible detrimental effects. Through thoughtful design and ethical guidelines, we seek to harness the potential for fostering well-being and memory support, while safeguarding against exploitation and privacy intrusions.

# 8

## CONCLUSION

Episodic memory is a core component of human cognition. It is a neural system that supports the management of episodic memories, which are summary records of sensory-perceptual-conceptual-affective processing of experiences. Augmenting episodic memory can have a wide range of benefits ranging from deeper interpersonal relationships, improved mental well-being, and enhanced creativity and problem-solving. While recent advances in artificial intelligence through large language models offer opportunities to have digital systems adapt to, assist, and augment episodic memory, it is critical to design suitable interfaces to ensure these systems enrich people's lives. In this thesis, we explore the design, development, and evaluation of two such interfaces, MEMORO and RESONANCE. While MEMORO offers real-time memory assistance through a wearable system that ensures minimal disruption to the user's task, RESONANCE provides reflective memory assistance to enable a user to be more deeply connected with their past and future self and to improve their mental wellbeing. This work highlights the importance and shows examples of systems powered by artificial intelligence centered around enriching people's real-life experiences by augmenting their episodic memory.

# A | APPENDIX

## A.1 EXAMPLE QUERIES

### A.1.1 Baseline

**Query:** What were Sarah's favorite soccer teams?

**Answer:** Sarah's favorite soccer teams were Everton and Wrexham.

**Query:** How does Benjamin get to work from the Santa Fe bus depot?

**Answer:** Benjamin boards a commuter train from the Santa Fe Depot that takes him to the EarthGuard office located near the Sorrento Valley station.

**Query:** What does Emily's house look like?

**Answer:** Emily's house is a charming, renovated Victorian house in a vibrant neighborhood. It features four spacious bedrooms, intricately detailed walnut wooden staircases, and original and black stained glass windows. The house is surrounded by a lush tomato garden and an outdoor.

**Query:** What vegetable starts does, Benjamin grow in his garden.

**Answer:** Benjamin grows tomatoes, kale, and bell peppers in his garden.

**Query:** What does William teaches kids?

**Answer:** Benjamin boards a commuter train from the Santa Fe Depot that takes him to the EarthGuard office located near the Sorrento Valley station.

### A.1.2 Query Mode

**Context:** went to Bangladesh, Pictures of Bengal tiger,

**Query:**What else did she do?

**Answer:** Dhaka, Jatiya Sangsad Bhaban.

**Context:** pet's name is Luna. It's some sort of dog **Query:**What type of dog does she have?

**Answer:** Husky.

**Context:** cycling going to get some bus 36.

**Query:** How does he go to work after Bus 36?

**Answer:** Train.



**Context:** She's 7 years old I don't remember her name

**Query:** What was her name?

**Answer:** Sophie.

**Context:** San Diego Music Festival

**Query:** What is the place called again, at the bar?

**Answer:** Lestat's.

### A.1.3 Queryless Mode

**Context:** Hardware, sincerity and dedication, He made two apps. The names are

**Inferred Query:** What are the two apps that William Thompson created?

**Answer:** BuzzPal, FoodMingle

**Context:** now he's very fond of being outdoors. He likes hiking mountain biking and

**Inferred Query:** What other outdoor activities does William Thompson enjoy?

**Answer:** Camping, fishing.

**Context:** okay, her house has full Venture bedroom and also and

**Inferred Query:** What are the features of Emily Johnson's house, such as the walnut wooden staircases, blue stained glass windows, and outdoor seating area?

**Answer:** Walnut staircases, blue windows, outdoor seating.

**Context:** saw a scary tiger in the tiger is called, it's like a famous tiger called

**Inferred Query:** What is the name of the famous tiger that was seen in the Sundarbans?

**Answer:** Bengal Tiger.

**Context:** She likes to paint on a canvas, And use.

**Inferred Query:** What type of paints does Sarah Lee prefer to use for her paintings?

**Answer:** Acrylic.

## A.2 PROMPT TEMPLATES

### A.2.1 Contextual Query

You are a helpful assistant that provides memory cues to a human. The human is engaged in a conversation with another human, and asks you in the middle for assistance. The answer can be found in the relevant memories. If it is not found in the relevant memories, you should truthfully answer that you do not know the answer.

Relevant memories: *<External Memories>*

The current context contains the conversation between the two humans.

Current context: *<Current Context>*

The query is the question asked by the human to you.

Query: *<Query>*

Answer: *[Generated Answer]*

### A.2.2 Concise Suggestions

Make the answer more concise, such that it only contains the words needed to answer the query. It should NOT contain any information that is already present in the current context.

Current context: *<Current Context>*

Query: *<Query>*

Answer: *<Retrieved Answer>*

Concise answer: *[Generated Answer]*

### A.2.3 Queryless Search

You are an assistant interface between user and a memory system. The user is engaged in a conversation with another human, and asks you in the middle for assistance. The assistant frames a query that the user would like to ask the memory system next at the end of the conversation. The recent conversation between the two humans is related to

the relevant memories. The answer that the user would like to retrieve would not be in the recent conversation. The query should be very relevant to the end of the last sentence of the recent conversation.

Recent conversation: *<Current Context>*

What do you think that the user would like to ask the memory system to finish or clarify his last sentence?

Query: [Generated Query]

## A.3 QUESTIONNAIRES

### A.3.1 User experiences and Perception

We measured eight aspects using a 7-point Likert scale (1=strongly disagree, 7=strongly agree).

1. **Length of Responses:** "I felt that the length of the answers was appropriate."
2. **Adaptiveness of the System:** "I felt that the system adapted to my needs in the conversation."
3. **Interruption to Conversation:** "The device manipulation by me interrupted the conversation."
4. **Helpfulness of Response:** "The answers from the system were helpful."
5. **Usefulness:** "The system would be useful in my everyday life."
6. **Politeness:** "I felt it was polite to use the system during the conversation."
7. **Naturalness:** "I acted naturally at all times while focusing on the researcher's face and using the system."
8. **Ease of Ignoring the Device:** "It was easy to ignore the fact that I was wearing the device."

### A.3.2 Conversation Quality

We measured six aspects using a 7-point Likert scale (1=strongly disagree, 7=strongly agree). [12]:

1. **Listening to the Conversational Partner:** "When the other person was speaking, I was always listening to them."
2. **Concentration on the Conversation:** "I was always concentrating on the conversation."
3. **Attention Towards Conversation Partner:** "When I was speaking, my attention was towards the other person."
4. **Eye Contact:** "When I was speaking, I maintained eye contact."
5. **Naturalness:** "I acted naturally at all times during the conversation."
6. **Feeling Relaxed:** "I felt relaxed during the conversation."

#### A.3.3 Task Performance/Recall Ability

We measured three aspects using a 100-point slider scale. It involved:

1. **Confidence in Memory:** "I was confident in my ability to recall the information of the person while answering the questions."
2. **Difficulty in Recall:** "I found it difficult in recalling the information of the person."
3. **Recalled Relevance:** "I recalled all the relevant information of the person with respect to the question."

#### A.4 QUESTIONS FOR SEED MEMORIES

1. Describe a travel experience that deeply moved or changed you. What was it about this experience that was so impactful?
2. What has been the biggest challenge you've faced in life, and how did you overcome it? What did you learn from this experience?
3. Is there a particular cultural event or practice you've experienced that left a lasting impression on you?"
4. What is your most cherished memory from your childhood, and why does it stand out to you?

5. Can you recall a family tradition that you particularly loved? How did it shape your understanding of family?

## **A.5 PROMPT TO USERS FOR EPISODIC MEMORY JOURNALING**

### **A.5.1 Memory Prompt**

Describe a meaningful event from the last 24 hours, emphasizing on how you felt.

Choose any memory you wish, but please make sure it is something you are comfortable with and is not distressing.

Provide as much detail as you can about the event: the time, the place, with whom the event happened with, and sensory perceptions.

Describe any thoughts or emotions you had while the event was happening.

### **A.5.2 Imagination Prompt**

Describe your imagination of the future event, emphasizing on how you felt.

Provide as much detail as you can about your imagination: the time, the place, the people that were in your imagination, and sensory perceptions.

Describe any thoughts or emotions you had while imagining the event.

## BIBLIOGRAPHY

- [1] Timothy A Allen and Norbert J Fortin. The evolution of episodic memory. *Proceedings of the National Academy of Sciences*, 110(supplement\_2):10379–10386, 2013.
- [2] Michael C. Anderson, Robert A. Bjork, and Elizabeth L. Bjork. Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5):1063, 1994.
- [3] Max Angenius and Maliheh Ghajargar. Design principles for interactive and reflective journaling with ai. In *Science and Information Conference*, pages 62–81. Springer, 2023.
- [4] Mathias Benedek, Roger E Beaty, Daniel L Schacter, and Yoed N Kenett. The role of memory in creative ideation. *Nature Reviews Psychology*, 2(4):246–257, 2023.
- [5] Carlos Bermejo, Tristan Braud, Ji Yang, Shayan Mirjafari, Bowen Shi, Yu Xiao, and Pan Hui. Vimes: A wearable memory assistance system for automatic information retrieval. page 3191–3200, 2020.
- [6] John Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.
- [7] Roger Brown and David McNeill. The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5(4):325–337, 1966.
- [8] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal ..., and Dario Amodei. Language models are few-shot learners, 2020.
- [9] Fred Bryant. Savoring beliefs inventory (sbi): A scale for measuring beliefs about savouring. *Journal of mental health*, 12(2):175–196, 2003.
- [10] Fred B. Bryant and Joseph Veroff. *Savoring: A New Model of Positive Experience*. Lawrence Erlbaum Associates Publishers, 2007.
- [11] Vannevar Bush et al. As we may think. *The atlantic monthly*, 176(1):101–108, 1945.

- [12] Runze Cai, Nuwan Nanayakkarasam Peru Kandage Janaka, Shengdong Zhao, and Minghui Sun. Paraglassmenu: Towards social-friendly subtle interactions in conversations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–21, 2023.
- [13] Sam Chan, Tamil Selvan Gunasekaran, Yun Suen Pai, Haimo Zhang, and Suranga Nanayakkara. Kinvoices: Using voices of friends and family in voice interfaces. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–25, 2021.
- [14] Samantha Chan. Biosignal-sensitive memory improvement and support systems. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2020.
- [15] Samantha Chan. *Augmenting Human Prospective Memory through Cognition-Aware Technologies*. PhD thesis, ResearchSpace@ Auckland, 2022.
- [16] Samantha Chan, Shardul Sapkota, Rebecca Mathews, Haimo Zhang, and Suranga Nanayakkara. Prompto: Investigating receptivity to prompts based on cognitive load from memory training conversational agent. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(4), dec 2020.
- [17] Samantha Chan, Haimo Zhang, and Suranga Nanayakkara. Prospero: A personal wearable memory coach. In *Proceedings of the 10th Augmented Human International Conference 2019*, pages 1–5, 2019.
- [18] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3):1–45, 2024.
- [19] Ted Chiang. The truth of fact, the truth of feeling. *Subterranean Press*, January 2013.
- [20] Martin A Conway. Memory and the self. *Journal of memory and language*, 53(4):594–628, 2005.
- [21] Martin A Conway. Episodic memories. *Neuropsychologia*, 47(11):2305–2313, 2009.
- [22] Martin A Conway, John M Gardiner, Timothy J Perfect, Stephen J Anderson, and Gillian M Cohen. Changes in memory awareness during learning: the acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General*, 126(4):393, 1997.

- [23] Ishita Dasgupta and Samuel J. Gershman. Memory as a computational resource. *Trends in Cognitive Sciences*, 25(3):240–251, 2021.
- [24] Nigel Davies, Adrian Friday, Sarah Clinch, Corina Sas, Marc Langheinrich, Geoff Ward, and Albrecht Schmidt. Security and privacy implications of pervasive memory augmentation. *IEEE Pervasive Computing*, 14(1):44–53, 2015.
- [25] Richard W. Devaul and Alex P. Pentland. *The Memory Glasses: Wearable Computing for Just-in-Time Memory Support*. PhD thesis, USA, 2004.
- [26] Roy Dings and Albert Newen. Constructing the past: The relevance of the narrative self in modulating episodic memory. *Review of Philosophy and Psychology*, 14(1):87–112, 2023.
- [27] STEVEN DONAGHY and WENDY WILLIAMS. A new protocol for training severely impaired patients in the usage of memory journals. *Brain Injury*, 12(12):1061–1076, 1998.
- [28] Thomas SC Farrell. Teacher self-awareness through journal writing. *Reflective Practice*, 14(4):465–471, 2013.
- [29] Carol L Flinchbaugh, E Whitney G Moore, Young K Chang, and Douglas R May. Student well-being interventions: The effects of stress management techniques and gratitude journaling in the management education classroom. *Journal of Management Education*, 36(2):191–219, 2012.
- [30] John M Gardiner. Episodic memory and auto-noetic consciousness: a first-person approach. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356(1413):1351–1361, 2001.
- [31] Olga Gelonch, Mireia Ribera, Núria Codern-Bové, Sílvia Ramos, Maria Quintana, Gloria ... Chico, and Maite Garolera. Acceptability of a lifelogging wearable camera in older adults with mild cognitive impairment: a mixed-method study. *BMC geriatrics*, 19(1):1–10, 2019.
- [32] Tanya Goyal, Junyi Jessy Li, and Greg Durrett. News summarization and evaluation in the era of gpt-3. *arXiv preprint arXiv:2209.12356*, 2022.
- [33] Cathal Gurrin, Alan F Smeaton, Aiden R Doherty, et al. Lifelogging: Personal big data. *Foundations and Trends® in information retrieval*, 8(1):1–125, 2014.



- [34] Juhye Ha, Dayoung Lee, and Changhoon Oh. You know what i'm saying: Designing conversational strategies of ai agent for tip of the tongue phenomenon. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2023.
- [35] Morgan Harvey, Marc Langheinrich, and Geoff Ward. Remembering through lifelogging: A survey of human memory augmentation. *Pervasive and Mobile Computing*, 27:14–26, 2016.
- [36] Demis Hassabis and Eleanor A Maguire. Deconstructing episodic memory with construction. *Trends in cognitive sciences*, 11(7):299–306, 2007.
- [37] Gillian R. Hayes, Shwetak N. Patel, Khai N. Truong, Giovanni Iachello, Julie A. Kientz, Rob Farmer, and Gregory D. Abowd. The personal audio loop: Designing a ubiquitous audio-based memory aid. In Stephen Brewster and Mark Dunlop, editors, *Mobile Human-Computer Interaction - MobileHCI 2004*, pages 168–179. Springer Berlin Heidelberg, 2004.
- [38] Roger Hiemstra et al. Uses and benefits of journal writing. *New directions for adult and continuing education*, 2001(90):19, 2001.
- [39] Yi-Ta Hsieh, Antti Jylhä, Valeria Orso, Luciano Gamberini, and Giulio Jacucci. Designing a willing-to-use-in-public hand gestural interaction technique for smart glasses. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4203–4215, 2016.
- [40] Nick Hunn. Hearables—the new wearables. *Wearable Technologies*, 2014.
- [41] Fabian Hutmacher, Linus Schläger, and Rinat Meerson. Autobiographical memory in the digital age: Insights based on the subjective reports of users of smart journaling apps. *Applied Cognitive Psychology*, 37(4):686–698, 2023.
- [42] Maria Carmen Inda, Elizaveta V Muravieva, and Cristina M Alberini. Memory retrieval and the passage of time: from reconsolidation and strengthening to extinction. *Journal of Neuroscience*, 31(5):1635–1643, 2011.
- [43] Muhammad Zahid Iqbal and Abraham G. Campbell. Adopting smart glasses responsibly: potential benefits, ethical, and privacy concerns with ray-ban stories. *AI and Ethics*, 3(1):325–327, 2023.
- [44] Lilian Jans-Beken, Nele Jacobs, Mayke Janssens, Sanne Peeters, Jennifer Reijnders, Lilian Lechner, and Johan Lataster. Gratitude and health: An updated review. *The Journal of Positive Psychology*, 15(6):743–782, 2020.

- [45] Shiqi Jiang, Zhenjiang Li, Pengfei Zhou, and Mo Li. Memento: An emotion-driven lifelogging system with wearables. *ACM Transactions on Sensor Networks (TOSN)*, 15(1):1–23, 2019.
- [46] Helen G Jing, Kevin P Madore, and Daniel L Schacter. Worrying about the future: An episodic specificity induction impacts problem solving, reappraisal, and well-being. *Journal of Experimental Psychology: General*, 145(4):402, 2016.
- [47] Paul E Jose, Bee T Lim, and Fred B Bryant. Does savoring increase happiness? a daily diary study. *The Journal of Positive Psychology*, 7(3):176–187, 2012.
- [48] Daniel Kahneman. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.
- [49] Norene Kelly and Stephen B. Gilbert. The wearer, the device, and its use: advances in understanding the social acceptability of wearables. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 62, pages 1027–1031. SAGE Publications Sage CA: Los Angeles, CA, 2018.
- [50] Mina Khan, Glenn Fernandes, Utkarsh Sarawgi, Prudhvi Rampey, and Pattie Maes. PAL: A wearable platform for real-time, personalized and context-aware health and cognition support. *CoRR*, abs/1905.01352, 2019.
- [51] Zafar U Khan, Elisa Martín-Montañez, Irene Navarro-Lobato, and E Chris Muly. Memory deficits in aging and neurological diseases. *Progress in molecular biology and translational science*, 122:1–29, 2014.
- [52] Torkel Klingberg. *The overflowing brain: Information overload and the limits of working memory*. Oxford University Press, 2009.
- [53] Marion Koelle, Swamy Ananthanarayan, and Susanne Boll. Social acceptability in hci: A survey of methods, measures, and design strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2020.
- [54] Kurt Kroenke, Tara W Strine, Robert L Spitzer, Janet BW Williams, Joyce T Berry, and Ali H Mokdad. The phq-8 as a measure of current depression in the general population. *Journal of affective disorders*, 114(1-3):163–173, 2009.

- [55] Amel Ksibi, Ala Saleh D Alluhaidan, Amina Salhi, and Sahar A El-Rahman. Overview of lifelogging: current challenges and advances. *IEEE Access*, 9:62630–62641, 2021.
- [56] Eyal Kushilevitz, Rafail Ostrovsky, and Yuval Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 614–623, 1998.
- [57] M. Lamming, P. Brown, K. Carter, M. Eldridge, M. Flynn, G. Louie, P. Robinson, and A. Sellen. The design of a human memory prosthesis. *The Computer Journal*, 37(3):153–163, 01 1994.
- [58] Andrew K Lampinen, Ishita Dasgupta, Stephanie CY Chan, Kory Matthewson, Michael Henry Tessler, Antonia Creswell, James L McClelland, Jane X Wang, and Felix Hill. Can language models learn from explanations in context? *arXiv preprint arXiv:2204.02329*, 2022.
- [59] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal ..., and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks, 2021.
- [60] Elizabeth F Loftus and Jacqueline E Pickrell. The formation of false memories, 1995.
- [61] Yangmei Luo, Xuhai Chen, Senqing Qi, Xuqun You, and Xiting Huang. Well-being and anticipation for future positive events: Evidences from an fmri study. *Frontiers in psychology*, 8:2199, 2018.
- [62] Christopher R Madan. Rethinking the definition of episodic memory. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 74(3):183, 2020.
- [63] Kevin P Madore, Donna Rose Addis, and Daniel L Schacter. Creativity and memory: Effects of an episodic-specificity induction on divergent thinking. *Psychological science*, 26(9):1461–1468, 2015.
- [64] Kevin P Madore and Daniel L Schacter. Remembering the past and imagining the future: Selective effects of an episodic specificity induction on detail generation. *Quarterly Journal of Experimental Psychology*, 69(2):285–298, 2016.
- [65] Steve Mann. Wearable tetherless computer-mediated reality: Wearcam as a wearable face-recognizer, and other applications for the disabled papers. 1996.

- [66] Natalia Marmasse. commotion: a context-aware communication system. In *CHI'99 Extended Abstracts on Human Factors in Computing Systems*, pages 320–321, 1999.
- [67] Lindsey A McCarty and Melissa S Faulkner. Integrating writing and mathematics: journaling to increase learning and enjoyment while reducing anxiety. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 39(3):145–159, 2020.
- [68] Amanda Mercer, Elizabeth Warson, and Jenny Zhao. Visual journaling: An intervention to influence stress, anxiety and affect levels in medical students. *The Arts in Psychotherapy*, 37(2):143–148, 2010.
- [69] Jennifer A Moon. *Learning journals: A handbook for reflective practice and professional development*. Routledge, 2006.
- [70] Neville Moray. Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly journal of experimental psychology*, 11(1):56–60, 1959.
- [71] Khalid Nassiri and Moulay Akhloufi. Transformer models used for text-based question answering systems. *Applied Intelligence*, 53(9):10602–10635, 2023.
- [72] Subigya Nepal, Arvind Pillai, William Campbell, Talie Massachi, Eunsol Soul Choi, Orson Xu, Joanna Kuc, Jeremy Huckins, Jason Holden, Colin Depp, et al. Contextual ai journaling: Integrating llm and time series behavioral sensing technology to promote self-reflection and well-being using the mindscape app. *arXiv preprint arXiv:2404.00487*, 2024.
- [73] Hannah M O’rourke, Wendy Duggleby, Kimberly D Fraser, and Lauren Jerke. Factors that affect quality of life from the perspective of people with dementia: a metasynthesis. *Journal of the American Geriatrics Society*, 63(1):24–38, 2015.
- [74] Lynn Ossher, Kristin E. Flegal, and Cindy Lustig. Everyday memory errors in older adults. *Aging, Neuropsychology, and Cognition*, 20(2):220–242, 2013. PMID: 22694275.
- [75] Frederick L Philippe, Richard Koestner, Genevieve Beaulieu-Pelletier, Serge Lecours, and Natasha Leves. The role of episodic memories in current and future well-being. *Personality and Social Psychology Bulletin*, 38(4):505–519, 2012.

- [76] Aung Pyae and Tapani N. Joelsson. Investigating the usability and user experiences of voice user interface: a case of google home smart speaker. In *Proceedings of the 20th international conference on human-computer interaction with mobile devices and services adjunct*, pages 127–131, 2018.
- [77] Björn Rasch and Jan Born. About sleep’s role in memory. *Physiological reviews*, 2013.
- [78] Ginger Raterink. Reflective journaling for critical thinking development in advanced practice registered nurse students. *Journal of Nursing Education*, 55(2):101–104, 2016.
- [79] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.
- [80] Douglas A. Reynolds, Thomas F. Quatieri, and Robert B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital signal processing*, 10(1-3):19–41, 2000.
- [81] B. J. Rhodes and P. Maes. Just-in-time information retrieval agents. *IBM Systems Journal*, 39(3.4):685–704, 2000.
- [82] Bradley J. Rhodes. The wearable remembrance agent: A system for augmented memory. *Personal Technologies*, 1(4):218–224, Dec 1997.
- [83] Bradley James Rhodes and Pattie Maes. Just-in-time information retrieval agents. *IBM Systems journal*, 39(3.4):685–704, 2000.
- [84] Utkarsh Sarawgi, Wazeer Zulfikar, Nouran Soliman, and Pattie Maes. Multimodal inductive transfer learning for detection of alzheimer’s dementia and its severity. *arXiv preprint arXiv:2009.00700*, 2020.
- [85] Daniel L Schacter and Donna Rose Addis. The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):773–786, 2007.
- [86] Valentin Schwind and Niels Henze. Anticipated user stereotypes systematically affect the social acceptability of mobile devices. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, pages 1–12, 2020.
- [87] Mohit Shah, Brian Mears, Chaitali Chakrabarti, and Andreas Spanias. Lifelogging: Archival and retrieval of continuously recorded audio using

- wearable devices. In *2012 IEEE International Conference on Emerging Signal Processing Applications*, pages 99–102, 2012.
- [88] Signy Sheldon, Mary Pat McAndrews, and Morris Moscovitch. Episodic memory processes mediated by the medial temporal lobes contribute to open-ended problem solving. *Neuropsychologia*, 49(9):2439–2447, 2011.
- [89] Shamane Siriwardhana, Rivindu Weerasekera, Elliott Wen, Tharindu Kaluarachchi, Rajib Rana, and Suranga Nanayakkara. Improving the domain adaptation of retrieval augmented generation (rag) models for open domain question answering. *Transactions of the Association for Computational Linguistics*, 11:1–17, 2023.
- [90] Monika Sohal, Pavneet Singh, Bhupinder Singh Dhillon, and Harbir Singh Gill. Efficacy of journaling in the management of mental illness: a systematic review and meta-analysis. *Family medicine and community health*, 10(1), 2022.
- [91] Felix Stahlberg, Aashish Kumar, Chris Alberti, and Shankar Kumar. Conciseness: An overlooked language task, 2022.
- [92] Madiha Tabassum, Tomasz Kosiński, Alisa Frik, Nathan Malkin, Primal Wijesekera, Serge Egelman, and Heather Richter Lipford. Investigating users’ preferences and expectations for always-listening voice assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(4), sep 2020.
- [93] Endel Tulving. What is episodic memory? *Current directions in psychological science*, 2(3):67–70, 1993.
- [94] Endel Tulving. Episodic memory: From mind to brain. *Annual review of psychology*, 53(1):1–25, 2002.
- [95] Sunil Vemuri, Chris Schmandt, Walter Bender, Stefanie Tellex, and Brad Lassey. An audio-based personal memory aid. pages 400–417, 2004.
- [96] Colleen Walsh. Are google and smartphones degrading our memories? *Harvard Gazette*, October 2021.
- [97] Philip C Watkins. Implicit memory bias in depression. *Cognition & Emotion*, 16(3):381–402, 2002.
- [98] Jason Wei, Maarten Bosma, Vincent Y. Zhao, Kelvin Guu, Adams Wei Yu, Brian ... Lester, and Quoc V. Le. Finetuned language models are zero-shot learners. *arXiv preprint arXiv:2109.01652*, 2021.

- [99] Helen L Williams, Martin A Conway, and Alan D Baddeley. The boundaries of episodic memories. 2008.
- [100] Julie R. Williamson, Andrew Crossan, and Stephen Brewster. Multimodal mobile interactions: Usability studies in real world settings. In *Proceedings of the 13th International Conference on Multimodal Interfaces, ICMI '11*, page 361–368, New York, NY, USA, 2011. Association for Computing Machinery.
- [101] Henry H Wilmer, Lauren E Sherman, and Jason M Chein. Smartphones and cognition: A review of research exploring the links between mobile technology habits and cognitive functioning. *Frontiers in psychology*, 8:605, 2017.
- [102] Kiichiro Yamano and Katunobu Itou. Browsing audio life-log data using acoustic and location information. In *2009 Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, pages 96–101, 2009.
- [103] Andreea Zaman, Roni Setton, Caroline Catmur, and Charlotte Russell. What is auto-noetic consciousness? examining what underlies subjective experience in memory and future thinking. *Examining What Underlies Subjective Experience in Memory and Future Thinking*, 2023.