

The Goldilocks Time Window for Proactive Interventions in Wearable AI Systems

Cathy Mengying Fang

Wazeer Zulfikar

Pattie Maes

{catfang,wazeer,pattie}@media.mit.edu

MIT Media Lab

Cambridge, USA

Yasith Samaradivakara

Suranga Nanayakkara

{yasith,suranga}@ahlab.org

Augmented Human Lab, National University of Singapore

Singapore, Singapore

ABSTRACT

As AI systems become increasingly integrated into our daily lives and into wearable form factors, there's a fundamental tension between their potential to proactively assist us and the risk of creating intrusive, dependency-forming experiences. This work proposes the concept of a Goldilocks Time Window—a contextually adaptive time window for proactive AI systems to deliver effective interventions. We discuss the critical factors that determine the time window, and the need of a framework for designing and evaluating proactive AI systems that can navigate this tension successfully.

ACM Reference Format:

Cathy Mengying Fang, Wazeer Zulfikar, Pattie Maes, Yasith Samaradivakara, and Suranga Nanayakkara. 2025. The Goldilocks Time Window for Proactive Interventions in Wearable AI Systems. In *The CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/nnnnnnnnnnnnnnn>

1 INTRODUCTION

Most systems for assistance and behavior change interact with users at a System 2 level—engaging individuals in slow, effortful thinking [1]. We argue that in order for wearable AI systems to effectively engage users at a System 1 level—where thinking is fast and automatic [8]—they need to be proactive. For instance, a person who wants to be healthy but subconsciously reaches out for a tasty yet unhealthy dessert would have liked an external intervention at that moment. “Just-in-time adaptive interventions” (JITAIs) solidified this concept by focusing on delivering the right type and amount of support at the right moment, adapting to an individual’s dynamic internal and contextual state [11, 14].

Current proactive frameworks, however, mainly focus on conversational AI in text-based systems [2, 5, 6]. Proactive interventions now take the form of various modalities such as audio, visual, haptic and more. Wearable multi-modal devices, such as smart glasses and head-mounted displays, allow the system to reference data in real-time and possibility have more avenues of intervening [3, 9, 15]. Thus, wearable AI can detect and respond to behavioral patterns

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI '25, April 26-May 1, 2025, Yokohama, Japan

© 2025 Copyright held by the owner/author(s).

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.

<https://doi.org/10.1145/nnnnnnnnnnnnnnn>

in real-time, offering timely interventions when users have not consciously decided to seek external support but would still benefit from it. Building on these advancements, it is crucial to examine not only the capabilities of proactive wearable AI interventions but also how they can optimize intervention timing to maximize their effectiveness in real-world contexts.

This paper explores the critical timing dimension of proactive AI interventions, introducing the concept of the “Goldilocks Time Window”—the optimal window of time when interventions are neither too early (risking false positives and unnecessary interruptions) nor too late (when effectiveness diminishes). Specifically, we identify several key factors that determine effective intervention timing, such as contextual environment, delivery modality, and social awareness. We argue that successful proactive AI must balance predictability with strategic unexpectedness and maintain alignment with user intent while preserving agency. We call for a standardized method for evaluating the timing dimension of proactive systems and assess intervention effectiveness, offering designers and researchers a structured approach to creating AI systems that know not just what to say, but precisely when to say it.

2 THE GOLDILOCKS TIME WINDOW

With the introduction of foundation models like Large Language Models, interactions between users and systems are no longer constrained by graphical user interfaces (GUI); the communication with these systems becomes more **temporal** and the interactions becomes implicit, where the systems infer user intent based on additional information [10]. This shift introduces a critical challenge: determining when AI systems should intervene without being explicitly prompted by the user. Imagine a helicopter parent constantly hovering over your shoulder—this metaphor captures the risk of poorly-timed proactive AI.

From the user’s thought to action, there is an optimum time window for intervention, which we call the “Goldilocks Time Window” (Figure 1). It has paradoxical properties: the efficacy of the intervention reduces as the timing gets closer to the user taking an

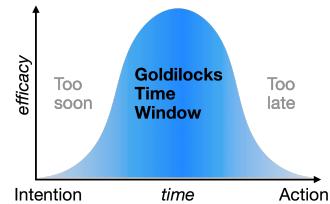


Figure 1: The Goldilocks Time Window

undesirable action, but if the intervention happens too early, the probability of a false positive rate increases and may cause unnecessary interruptions. Consider fire alarms. In high-stakes situations, the fire alarm abides the user's trust that it must sound during genuine emergencies. However, a false positive is irritating, and repeated false positives lose the user's trust akin to "The Boy who Cried Wolf" effect.

Effective intervention systems must align with the user's greater intention or higher-level goals. At the same time, we argue the intervention's efficacy depends on unexpectedness. For example, the system chimes in and reminds the user about their health goals when the user subconsciously reaches for a chocolate bar. Thus, the intervention's specific timing provides value through their unexpectedness. Current user experience frameworks emphasize the importance of predictability of user interfaces in building trust with the user [12]. Proactive interventions seemingly contradict but complement the predictability of user interfaces with its strategic unexpectedness.

In addition, well-timed interventions enhance rather than diminish user agency. Poorly-timed or overly frequent interventions can create dependency, where users rely on the system rather than developing their own capabilities. For example, a language learning tool that automatically corrects every error without requiring user engagement. Over time, the user's agency diminishes as the technology shifts from empowering tool to essential prosthetic.

3 FACTORS THAT AFFECT THE GOLDILOCKS TIME WINDOW

This optimal timing is influenced by several interconnected factors that work together to determine when an intervention will be most effective. First, the system should have social sensitivity. An intervention appropriate when someone is alone at home might be completely inappropriate during a job interview or important meeting. The delivery modality—whether through visual alerts, sounds, or haptic feedback— influences optimal timing, as different channels have varying levels of intrusiveness and attention-capturing ability.

The alignment between immediate actions and longer-term goals creates tension that timing must navigate. Effective interventions balance immediate needs (like stress relief) with long-term objectives (such as health improvement). This balance shifts depending on the magnitude of potential consequences - interventions for minor dietary choices warrant different timing considerations than those addressing potential safety hazards. User history creates important precedents that shape future intervention effectiveness. Previous experiences with similar interventions establish expectations and response patterns that vary even between seemingly similar situations.

4 EXAMPLES

Here we use three projects done by the authors as examples of where proactive responses play a significant role in helping the user to achieve their respective goals.

Mirai[7] is a wearable AI system with an integrated camera, real-time speech processing, and personalized voice-cloning to provide proactive and contextual nudges for positive behavior change (Figure 2 in Appendix A). Mirai continuously monitors and analyzes

the user's environment to anticipate their intentions, generating contextually-appropriate responses delivered in the user's own cloned voice. Noteably, the system implements a "debouncer" mechanism (Appendix A.2) and a proactive agent (Appendix A.1) to infer the right timing to intervene. The debouncer mechanism ensures that responses are triggered only when there is a significant context change or at controlled intervals during stable states. Upon detecting new information, the proactive agent evaluates whether the update reflects a significant change in user behavior or context. It also determines the appropriate timing for a response—deciding whether to respond immediately or delay the response. If a response is warranted, it incorporates the updated understanding into the reply.

Memoro[16] is a wearable audio-based memory assistant. Memoro uses a large language model (LLM) to infer the user's memory needs in a conversational context, semantically search memories, and present minimal suggestions. The assistant has two interaction modes: Query Mode for voicing queries and Queryless Mode for on-demand predictive assistance, without explicit query Figure 3. The Queryless mode is facilitated by the user requesting the memory assistant to understand the ongoing flow of the conversation and infer their precise memory need.

AiSee[4] is an assistive wearable interface that helps people with visual impairment overcome daily challenges, such as grocery shopping. The prototype consisted of a bone-conduction headset with a camera capable of processing and extracting features such as text, logos, and labels from the captured images. While evaluating the device with participants, the researchers observed that providing excessive information that is not 100% relevant for the recognition produced confusion, and made the users lose their concentration on the outcome. Thus, future work on how a proactive version of the AI system could infer user's needs and focus and deliver only the critical information at the right moment.

5 CALL TO ACTION

This paper presents a case for the importance of timing in interventions delivered by proactive, wearable AI systems. We define the term "Goldilocks Time Window" to illustrate that the maximum efficacy of a proactive intervention happens when the timing is not too soon nor too late, and we identified factors that influence the Goldilocks Time Window. We used the authors' research prototypes as illustrative examples. Current evaluations of wearable systems are primarily centered around wearability and privacy [13, 17]. We propose that future systems that involve a proactive approach should evaluate the systems' efficacy at achieving the right time window. In addition, because these wearable systems often require communications among multiple components, latency is also often used as part of the systems' technical evaluation, which primarily focuses on the engineering aspects of achieving low latency. We advocate for a type of latency evaluation that focuses on the timing of the intervention delivery with respect to human needs. This ensures that the success metrics have ecological validity and are grounded in practical, safe, and ethical use [10]. This new benchmark might look like a mix of qualitative and quantitative evaluations where the user provides open-ended feedback, given the variety in individuals' prior experience and contexts.

REFERENCES

- [1] Alexander T Adams, Jean Costa, Malte F Jung, and Tanzeem Choudhury. 2015. Mindless computing: designing technologies to subtly influence behavior. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 719–730.
- [2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.
- [3] Riku Arakawa, Jill Fain Lehman, and Mayank Goel. 2024. PrISM-Q&A: Step-Aware Voice Assistant on a Smartwatch Enabled by Multimodal Procedure Tracking and Large Language Models. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 4, Article 180 (Nov. 2024), 26 pages. <https://doi.org/10.1145/3699759>
- [4] Roger Boldu, Denys JC Matthies, Haimo Zhang, and Suranga Nanayakkara. 2020. AiSee: an assistive wearable device to support visually impaired grocery shoppers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–25.
- [5] Yang Deng, Lizi Liao, Wenqiang Lei, Grace Yang, Wai Lam, and Tat-Seng Chua. 2025. Proactive Conversational AI: A Comprehensive Survey of Advancements and Opportunities. *ACM Transactions on Information Systems* (2025).
- [6] Christopher Diebel, Marc Goutier, Martin Adam, and Alexander Benlian. 2025. When AI-Based Agents Are Proactive: Implications for Competence and System Satisfaction in Human–AI Collaboration. *Business & Information Systems Engineering* (2025), 1–20.
- [7] Cathy Mengying Fang, Yasith Samaradivakara, Pattie Maes, and Suranga Nanayakkara. 2025. Mirai: A Wearable Proactive AI "Inner-Voice" for Contextual Nudging. *arXiv preprint arXiv:2502.02370* (2025).
- [8] Daniel Kahneman. 2011. *Thinking, fast and slow*. macmillan.
- [9] Jaewook Lee, Jun Wang, Elizabeth Brown, Liam Chu, Sebastian S Rodriguez, and Jon E Froehlich. 2024. GazePointAR: A context-aware multimodal voice assistant for pronoun disambiguation in wearable augmented reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.
- [10] Meredith Ringel Morris. 2025. HCI for AGI. *Interactions* 32, 2 (2025), 26–32.
- [11] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2018. Just-in-time adaptive interventions (JITAs) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine* (2018), 1–17.
- [12] Jakob Nielsen. 10. Usability heuristics for user interface design.
- [13] Alex Olwal, Kevin Balke, Dmitrii Votintsev, Thad Starner, Paula Conn, Bonnie Chinh, and Benoit Corda. 2020. Wearable subtitles: Augmenting spoken communication with lightweight eyewear for all-day captioning. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 1108–1120.
- [14] Adiba Orzikulova, Han Xiao, Zhipeng Li, Yukang Yan, Yuntao Wang, Yuanchun Shi, Marzyeh Ghassemi, Sung-Ju Lee, Anind K Dey, and Xuhai Xu. 2024. Time2Stop: Adaptive and Explainable Human-AI Loop for Smartphone Overuse Intervention. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20.
- [15] Peng Wang and Alan F. Smeaton. 2013. Using visual lifelogs to automatically characterize everyday activities. *Inf. Sci.* 230 (May 2013), 147–161. <https://doi.org/10.1016/j.ins.2012.12.028>
- [16] Wazeer Deen Zulfikar, Samantha Chan, and Pattie Maes. 2024. Memoro: Using Large Language Models to Realize a Concise Interface for Real-Time Memory Augmentation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.
- [17] Wazeer Deen Zulfikar, Cayden Pierce, and Pattie Maes. 2024. MeMic: Towards Social Acceptability of User-Only Speech Recording Wearables. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–9.

A MIRAI

A.1 Proactive prompt snippet

These are some contextual rules: If you receive new information tagged as [NEW INFO] from "system", first consider whether this information indicates a new behavior of the user, and only if so you should initiate a response. If you are asked to remind the user, determine the right moment to say the reminder, do so succinctly under 5 words without causing much disruption. A good rule of thumb for the right time to intervene is when the user hasn't been speaking for a while AND when you don't see someone else talking.

The full prompt can be found in [7].

A.2 Debouncer overview

The debouncer mechanism is mathematically represented as follows:

$$U_t = \begin{cases} 1 & \text{if } S_t = \text{"YES"} \text{ and } S_t \neq S_{t-1}, \\ 1 & \text{if } S_t = \text{"YES"} \text{ and } R_t \bmod 3 = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Here, U_t determines whether a response is triggered at time t . The classifier's state is represented by S_t , with S_{t-1} denoting the previous state, and R_t representing the time step. A response is triggered under two conditions:

- (1) The current state S_t is "YES" and differs from the previous state ($S_t \neq S_{t-1}$), indicating a significant context change.
- (2) The current state S_t is "YES" and the time step R_t satisfies $R_t \bmod 3 = 0$, which prevents repeated interruptions during stable states.

This mechanism ensures that responses are contextually relevant, avoids excessive interruptions, and maintains responsiveness to meaningful changes.

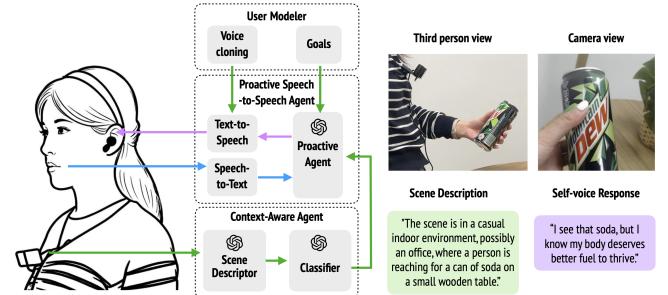


Figure 2: Overview of Mirai.

B MEMORO

B.1 Queryless prompt

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]

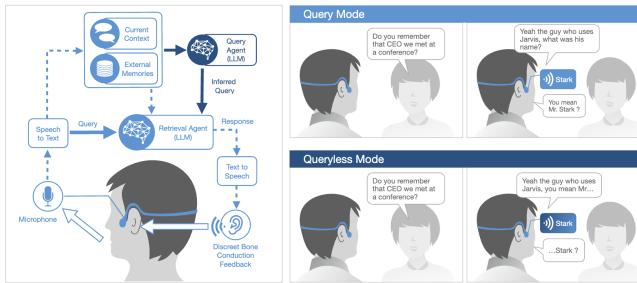


Figure 3: Memoro’s interaction modes: (1) **Query Mode** where the user can ask contextual questions (2) **Queryless Mode** where the user can request predictive assistance and skip query formation. In both modes, responses are discreetly played back to the user using a bone conduction headset.

C AISEE



Figure 4: An assistive bone-conduction headset for people with visual impairment.