Wearable Meets LLM for Stress Management: A Duoethnographic Study Integrating Wearable-Triggered Stressors and LLM Chatbots for Personalized Interventions

Sameer Neupane*†
Computer Science
University of Memphis
Memphis, Tennessee, USA
sameer.neupane@memphis.edu

Poorvesh Dongre†
Department of Computer Science
Virginia Tech
Blacksburg, Virginia, USA
poorvesh@vt.edu

Denis Gracanin
Department of Computer Science
Virginia Tech
Blacksburg, Virginia, USA
gracanin@vt.edu

Santosh Kumar
Computer Science
University of Memphis
Memphis, Tennessee, USA
santosh.kumar@memphis.edu

Abstract

We use a duoethnographic approach to study how wearable-integrated LLM chatbots can assist with personalized stress management, addressing the growing need for immediacy and tailored interventions. Two researchers interacted with custom chatbots over 22 days, responding to wearable-detected physiological prompts, recording stressor phrases, and using them to seek tailored interventions from their LLM-powered chatbots. They recorded their experiences in autoethnographic diaries and analyzed them during weekly discussions, focusing on the relevance, clarity, and impact of chatbot-generated interventions. Results showed that even though most events triggered by the wearable were meaningful, only one in five warranted an intervention. It also showed that interventions tailored with brief event descriptions were more effective than generic ones. By examining the intersection of wearables and LLM, this research contributes to developing more effective, user-centric mental health tools for real-time stress relief and behavior change.

CCS Concepts

• **Human-centered computing → User studies**; **Ubiquitous and mobile computing**.

Keywords

Duoethnography, large language models (LLM), stress relief, stressors, interventions, wearables

ACM Reference Format:
Sameer Neupane, Poorvesh Dongre, Denis Gracanin, and Santosh Kumar. 2025. Wearable Meets LLM for Stress Management: A Duoethnographic Study Integrating Wearable-Triggered Stressors and LLM Chatbots for Personalized Interventions. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3706599.3720197

*Corresponding Author
†Both authors contributed equally to this research.

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3720197

1 Introduction

Stress is defined as a state of mental or emotional strain resulting from adverse or demanding circumstances. Stress has become a pervasive challenge in modern life, with chronic stress posing significant risks to mental and physical health [4, 7, 15, 22, 31]. This issue holds considerable significance in Human-Computer Interaction (HCI), as technology plays an increasingly central role in daily life and, consequently, in mental health interventions.

Traditional approaches to stress management, such as therapy and structured programs, are valuable but often constrained by accessibility, cost, and the lack of immediacy. Wearable devices have emerged as promising tools for stress tracking, offering real-time monitoring of physiological stress indicators, such as heart rate variability (HRV), to provide timely insights into stress levels [12, 20, 34]. Complementing this, generative AI has been used to develop on-demand mental health support chatbots [21, 23]. While LLM-driven AI solutions have shown promise in supporting mental health diagnosis [26] and providing mental health first aid [17], they often exhibit behaviors typical of low-quality therapy [3]. Recent LLM-powered journaling approaches integrating passively collected behavioral patterns have enhanced self-awareness and well-being [24]. However, passive smartphone data may miss momentary events (e.g., anxiety, frustration) that wearables may catch from physiological arousal. In both cases, do all detected events need an intervention or only a subset? If the latter, which ones? We envision a future where wearables, smartphones, and generative AI work together to identify stressors automatically, determine which stressors require intervention, and provide tailored interventions.

In this paper, we take the first steps to investigate the opportunities and challenges in integrating wearable-triggered stressor journaling with LLM chatbots to provide personalized interventions. We employed a duoethnographic methodology, a collaborative autoethnography involving two researchers, to capture detailed, reflexive accounts of integrating these technologies. In a 22-day study, two PhD students (who also experience high self-reported stress) used CuesHub [1], an app for smartwatches and smartphones, offering real-time wearable-triggered stressor journaling. Simultaneously, both researchers used a personalized LLM-based stress intervention chatbot built with OpenAI's custom GPT. Each researcher engineered prompt templates tailored to their individual stress triggers, coping styles, and daily routines, following a lightweight framework informed by modern stress management strategies. This personalized approach allowed a more nuanced exploration of how wearables and LLM chatbot interventions can be integrated and adapted to user needs and preferences.

Our motivation lies in harnessing the complementary strengths of wearables and LLM chatbots for an integrated stress management solution. Our research focuses on two questions: (1) How often do users seek interventions for momentary events triggered by wearables? and (2) How does integrating very brief descriptions of momentary events into LLMs enhance stress management interventions? By documenting our lived experiences, we aim to inform the design of multi-modal AI solutions that can track stressors in real time, offer interventions, and adapt to the complexities of individual user needs. Our final goal is to articulate design considerations and insights that can guide broader HCI efforts to develop integrated AI systems for stress management.

2 Background and Related Works

Technology-powered interventions have become increasingly popular for supporting mental health and well-being. They provide real-time data, adapt to personal needs, and reduce barriers such as cost and accessibility.

2.1 Wearables for Stress Tracking, Stressor Journaling, and Interventions

Wearable devices have emerged as one of the most widely used AI-driven tools for recognizing affect states, including stress from physiological signals. During the past two decades, the scientific community has made significant advances in stress detection, using wearable technology to monitor physiological responses in real-time, indicative of stress [14, 29, 40]. Building on this foundation, commercial devices such as Fitbit [34], Garmin [12], Whoop [20], and Empatica [11] have integrated stress tracking capabilities. Passively collected data has been leveraged to create innovative visualizations for self-reflection [18, 19, 30, 35, 41]. Furthermore, passive stress detection has effectively delivered real-time interventions, addressing stress precisely when it occurs [2, 16]. However, participants in these studies have expressed the desire for interventions to match the source of stress [16, 36]. Recent work has shown how stress detection can be used to collect stressors and how incorporating stressors into self-reflective visualizations can lead to behavioral changes [25]. This study takes the next step of integrating stressors with chatbot interactions for personalized stress interventions.

2.2 LLM for Mental Health Support

Mental health support chatbots trace their origins back to ELIZA, one of the first computer programs to simulate human conversation [38]. Since then, several rule-based and semi-automated chatbots have emerged that offer psycho-education, basic coping strategies, and limited therapeutic interactions. However, the emergence of LLMs has transformed the development of conversational agents and created novel opportunities to provide users with advanced mental health chatbots. Liu et al. [23] developed an LLM chatbot, ChatCounselor, by fine-tuning an open-source pre-trained LLM on a dataset prepared from 260 in-depth interviews between patients and psychologists. The LLM chatbot by Lai et al. [21] was developed by fine-tuning two Chinese pre-trained LLMs with real-world professional Q&A datasets from psychologists and psychological articles to develop their mental health chatbot. Dongre et al. [8, 9] also fine-tuned an open-source LLM on a Q&A dataset scraped from a mental health website to develop their LLM chatbot for stress management. CaiTI, a Conversational AI Therapist, uses LLMs, smart devices, and reinforcement learning to provide personalized psychotherapeutic interventions, enhancing mental health self-care [26]. BOLT, a framework for evaluating LLM therapists' conversational behavior, highlights the need for improvements to achieve high-quality care [3]. MindScape introduces an AI-powered journaling approach that integrates behavioral patterns like sleep, location, and engagement to deliver personalized, context-aware prompts [24]. These works show the vast potential of LLM to power tailored interventions on demand. The novelty of this study lies in its integration with wearable AI to deliver automated, personalized, and context-aware stress interventions.

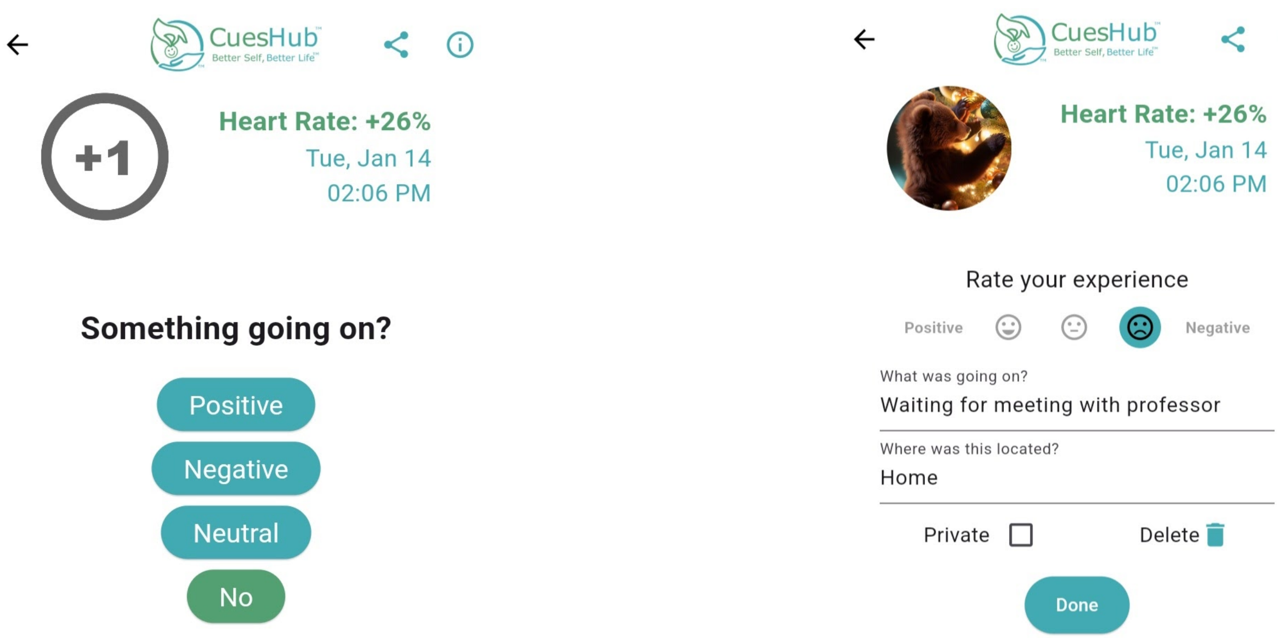
3 Methods

3.1 Duoethnography

This study employs a *duoethnographic approach*, a qualitative method exploring how two or more researchers interpret and give meaning to shared experiences [5, 32]. It builds on autoethnographic methods [6, 10, 27, 37], which have been applied in diverse contexts, including the use of LLM chatbots for thesis writing, creative expression, and to understand the humanistic elements of a Chatbot's self-identity [13, 28, 33]. For this study, two researchers independently interacted with the wearables and personalized LLM chatbots for stress management, documenting their experiences and reflections in autoethnographic diaries. This dual autoethnographic approach provided introspective data on how integrating the two AI systems influenced perceived effectiveness in managing everyday stress. Weekly meetings between the researchers enabled shared reflection, discussion, and comparative analysis of individual perspectives. The study's insights into integrating wearables and LLM chatbots for stress management were enriched by combining individual depth with collaborative breadth.

3.2 Wearable-triggered Stressor Journaling

Both researchers wore Samsung Galaxy Watch 6 devices equipped with the CuesHub smartwatch app [1], serving as the wearable AI platform for detecting physiological events associated with stress. Upon detecting a physiological event, the smartwatch app prompted



(a) Valence rating screen of the prompted event

(b) Event description screen of the prompted event

Figure 1: CuesHub app screenshots for recording valence and descriptions for events detected by the smartwatch app

researchers via their smartphones, requesting them to provide brief descriptions of the event. Researchers rated the events using one of four options: *Positive*, *Negative*, *Neutral*, or *No* (see Figure 1a). For responses indicating *Positive*, *Negative*, or *Neutral*, researchers were further prompted to answer two questions: *What was going on?* and *Where was this located?* (see Figure 1b).

3.3 Custom GPT for Stress Intervention

The chatbots employed in this study used OpenAI's GPT-4o [1] as their underlying language model. To investigate the effectiveness of different interaction paradigms, we explored two contrasting approaches to chatbot design, emphasizing distinct user experiences and intervention delivery styles. *Researcher A* (R_A) designed their chatbot, named *DeStressify*, to emulate a zero-shot prompting approach akin to intervention systems that automatically deliver stress interventions upon detecting a user's stress [2, 16]. When experiencing the need for intervention, *Researcher A* prompted the chatbot with their stressor and location. Most interactions were transactional, consisting of a single prompt-response exchange where the chatbot provided an intervention in the traditional instructive manner often used in text messaging.

Researcher B (R_B), on the other hand, adopted the modern approach for their chatbot, named *StressGPT*. This chatbot was structured to facilitate a more therapy-like interaction, allowing them to engage in dynamic, multi-turn dialogues. This conversational style aimed to simulate collaborative interactions, with *StressGPT* mimicking personalized coaching sessions by refining suggestions or providing layered support through iterative exchanges, enabling deeper reflection and tailored advice. Details about the chatbot designs for both researchers are provided in Appendix A.1.

[1]https://openai.com/index/gpt-4/

3.4 Data Collection and Analysis

Both researchers used the CuesHub app for 22 days. For stressors where researchers felt the need for interventions, they voluntarily engaged with the LLM chatbots. R_A, using *DeStressify*, engaged with the chatbot immediately (or soon) after prompts in the CuesHub app, i.e., following a *just-in-time zero-shot* approach. R_A incorporated the most recently logged stressor and location from the CuesHub app into the chatbot prompt for each interaction.

R_B, in contrast, engaged with *StressGPT* predominantly at the end of the day, consolidating the day's events and reflectively interacting with *StressGPT*. R_B used the logged stressor entries from the CuesHub app as a reference to interact with the chatbot. We note that even in this *end-of-day interactive* approach, prompts helped capture events that may have been missed otherwise.

After each engagement, researchers rated the intervention on a scale from *Very poor* to *Very good*. Ratings were based on the immediate relevance and usefulness of the interventions in addressing specific stressors. Both researchers employed distinct prompt engineering strategies, guided by catalog patterns described in [39], ensuring tailored and meaningful interactions with their respective LLM chatbots. Both researchers maintained a running journal after each chatbot interaction, reflecting and documenting their perceptions of its relevance, clarity, and impact.

Both researchers conducted a thematic analysis of data from weekly meeting transcripts, journals, and stressors logged through

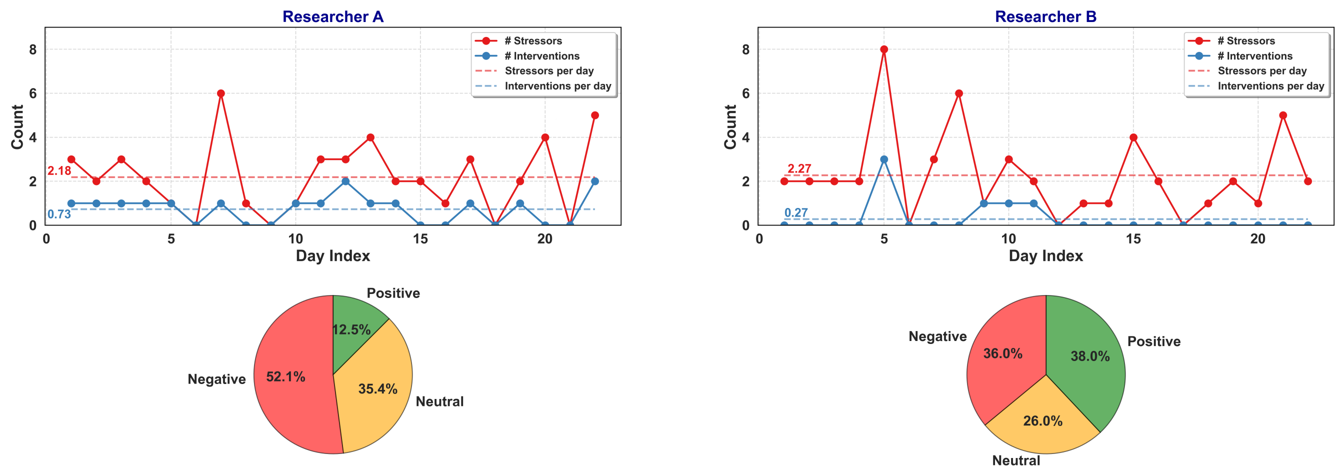


Figure 2: Frequency of stressors and interventions per day and valence distribution of events detected by the CuesHub app

the CuesHub app to examine chatbot usage, intervention effectiveness, and the connection between wearable-detected stress experiences and chatbot interactions. The analysis included design documentation detailing chatbot customization and expectations, daily journals reflecting on the relevance and effectiveness of interactions, and engagement logs tracking the frequency and content of interactions.

4 Findings

We first describe the stressors logged by researchers and how many of these needed an intervention. We then compare and contrast our intervention experience between the just-in-time zero-shot and the end-of-day interactive approaches.

4.1 Diversity in Stressors Detected and the Need for Interventions

4.1.1 Diversity in Stressors Detected. The stressors detected by the app and recorded by the researchers ranged from work-related challenges, such as preparing complex drafts or debugging code, to social stressors, like navigating difficult conversations or attending gatherings. Daily life stressors, including routine tasks such as cooking or dealing with unexpected delays, were also captured, along with positive stressors, such as celebrating achievements or enjoying successful interactions. Work-related and social stressors often evoked a sense of urgency and anxiety, whereas daily life stressors tended to build gradually, creating cumulative stress. Positive stressors, on the other hand, brought excitement and a sense of fulfillment, often tied to personal growth or future rewards.

4.1.2 Proportion of stress events that require interventions. The researchers logged 98 events (48 by R_A and 50 by R_B). Of these, 43, 30, and 25 events were negative, neutral, and positive, respectively. Figure 2 shows the number of stressors and how many were used for interventions on each day during the study. Out of the 98 events, the researchers felt the need to engage with the chatbot for 22 events (R_A = 16, R_B = 6). Of those, 20 were negative, and 2 were neutral. Researchers required an intervention on only half of those days when at least one stress event was reported. The rating distributions of interventions were: for R_A, *Very poor* (6.25%), *Poor* (18.75%), *Acceptable* (25.00%), *Good* (43.75%), and *Very good* (6.25%); for R_B, *Acceptable* (16.67%), *Good* (66.67%), and *Very good* (16.67%).

Interestingly, certain negative events, such as *"Working on a paper," "Rushing to catch a flight," "Completing pending work,"* and *"Feeling hungry,"* did not require intervention. This was attributed to various factors, including the perception that the event was manageable without assistance, the temporary or fleeting nature of the stressor, or a personal preference for self-reliance in handling the situation. This observation highlights that merely detecting a stress state is insufficient; recognizing the specific stressor is critical for delivering tailored interventions. Additionally, stressors like *"coding issues"* sometimes warranted intervention but not consistently in every occurrence. This highlights that simply identifying a stressor may also not be sufficient; participant feedback or additional contextual cues are necessary to determine when an intervention is genuinely needed.

4.2 Common Themes in Our Experience with *DeStressify* and *StressGPT*

We begin by highlighting the common themes that emerged from both researchers' experiences with their respective LLM chatbots (*DeStressify* and *StressGPT*) and wearable-triggered stressor logging, offering insights into the similarities in how these systems functioned and how they were perceived during interactions.

4.2.1 Need for Continuity and Stressor Awareness in Chatbots. The chatbot's effectiveness was significantly hindered by its inconsistent integration of continuity and contextual awareness, despite having access to the stressors. While it occasionally referenced past interactions, it failed to do so regularly, making conversations feel less personalized and lacking a sense of ongoing support. Instead of tailoring responses based on the history of stressors shared with the chatbot, it often provided generic solutions that did not address the evolving nature of the user's challenges. R_B noted, *It felt like the chatbot did not really know who I am and would have given similar responses to anyone asking similar questions.* This absence of context integration emphasized the need for chatbots to track and link past stressors with current interactions, ensuring a more dynamic, responsive system that delivers consistent, relevant support.

4.2.2 Striking the Perfect Balance Between Emotional Support and Practicality.

A key insight was the need for chatbots to balance emotional and practical support. While strategies like time management and breathing exercises were helpful, they were incomplete without emotional support. Purely emotional responses, without actionable advice, were ineffective in addressing stressors. R_A valued interventions that reframed emotional states, such as linking frustration to the PhD journey, while R_B appreciated the chatbot's emojis, which felt like a human-like touch to the interaction. This underscores the need for a holistic approach, where chatbots offer both emotional support and practical guidance, ensuring users feel fully supported in managing their stress.

4.2.3 Clarity and precision in chatbot responses.

Both researchers preferred concise responses that directly addressed their issues, as they helped reduce cognitive load and maintained efficiency in the interaction. R_B noted that the chatbot often provided long, structured responses to simple questions, which made the conversation feel more mechanical and less human-like. On the other hand, R_A particularly appreciated interventions that offered an explicit, singular action or suggestion rather than multiple options or vague guidance. This preference for specificity was especially apparent in emotionally charged or complex situations, where a precise, direct solution was more helpful than a broad range of solutions [3].

4.3 Contrasting Real-Time Zero-Shot vs. End-of-Day Interactive Chatbot Experiences

Next, we focus on the differences in researchers' experiences between real-time zero-shot interventions provided by *DeStressify* and the end-of-day interactive sessions facilitated by *StressGPT*.

4.3.1 Number of Interventions Required.

R_A primarily used *DeStressify* in real-time immediately after logging stressors that required interventions, whereas R_B engaged in conversation with *StressGPT* towards the end of the day. Figure 2 illustrates that while the average number of stressors logged per day was comparable for R_A (2.18) and R_B (2.27), R_A engaged with interventions almost three times more frequently per day (0.73) compared to R_B (0.27). This suggests that real-time interventions are more likely to be utilized when stressors are momentary. Factors such as the transient nature of certain stressors may make immediate intervention more relevant and actionable, providing support precisely when needed. In contrast, engaging with *StressGPT* at the end of the day may reduce the perceived need for interventions, as some stressors might have already dissipated or been resolved by that time. Although stressors such as *meeting with professor* or *long wait time* were common for both researchers, only R_A asked for interventions.

4.3.2 Targeted Interventions for DeStressify vs Human-Like Yet Not Domain-Specific for StressGPT.

R_A found that the wearable-triggered stressor logging app, paired with the LLM chatbot, provided targeted interventions tailored to specific stressors with effective, actionable strategies. For work challenges, breaking tasks into smaller steps improved clarity, while physical activities and environmental changes helped reset mental states. Social stressors were addressed through confidence-building interventions, like affirmations and reframing negative thoughts. Positive perspectives on frustrating situations also enhanced engagement and effectiveness. For example, targeted interventions, such as structured debugging for stressors like replicating a paper, were more effective than generic strategies like Progressive Muscle Relaxation (see A.2). In contrast, generic responses, like mindfulness exercises or repeated advice (e.g., "take a deep breath"), often caused frustration, reducing the chatbot's perceived usefulness. R_A preferred more dynamic, contextually relevant responses.

Although R_B found *StressGPT* more human-like compared to typical rule-based chatbots, the chatbot often acted like a general-purpose AI, responding to queries outside the stress management domain, which reduced its relevance. Additionally, it was quick to offer suggestions without fully understanding the user's stressor or context, generating lengthy responses without asking clarifying questions. This lack of active listening made the interaction feel less personal and empathetic. The structured, verbose responses felt robotic and detached, resembling any other general-purpose LLM chatbot, than a specialized stress management assistant. Furthermore, the fast text generation speed disrupted the conversational flow, making it challenging to stay engaged as responses were quickly produced before they could be fully read.

4.3.3 Privacy Concerns.

For *DeStressify*, R_A only shared stressors and locations that did not carry sensitive or risky identifiable information. *StressGPT* raised privacy concerns for R_B, especially when discussing sensitive or deeply personal topics. R_B expressed hesitation in fully opening up during interactions, fearing that private information might not be adequately protected or might be misused. This reluctance emphasized the need for stress management chatbots to establish trust and security, ensuring that users feel confident sharing personal information without the fear of breach or misuse. Stress management chatbots must ensure privacy, clear data policies, and transparency, fostering trust and engagement for more effective and personalized support.

5 Discussions

5.1 Challenges in Integrating Wearables and LLMs for Stress Management

Advances in wearable AI for real-time stress detection and LLMs capable of generating adaptive interventions hold significant promise for stress management. A natural assumption might be that seamlessly connecting wearable AI to LLMs could create the ultimate stress management system. However, this study's findings revealed several challenges in effectively integrating these technologies.

While LLMs excel at generating one-time interventions and engaging in extended conversational support, their integration with wearable AI raises significant questions about whether interventions align with users' event descriptions or stressors. Firstly, our findings reveal a discrepancy between stress detected by wearable AI and the user's actual need for intervention. Out of 98 confirmed stress events, only 22 (approximately 20%) required an intervention, indicating that delivering interventions for every detected event may be counterproductive.

Wearable AI detects both positive and negative stress events. Interestingly, not all negative events need interventions, either. In this study, fewer than half of the negative events prompted a desire for intervention. Systems that push out interventions based solely on stress detection (even if the AI were powerful enough to automatically distinguish negative events from positive or neutral ones) can create unnecessary burdens, leading to disengagement and reduced utility.

5.2 Opportunities in Integrating Wearables and LLMs for Stress Management

Prior research underscores the importance of interventions tailored to users' specific stressors [16, 36]. Our findings highlight the potential of wearable stressor-tracking apps like CuesHub, which combine physiological data with contextual cues, such as location and emotional state, to personalize interventions. By integrating these insights into LLMs, stress management systems can deliver more precise and responsive support that aligns with users' immediate needs. Effective stress management systems must transition from automatic, event-triggered responses to more selective approaches that prioritize user needs and preferences. By analyzing stressor descriptions, systems can better distinguish events that truly need intervention, improving relevance and impact.

Our study explored two approaches to engaging with the LLM chatbot: real-time and reflective interactions. In the real-time approach, delays in the delivery of interventions can significantly reduce their effectiveness, as stressors often require immediate attention. Although the reflective approach can tolerate delays, it may still benefit from faster interactions, as timely responses can support users in processing their experiences more effectively. By incorporating proactive and predictive mechanisms, wearable AI-LLM systems can anticipate potential stressors and intervene early shifting from reactive to preemptive support and enhancing overall stress management.

We also observed that the LLM chatbots sometimes successfully generated interventions based on previously shared events but were limited to events for which users explicitly requested support. Expanding its access to the full history of wearable-detected events could provide more context, enabling more personalized and informed interventions. Moreover, integrating memory and user history into LLM chatbot systems could significantly enhance the continuity of interactions, allowing for more personalized and consistent support over time across multiple sessions. This continuity can foster a deeper sense of trust and connection between the user and the chatbot, which is essential for building long-term engagement and effectively addressing recurring stressors.

Lastly, our findings highlighted the potential of a physiology-driven, emotion-aware chatbot to detect heightened stress cues and respond with a combination of practical solutions and emotional comfort, while features like adaptive dialogue could further enrich the user experience. Stress management chatbots could improve by emulating human-like interactions with empathetic communication and emotional intelligence, thus better addressing users' emotional needs, and fostering more compassionate, relevant exchanges.

In conclusion, this study highlights both the critical need and the opportunity for wearable AI-LLM systems to focus on predicting stressors and aligning interventions with the stressor and the surrounding context to ensure they provide meaningful support without adding unnecessary cognitive load.

6 Limitations

This study has several limitations that impact its generalizability. The small sample size of two researchers, both with expertise in wearables and AI, introduces potential bias and restricts broader applicability. Additionally, our perspectives—shaped by both optimism about AI's potential and caution regarding ethical concerns—may have influenced our interpretation of results. The absence of control groups and standardized conditions limits our ability to isolate the effects of wearable-triggered interventions, particularly given variations in chatbot design and interaction timing. Furthermore, the study relies primarily on subjective experiences and qualitative data, making it difficult to objectively assess intervention effectiveness. While we have expertise in AI and wearable technologies, our lack of formal training in stress management and mental health presents another limitation.

Future research should address these limitations by including larger, more diverse participant samples, implementing controlled comparisons, and integrating objective evaluation metrics. Additionally, to fully realize the potential of AI-driven stress interventions, future systems should enable seamless, automated integration between wearables and LLMs, reducing the need for manual input and ensuring real-time, personalized support.

7 Conclusion and Future Works

This study demonstrates the potential of wearable AI integrated LLM chatbots like *DeStressify* and *StressGPT* to offer valuable support for stress management, while also highlighting key areas for improvement. The findings indicate that the perceived effectiveness and impact of interventions is heavily influenced by factors such as timing, personalization, and contextual relevance. Although the chatbots provided useful assistance, limitations like repetitive suggestions, long responses, and a lack of memory retention from past interactions often compromised their ability to offer truly impactful support. These challenges highlight the need for advancements in chatbot systems, particularly with wearable integration. Looking ahead, future LLM chatbots, combined with wearable devices, have the potential to become more adaptive, effective, and user-centered, offering real-time, comprehensive, and personalized support for stress management.

Acknowledgments

We thank the anonymous reviewers for greatly improving the paper. Research reported here was supported by the National Institutes of Health (NIH) under award P41EB028242. The opinions expressed in this article are the authors' own and do not reflect that of NIH.

References

[1] 2024. *CuesHub, PBC.* https://www.cueshub.com/

[2] Samuel L Battalio, David E Conroy, Walter Dempsey, Peng Liao, Marianne Menictas, Susan Murphy, Inbal Nahum-Shani, Tianchen Qian, Santosh Kumar, and Bonnie Spring. 2021. Sense2Stop: a micro-randomized trial using wearable sensors to optimize a just-in-time-adaptive stress management intervention for smoking relapse prevention. *Contemporary Clinical Trials* 109 (2021), 106534.

[3] Yu Ying Chiu, Ashish Sharma, Inna Wanyin Lin, and Tim Althoff. 2024. A computational framework for behavioral assessment of llm therapists. *arXiv preprint arXiv:2401.00820* (2024).

[4] George P Chrousos. 2009. Stress and disorders of the stress system. *Nature reviews endocrinology* 5, 7 (2009), 374–381.

[5] Marika Cifor and Patricia Garcia. 2020. Gendered by design: A duoethnographic study of personal fitness tracking systems. *ACM Transactions on Social Computing* 2, 4 (2020), 1–22.

[6] John W Creswell and Cheryl N Poth. 2016. *Qualitative inquiry and research design: Choosing among five approaches.* Sage publications.

[7] Shirui Dai, Yongzhen Mo, Yumin Wang, Bo Xiang, Qianjin Liao, Ming Zhou, Xiaoling Li, Yong Li, Wei Xiong, Guiyuan Li, et al. 2020. Chronic stress promotes cancer development. *Frontiers in oncology* 10 (2020), 1492.

[8] Poorvesh Dongre. 2024. Physiology-Driven Empathic Large Language Models (EmLLMs) for Mental Health Support. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems.* 1–5.

[9] Poorvesh Dongre, Majid Behravan, Kunal Gupta, Mark Billinghurst, and Denis Gračanin. 2024. Integrating Physiological Data with Large Language Models for Empathic Human-AI Interaction. *arXiv preprint arXiv:2404.15351* (2024).

[10] Carolyn Ellis, Tony E Adams, and Arthur P Bochner. 2011. Autoethnography: an overview. *Historical social research/Historische sozialforschung* (2011), 273–290.

[11] Empatica. 2023. Empatica Embrace Plus. https://www.empatica.com/embraceplus/ [Accessed: (Feb 18, 2024)].

[12] Garmin. 2022. Stress tracking. https://www.garmin.com/en-US/garmin-technology/health-science/stress-tracking/

[13] Maliheh Ghajargar, Jeffrey Bardzell, and Love Lagerkvist. 2022. A redhead walks into a bar: experiences of writing fiction with artificial intelligence. In *Proceedings of the 25th international academic MindTrek conference.* 230–241.

[14] Martin Gjoreski, Mitja Luštrek, Matjaž Gams, and Hristijan Gjoreski. 2017. Monitoring stress with a wrist device using context. *Journal of biomedical informatics* 73, 159–170.

[15] Thomas G Guilliams and Lena Edwards. 2010. Chronic stress and the HPA axis. *The standard* 9, 2 (2010), 1–12.

[16] Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abdin, Gonzalo Ramos, Tracy Tran, and Mary P Czerwinski. 2022. Design of digital workplace stress-reduction intervention systems: Effects of intervention type and timing. In *CHI Conference on Human Factors in Computing Systems.* 1–16.

[17] Sijie Ji, Xinzhe Zheng, Jiawei Sun, Renqi Chen, Wei Gao, and Mani Srivastava. 2024. MindGuard: Towards Accessible and Sitgma-free Mental Health First Aid via Edge LLM. *arXiv preprint arXiv:2409.10064* (2024).

[18] Rafal Kocielnik and Natalia Sidorova. 2015. Personalized stress management: enabling stress monitoring with lifelogexplorer. *KI-Künstliche Intelligenz* 29 (2015), 115–122.

[19] Rafal Kocielnik, Natalia Sidorova, Fabrizio Maria Maggi, Martin Ouwerkerk, and Joyce HDM Westerink. 2013. Smart technologies for long-term stress monitoring at work. In *proceedings of the 26th IEEE international symposium on computer-based medical systems.* IEEE, 53–58.

[20] Tatiana Kuzmowycz. 2023. Introducing stress monitor: A new way to monitor and manage stress. https://www.whoop.com/thelocker/introducing-stress-monitor-a-new-way-to-monitor-manage-stress/

[21] Tin Lai, Yukun Shi, Zicong Du, Jiajie Wu, Ken Fu, Yichao Dou, and Ziqi Wang. 2023. Psy-llm: Scaling up global mental health psychological services with ai-based large language models. *arXiv preprint arXiv:2307.11991* (2023).

[22] Noelle R Leonard, Marya V Gwadz, Amanda Ritchie, Jessica L Linick, Charles M Cleland, Luther Elliott, and Michele Grethel. 2015. A multi-method exploratory study of stress, coping, and substance use among high school youth in private schools. *Frontiers in psychology* (2015), 1028.

[23] June M Liu, Donghao Li, He Cao, Tianhe Ren, Zeyi Liao, and Jiamin Wu. 2023. Chatcounselor: A large language models for mental health support. *arXiv preprint arXiv:2309.15461* (2023).

[24] Subigya Nepal, Arvind Pillai, William Campbell, Talie Massachi, Michael V Heinz, Ashmita Kunwar, Eunsol Soul Choi, Xuhai Xu, Joanna Kuc, Jeremy F Huckins, et al. 2024. MindScape Study: Integrating LLM and Behavioral Sensing for Personalized AI-Driven Journaling Experiences. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 8, 4 (2024), 1–44.

[25] Sameer Neupane, Mithun Saha, Nasir Ali, Timothy Hnat, Shahin Alan Samiei, Anandatirtha Nandugudi, David M Almeida, and Santosh Kumar. 2024. Momentary Stressor Logging and Reflective Visualizations: Implications for Stress Management with Wearables. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–19.

[26] Jingping Nie, Hanya Shao, Yuang Fan, Qijia Shao, Haoxuan You, Matthias Preindl, and Xiaofan Jiang. 2024. LLM-based Conversational AI Therapist for Daily Functioning Screening and Psychotherapeutic Intervention via Everyday Smart Devices. *arXiv preprint arXiv:2403.10779* (2024).

[27] Aisling Ann O'Kane, Yvonne Rogers, and Ann E Blandford. 2014. Gaining empathy for non-routine mobile device use through autoethnography. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems.* 987–990.

[28] Marta Olasik. 2023. "Good morning, ChatGPT, can we become friends?": an interdisciplinary scholar's experience of 'getting acquainted'with the OpenAI's Chat GPT: an auto ethnographical report. *European Research Studies Journal* (2023).

[29] Simon Ollander, Christelle Godin, Aurélie Campagne, and Sylvie Charbonnier. 2016. A comparison of wearable and stationary sensors for stress detection. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC).* IEEE, 004362–004366.

[30] Pedro Sanches, Kristina Höök, Elsa Vaara, Claus Weymann, Markus Bylund, Pedro Ferreira, Nathalie Peira, and Marie Sjölinder. 2010. Mind the body! Designing a mobile stress management application encouraging personal reflection. In *Proceedings of the 8th ACM conference on designing interactive systems.* 47–56.

[31] Robert M Sapolsky. 2004. *Why zebras don't get ulcers: The acclaimed guide to stress, stress-related diseases, and coping.* Holt paperbacks.

[32] Richard D Sawyer and Darren Lund. 2012. *Duoethnography: Dialogic Methods for Social, Health, and Educational Research.* Vol. 7. Left Coast Press.

[33] Nicolas Schwenke, Heinrich Söbke, and Eckhard Kraft. 2023. Potentials and challenges of chatbot-supported thesis writing: An autoethnography. *Trends in Higher Education* 2, 4 (2023), 611–635.

[34] Fitbit Staff. 2022. Learn more about Fitbit's new body response feature on sense 2. https://blog.fitbit.com/fitbit-new-stress-detection/

[35] Stefan Stepanovic, Vadym Mozgovoy, and Tobias Mettler. 2019. Designing visualizations for workplace stress management: results of a pilot study at a swiss municipality. In *International Conference on Electronic Government.* Springer, 94–104.

[36] Xin Tong, Matthew Louis Mauriello, Marco Antonio Mora-Mendoza, Nina Prabhu, Jane Paik Kim, and Pablo E Paredes Castro. 2023. Just Do Something: Comparing Self-proposed and Machine-recommended Stress Interventions among Online Workers with Home Sweet Office. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* 1–20.

[37] Jaakko Väkevä, Elisa D Mekler, and Janne Lindqvist. 2024. From Disorientation to Harmony: Autoethnographic Insights into Transformative Videogame Experiences. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–20.

[38] Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 1 (1966), 36–45.

[39] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C Schmidt. 2023. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382* (2023).

[40] Johnny Chun Yiu Wong, Jun Wang, Eugene Yujun Fu, Hong Va Leong, and Grace Ngai. 2019. Activity recognition and stress detection via wristband. In *Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia.* 102–106.

[41] Mengru Xue, Rong-Hao Liang, Bin Yu, Mathias Funk, Jun Hu, and Loe Feijs. 2019. AffectiveWall: Designing collective stress-related physiological data visualization for reflection. *IEEE Access* 7 (2019), 131289–131303.

A Appendix

A.1 Design Templates of custom GPT

A.1.1 DeStressify. I am pursuing a PhD in computer science. I primarily work remotely from home on my research, which involves multiple virtual meetings each week, including two research discussions with my professor to review progress and receive guidance. As a PhD student, I often experience stress related to managing complex research tasks, anxiety about future career prospects, and maintaining work-life balance. The goal of this GPT is to assist in developing effective strategies for managing my stress and stressors, maintaining motivation, and sustaining productivity throughout my academic journey. Below are the instructions for this GPT.

(1) **Role Definition:** This GPT should act as a stress intervention specialist trained to provide personalized, actionable advice to help manage stress effectively.
(2) **Tone and Approach:** Ensure the GPT outputs are empathetic, supportive, and practical. Emphasize understanding and acknowledging the user's stressors to foster trust and relatability. Avoid complex jargon; prioritize simplicity and clarity in responses.
(3) **Core Functionality:** Assess the user's input for specific stressors and location details to tailor its advice. Suggest interventions that are realistic and achievable within the user's context.
(4) **Output Requirements:** Provide stress management techniques that directly address the described stressor. Adjust recommendations based on location.
(5) **Constraints and Best Practices:** Avoid generic or overly broad advice; focus on specific, user-centered recommendations. Ensure responses are concise yet detailed enough to guide the user through the intervention.
(6) **Continuous Improvement:** Incorporate feedback mechanisms to allow the user to indicate whether the advice was helpful. Adjust future outputs based on user preferences and past interactions.

After each intervention, the GPT should ask different questions to gather feedback about the relevance and effectiveness of the intervention. Use this feedback to refine future recommendations.

A.1.2 StressGPT. This GPT is a supportive guide to help users manage everyday stress, grounded in the principles of Cognitive Behavioral Therapy (CBT). It focuses on specific CBT techniques, including Cognitive Restructuring, Problem-Solving Skills, Journaling and Thought Records, Socratic Questioning, Reframing, and Stress Inoculation Training (SIT). Stress GPT helps users identify and challenge negative thought patterns, develop healthier coping mechanisms, and build resilience through practical, actionable strategies. It provides personalized advice aligned with the user's preferences, lifestyle, and unique stressors, creating a calm and uplifting space. Responses are empathetic, clear, and aimed at fostering self-awareness and emotional well-being.

A.2 An Illustration of Generic vs Targeted Prompts for DeStressify

Generic Prompt: ***I am stressed right now. Can you help me?***
Response: *It sounds like you're experiencing stress tied to managing the complexities of your PhD journey. The combination of remote work, high-level research demands, career uncertainties, and balancing your personal life in a solitary environment can definitely feel overwhelming. Here's a targeted approach to help you decompress right now:*

Quick Intervention: Progressive Muscle Relaxation (PMR)

This exercise is especially helpful when you're working from home and need to release tension without leaving your desk.

(1) *Sit comfortably in your chair, or lie down if you prefer.*
(2) *Focus on one muscle group at a time, starting from your toes:*
    - *Curl your toes tightly and hold for 5 seconds, then release.*
    - *Move up to your calves: tighten them, hold, and release.*
    - *Work through your thighs, abdomen, chest, arms, and shoulders, ending with your face.*
(3) *With each release, breathe deeply, imagining the tension leaving your body.*
(4) *End by taking three slow, deep breaths and noticing how your body feels.*

This technique not only reduces physical tension but can also clear your mind to approach your tasks more calmly.

Targeted Prompt: ***I am stressed right now because of trying to replicate a research paper. Can you help me?***
Response: *Struggling with understanding and implementing a paper is a common stress point for researchers, and it's entirely valid to feel overwhelmed. Let's break this down into manageable steps to alleviate the stress and help you move forward effectively.*

Intervention: Structured Debugging and Support

(1) *Step Back and Breathe:*
    - *Take a 5-minute break to reset. Practice box breathing:*
        – *Inhale for 4 seconds, hold for 4 seconds, exhale for 4 seconds, and hold again for 4 seconds. Repeat this cycle three times.*
    - *This will calm your nervous system, helping you approach the problem with a clearer mind.*
(2) *Decompose the Problem:*
    - *Break the challenging portion of the implementation into the smallest possible steps or components. Write down exactly what you understand and what is unclear.*
    - *Identify the precise point where the confusion or difficulty arises.*
(3) *Seek Clarification:*
    - *Check for supplementary materials: many papers have public GitHub repositories, appendices, or links to datasets that might provide additional insight.*
    - *If possible, reach out to the authors with a clear, concise email explaining your specific question. Authors often appreciate when their work is being engaged with thoughtfully.*
(4) *Leverage Support Networks:*
    - *Post your issue on forums like Stack Overflow, Reddit's r/MachineLearning, or any academic community.*
    - *If you have peers or colleagues, schedule a quick call or chat to explain your understanding and hear their perspective. Often, talking it out helps untangle the problem.*
(5) *Make Peace with Partial Progress:*
    - *Remind yourself that replication can be iterative. Even understanding just one more piece today is progress.*

Would you like help drafting a specific message to an author or academic forum? Alternatively, we could outline the problematic portion together here, and I could help brainstorm possible explanations or approaches.