Integrating Digital Calendars with Large Language Models for Stress Management Interventions

Pranav Rao†, Sarah Yi Xu†, Ananya Bhattacharjee, Yuchen Zeng, Alex Mariakakis and Joseph Jay Williams

Computer Science, University of Toronto, Toronto, Ontario, Canada

Abstract

Traditional personalized support tools rely on generalized interventions such as scheduled and prewritten notifications and text messages. However, they overlook the dynamic nature of individuals' schedules, which may prominently influence how users receive and react to the intervention. Intervening at the wrong time may lead to less effective outcomes or those deviating from the intended impact. This paper proposes an approach that leverages dynamic contexts from users' digital calendar data, interactions between users and our system, and OpenAI's GPT-4 large language model (LLM) to create time-sensitive, context-aware text messaging interventions that proactively manage stress. Through future work on an associated study involving 20 participants, we will investigate the impact of LLM- versus expert-generated text messages on stress reduction to gather insights on how expert-generated messages can be enhanced and reveal the potential benefits of such an intervention design. Potential findings will highlight necessary factors to consider for a helpful, time-sensitive intervention and important design considerations for LLM-supported tools for stress reduction.

Keywords

Stress Management, Text Messaging, Calendar Data, Large Language Models, Personalized Reflections, GPT-4

1. INTRODUCTION

Stress is a state of mental or emotional strain resulting from demanding circumstances. While stress is a common experience in the workplace [1], academic settings [2], and even in one's personal life, chronic or excessive stress can cause mental and physical illnesses [3]. Common existing stress management methods employ a one-size-fits-all approach, where individual differences in stress responses and preferred coping methods are not accounted for [4, 5, 6].

Digital calendars have evolved into indispensable tools, serving as repositories of our commitments in both personal and professional spheres. They offer a comprehensive snapshot of our daily routines, indicating periods of availability and potential stress triggers, such as frequent

In: Kiemute Oyibo, Wenzhen Xu, Elena Vlahu-Gjorgievska (eds.): The Adjunct Proceedings of the 19th International Conference on Persuasive Technology, April 10, 2024, Wollongong, Australia

† These authors contributed equally.

pranav.rao@mail.utoronto.ca (P. Rao); sarahyixu@gmail.com (S. Y. Xu); ananya@cs.toronto.edu (A. Bhattacharjee); yuchen.zeng@mail.utoronto.ca (Y. Zeng); mariakakis@cs.toronto.edu (A. Mariakakis); williams@cs.toronto.edu (J. J. Williams)

0009-0007-2355-1445 (P. Rao); 0009-0006-9745-4843 (S. Y. Xu); 0000-0002-9116-3766 (A. Bhattacharjee); 0009-0006-3452-5536 (Y. Zeng); 0000-0002-9986-3345 (A. Mariakakis); 0000-0002-9122-5242 (J. J. Williams)

meetings and brief breaks, which have been found to be an indicator of increased workplace stress [7, 8]. Integrating text messaging systems with digital calendars presents a unique opportunity to leverage this wealth of information for personalized stress management interventions. For example, by analyzing calendar data, moments for intervention can be identified, prompting users to take short breaks during hectic schedules.

Traditional approaches to stress reduction often rely on generalized interventions or consistent user input [9], overlooking the nuances of individual schedules and routines [10]. However, a promising approach lies in the intersection of leveraging digital calendar data, Large Language Models (LLMs), and text message interventions. LLMs have demonstrated promise in providing personalized support in practicing mindfulness [11], providing learning feedback [12], and intelligent writing [13]. By utilizing text and time-based digital calendar data, LLMs may possess the power to compose text messages tailored to the user's daily schedules by seeding relevant information from calendars within the messages and sending time-sensitive reminders. Text messaging services have shown success in aiding behavioral change for physical and mental health challenges, including sending motivational quotes to motivate physical exercise [14, 15], smoking cessation [16], and reducing alcohol consumption [17].

We aim to run a study enrolling 20 participants in our text messaging system to investigate the impact of LLM- versus expert-generated text messages on stress reduction. By collecting participants' ratings of text messages, we will analyze factors (such as message content, length, delivery time, selected event) that contribute to a stress-reducing intervention, and gather insights on how expert-generated messages can be enhanced using LLMs. By investigating the potential benefits and challenges of this approach, we aim to contribute to efforts in designing adaptive algorithms that can support behavior change.

2. RELATED WORK

The integration of digital interventions into lifestyle management and mental health care has garnered significant attention from researchers [18]. Informed by previous works, our study combines text message interventions with the use of calendar data and LLMs, each element playing a vital role.

Text messaging interventions are not only cost-effective and prevalent tools, but can successfully induce behavior change in a variety of contexts [7, 19]. Contexts can be internal, such as a person's emotional state, self-efficacy, or motivation; or external, such as their environment and daily schedule. Although text messaging interventions have shown promise in several applications, MacDougall et al. [18] note a clear lack of discussion in existing research about implementation details, including considerations about the "bidirectionality of texting" and the level of tailoring and personalization that current text message interventions offer. In our study, we introduce a bidirectional channel of communication with the user, where each subsequent interaction from the user will enable more potential for personalization in future text messages sent.

Numerous studies have also explored the use of digital calendar data to provide personalized interventions [7, 20, 21, 22]. Howe et al. [7] and Kocielnik et al. [20] analyzed data from popular calendar platforms like Microsoft Outlook to predict patterns of stress and deliver interventions accordingly. Baras et al. [21] used calendar data and mood indicators to send a wide range of encouraging personalized push notifications to students. Tateyama et al. [22] even propose developing a deep-learning model that can predict a subject's mood based on their calendar information. Given the effectiveness of utilizing calendar data to deliver interventions from these previous works, we have incorporated the same into our research alongside tools like LLMs in hopes of maximizing the personalization of text messages sent by our system.

LLMs, despite being evolving tools, have been used to personalize content for users based on provided data in several contexts: as educational chatbots utilizing student data [23], coding assistants utilizing code and prompt data [24], and procrastination-management interventions utilizing users' self-reported circumstances [25]. These existing use cases achieve decent user-reported satisfaction and user experience [23], as well as accuracy and relevance in generated content [24]. This is an encouraging opportunity to leverage LLMs with contextual factors, like calendar data, to tailor messages that are relevant, helpful, and potentially aid in stress relief.

Howe et al. [7] found that, even though digital micro-interventions can effectively reduce short-term stress, personalization of delivery timing and content type can improve user engagement and stress reduction outcomes. The importance of contextual factors is also emphasized by many other previous studies [26, 27, 28]. Bhattacharjee et al. [27] revealed how dynamic context factors, like one's daily schedule, are prominent in influencing how users receive and react to text messages. Their study participants highlighted the importance of message volume and time sensitivity as adaptive factors for a successful text messaging system. Our research into combining the dynamic contexts mentioned has the potential to create time-sensitive, context-aware interventions that proactively manage stress.

3. SYSTEM DESIGN

We are developing a system to help us further understand how personalized text message interventions can reduce stress and improve general mental well-being. Because LLMs are still an emerging tool, we are initially targeting the improvement of general well-being before exploring how LLMs can be deployed in more sensitive contexts, such as ones where mental illnesses are involved.

We demonstrate the sequence of events of a sample user interaction within our system in Figure 1. After users provide their preferred phone number to receive text messages and access to their Google Calendar, they are directed to their dashboard, where they can view past messages received from our system.

The server selects events from a user's calendar, generates associated messages using OpenAI's GPT-4 model [29], and schedules them, all within study parameters (see Section 4). To ensure the relevance of scheduled messages, the system checks for updates to users' calendars every thirty minutes, adjusting scheduled times or message content if necessary. The server is also responsible for handling user feedback via text message and storing it in an external database for analysis.

Privacy was a focus when developing this application. Because all authentication is handled offsite with authentication providers like Google, we do not store any sensitive information such as usernames and passwords. In the future, we look forward to expanding support for alternate digital platforms such as Microsoft Outlook.

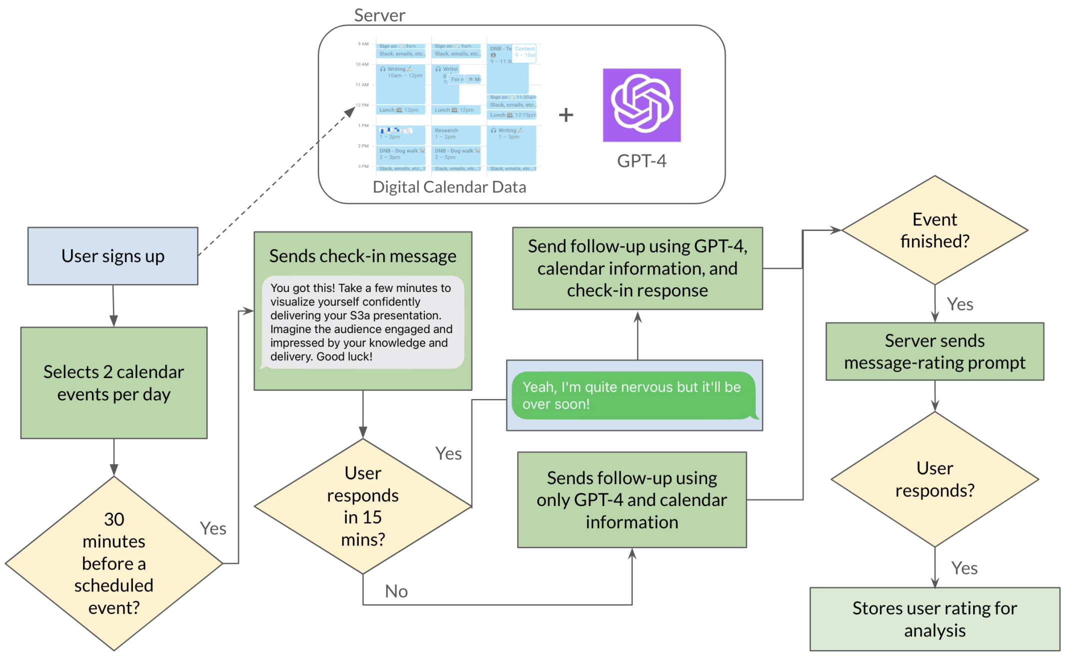

Figure 1: Text Messaging Experience Flow

4. STUDY DESIGN

We aim to investigate how using LLMs to enhance expert-written text messages can improve stress management among post-secondary students. Expert messages are messages designed by experts based on psychology literature, scientifically verified to manage stress. This study employs a randomized controlled trial design with two groups of participants.

The study will involve a total of 20 students enrolled in a large Canadian university. Participants will be recruited to use our text messaging service, for which they are required to provide permission to access their digital calendars. We will employ a within-subjects design, where they will receive personalized messages tailored to their schedules over the period of two weeks. Some messages will be written by experts, and others by an LLM that has been given a sample bank of expert messages and additional contextual information. While users will receive an equal number of messages under both conditions, the messages will be sent in randomized order to avoid the effects of early or late receipt.

Participants will receive two scheduled text messages per day. The events will be randomly selected, but one will be for an event before lunch and one will be for an event after to ensure sufficient spacing between messages. In the LLM condition, thirty minutes before a selected event on their calendar, participants will receive a prompt asking about their emotional and mental state. A 15-minute response window will be provided; if they reply, their response, as well as the summary and description of the calendar event, will be given to the LLM to inform the subsequent personalized message. If no response is received within the window, a follow-up message will be sent based solely on the user's calendar information.

In both conditions, following each scheduled event for which participants receive text messages, they will be prompted to rate the effectiveness of the message in aiding stress management following that event. Rating prompts will include questions such as, "How prepared did this text message make you feel for the event?" Although prompts may vary, ratings will be quantitative and on a numerical scale, though an opportunity for additional text feedback may be provided. At the end of the two weeks, we will conduct interviews with participants to collect qualitative feedback to gather more nuanced insights into their experiences.

This study design allows for an investigation into how LLMs can enhance expert-written intervention messages. Analyzing participant ratings and responses from the interviews will provide insights into the potential of LLMs in tailoring messages to individuals' schedules and routines to enhance stress management outcomes.

5. CHALLENGES AND NEXT STEPS

The progression of this study has encountered some notable obstacles, some of which merit further acknowledgement and discussion. For instance, the quality of text messages the system sends may depend on the actions of study participants within a time constraint. If a participant is consistently unable or unwilling to respond to the first text message within the 15-minute window, interventions will be solely based on the information from that participant's calendar, limiting the LLM's ability to respond in a personalized manner. To combat this, we intend to investigate further contextual factors we can collect to diversify sources of personalization. We are currently exploring collecting additional user feedback prior to the study and potentially incorporating data from mobile health sensors.

In the near future, we also hope to increase the robustness of the application itself by utilizing GPT-4 to analyze user interactions and dynamically select the events for which to send interventions, rather than simply retaining the data for manual analysis. This feature is one we believe will greatly boost the mental well-being of users, as it will help prioritize events for which they most require interventions.

6. WORKSHOP ACTIVITY

During the workshop, after going over our study, we would aim to engage participants' insights regarding the following questions:

- What are meaningful contextual factors we can manipulate with LLMs aside from calendar information and mobile sensor data?
- Is the two-week period long enough to capture diverse changes in user schedules?
- Aside from calendar information and explicit user feedback, what other information might we be able to analyze with an LLM to dynamically select events that are pertinent to the user?

- This paper proposes sending a message 30 minutes before an event and waiting for 15 minutes. Are there alternate timings that could be more effective for getting user responses, and why?
- How can we more concretely measure stress in users, aside from just using self-reported ratings and interviews?
- Given the closed-source nature of GPT-4 and ChatGPT, how can we effectively identify reproducible factors in successful or unsuccessful approaches within this study?

7. ACKNOWLEDGEMENTS

We thank Drs. Rachel Kornfield (Northwestern University) and Renwen Zhang (National University of Singapore) for their input on working on LLMs and digital well-being. This work was supported by grants from the National Institute of Mental Health (K01MH125172, R34MH124960), the Office of Naval Research (N00014-18-1-2755, N00014-21-1-2576), the Natural Sciences and Engineering Research Council of Canada (RGPIN-2019-06968), and the National Science Foundation (2209819). In addition, we acknowledge a gift from the Microsoft AI for Accessibility program (http://aka.ms/ai4a), and awards from the Wolfond Scholarship Program in Wireless Information Technology and Inlight Student Mental Health Research Initiative, an Institutional Strategic Initiative at the University of Toronto.

References

[1] C. Ryan, M. Bergin, T. Chalder, J. S. G. Wells, Web-based interventions for the management of stress in the workplace: Focus, form, and efficacy, Journal of Occupational Health 59 (2017) 215–236. doi:10.1539/joh.16-0227-RA.

[2] M. J. Chambel, L. Curral, Stress in academic life: work characteristics as predictors of student well-being and performance, Applied psychology 54 (2005) 135–147.

[3] N. Schneiderman, G. Ironson, S. D. Siegel, Stress and health: psychological, behavioral, and biological determinants, Annu. Rev. Clin. Psychol. 1 (2005) 607–628.

[4] I. Nahum-Shani, S. N. Smith, B. J. Spring, L. M. Collins, K. Witkiewitz, A. Tewari, S. A. Murphy, Just-in-time adaptive interventions (jitais) in mobile health: key components and design principles for ongoing health behavior support, Annals of Behavioral Medicine (2018) 1–17.

[5] R. Kornfield, D. C. Mohr, R. Ranney, E. G. Lattie, J. Meyerhoff, J. J. Williams, M. Reddy, Involving crowdworkers with lived experience in content-development for push-based digital mental health tools: lessons learned from crowdsourcing mental health messages, Proceedings of the ACM on Human-computer Interaction 6 (2022) 1–30.

[6] Y. Fukuoka, C. Gay, W. Haskell, S. Arai, E. Vittinghoff, et al., Identifying factors associated with dropout during prerandomization run-in period from an mhealth physical activity education study: The mped trial, JMIR mHealth and uHealth 3 (2015) e3928.

[7] E. Howe, J. Suh, M. Bin Morshed, D. McDuff, K. Rowan, J. Hernandez, M. I. Abdin, G. Ramos, T. Tran, M. P. Czerwinski, Design of digital workplace stress-reduction intervention

systems: Effects of intervention type and timing, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22, Association for Computing Machinery, New York, NY, USA, 2022. URL: https://doi.org/10.1145/3491102.3502027. doi:10.1145/3491102.3502027.

[8] G. Mark, S. Iqbal, M. Czerwinski, How blocking distractions affects workplace focus and productivity, in: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, UbiComp '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 928–934. URL: https://doi.org/10.1145/3123024.3124558. doi:10.1145/3123024.3124558.

[9] P. Paredes, R. Gilad-Bachrach, M. Czerwinski, A. Roseway, K. Rowan, J. Hernandez, Poptherapy: Coping with stress through pop-culture, in: Proceedings of the 8th international conference on pervasive computing technologies for healthcare, 2014, pp. 109–117.

[10] L. Wang, L. C. Miller, Just-in-the-moment adaptive interventions (jitai): a meta-analytical review, Health Communication 35 (2020) 1531–1544.

[11] H. Kumar, Y. Wang, J. Shi, I. Musabirov, N. A. S. Farb, J. J. Williams, Exploring the use of large language models for improving the awareness of mindfulness, in: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, CHI EA '23, Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3544549.3585614. doi:10.1145/3544549.3585614.

[12] S. Uchiyama, K. Umemura, Y. Morita, Large language model-based system to provide immediate feedback to students in flipped classroom preparation learning, 2023. arXiv:2307.11388.

[13] M. Lee, P. Liang, Q. Yang, Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities, in: CHI Conference on Human Factors in Computing Systems, CHI '22, ACM, 2022. URL: http://dx.doi.org/10.1145/3491102.3502030. doi:10.1145/3491102.3502030.

[14] L. Duro, P. F. Campos, T. Romão, E. Karapanos, Visual quotes: Does aesthetic appeal influence how perceived motivating text messages impact short-term exercise motivation?, in: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI EA '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 1–6. URL: https://doi.org/10.1145/3290607.3312830. doi:10.1145/3290607.3312830.

[15] G. N. Horner, S. Agboola, K. Jethwani, A. Tan-McGrory, L. Lopez, Designing patient-centered text messaging interventions for increasing physical activity among participants with type 2 diabetes: Qualitative results from the text to move intervention, JMIR Mhealth Uhealth 5 (2017) e54. URL: http://www.ncbi.nlm.nih.gov/pubmed/28438728. doi:10.2196/mhealth.6666.

[16] S. Haug, M. P. Schaub, V. Venzin, C. Meyer, U. John, Efficacy of a text message-based smoking cessation intervention for young people: a cluster randomized controlled trial, Journal of medical Internet research 15 (2013) e171.

[17] B. Suffoletto, T. Chung, F. Muench, P. Monti, D. B. Clark, et al., A text message intervention with adaptive goal support to reduce alcohol consumption among non-treatment-seeking young adults: non-randomized clinical trial with voluntary length of enrollment, JMIR

mHealth and uHealth 6 (2018) e8530.

[18] S. MacDougall, S. Jerrott, S. Clark, L. A. Campbell, A. Murphy, L. Wozney, Text message interventions in adolescent mental health and addiction services: Scoping review, JMIR Ment Health 8 (2021) e16508. URL: https://doi.org/10.2196/16508. doi:10.2196/16508.

[19] D. M. Smith, L. Duque, J. C. Huffman, B. C. Healy, C. M. Celano, Text message interventions for physical activity: A systematic review and meta-analysis, American Journal of Preventive Medicine 58 (2020) 142–151. URL: https://doi.org/10.1016/j.amepre.2019.08.014. doi:10.1016/j.amepre.2019.08.014.

[20] R. Kocielnik, N. Sidorova, F. M. Maggi, M. Ouwerkerk, J. H. Westerink, Smart technologies for long-term stress monitoring at work, in: proceedings of the 26th IEEE international symposium on computer-based medical systems, IEEE, 2013, pp. 53–58.

[21] K. Baras, L. Soares, N. Paulo, R. Barros, 'smartphine': Supporting students' well-being according to their calendar and mood, in: 2016 International multidisciplinary conference on computer and energy science (SpliTech), IEEE, 2016, pp. 1–7.

[22] N. Tateyama, R. Fukui, S. Warisawa, Mood prediction based on calendar events using multitask learning, IEEE Access 10 (2022) 79747–79759.

[23] P. Denny, V. Kumar, N. Giacaman, Conversing with copilot: Exploring prompt engineering for solving cs1 problems using natural language, in: Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1, SIGCSE 2023, Association for Computing Machinery, New York, NY, USA, 2023, p. 1136–1142. URL: https://doi.org/10.1145/3545945.3569823. doi:10.1145/3545945.3569823.

[24] P. Denny, V. Kumar, N. Giacaman, Conversing with copilot: Exploring prompt engineering for solving cs1 problems using natural language, in: Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1, SIGCSE 2023, Association for Computing Machinery, New York, NY, USA, 2023, p. 1136–1142. URL: https://doi.org/10.1145/3545945.3569823. doi:10.1145/3545945.3569823.

[25] A. Bhattacharjee, Y. Zeng, S. Y. Xu, D. Kulzhabayeva, M. Ma, R. Kornfield, S. I. Ahmed, A. Mariakakis, M. P. Czerwinski, A. Kuzminykh, M. Liut, J. J. Williams, Understanding the role of large language models in personalizing and scaffolding strategies to combat academic procrastination, 2023. arXiv:2312.13581.

[26] W. Chen, C. Yu, H. Wang, Z. Wang, L. Yang, Y. Wang, W. Shi, Y. Shi, From gap to synergy: Enhancing contextual understanding through human-machine collaboration in personalized systems, in: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, UIST '23, Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3586183.3606741. doi:10.1145/3586183.3606741.

[27] A. Bhattacharjee, J. J. Williams, J. Meyerhoff, H. Kumar, A. Mariakakis, R. Kornfield, Investigating the role of context in the delivery of text messages for supporting psychological wellbeing, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3544548.3580774. doi:10.1145/3544548.3580774.

[28] A. Bhattacharjee, J. Pang, A. Liu, A. Mariakakis, J. J. Williams, Design implications for one-way text messaging services that support psychological wellbeing, ACM Trans. Comput.-Hum. Interact. 30 (2023). URL: https://doi.org/10.1145/3569888. doi:10.1145/3569888.

[29] OpenAI, Gpt-4 technical report, 2024. arXiv:2303.08774.