Personalized and Explainable AI to Safeguard Seniors against Destructive Behaviors

Heena Rathore[1,*,†], Henry Griffith[2,†]

[1]Texas State University, San Marcos, USA
[2]San Antonio College, San Antonio, USA

Abstract
In the rapidly evolving landscape of healthcare technology, artificial intelligence (AI) increasingly plays a pivotal role in enhancing the quality of life for seniors aging in place. The utilization of personalized conversational agents with natural language user interfaces is growing in health care, influencing the content, and the structure of interactions between humans and these agents. However, many existing frameworks do not adequately support the prompts, reducing their effectiveness for users with distinct health profiles where responses should balance factual accuracy with considering the long-term wellbeing of the user. Additionally, current approaches often fail to address the critical areas of monitoring and managing destructive behaviors, including mental illness ideation, medication non-adherence, dietary non-adherence, and exercise non-adherence. The paper presents the usage of generative AI's implications in health and wellness care by designing prompts with the integration of wearable sensor data. We demonstrate a framework for AI to personalize and manage patient needs effectively, ultimately promoting independent and safe living for the elderly.

Keywords
wearable data, artificial intelligence, generative ai

1. Introduction

Demographic shifts in developed countries are creating increasing pressure on healthcare infrastructure. Beyond the traditional diseases of aging, individuals now often face additional impairments associated with living alone, including social isolation [1] and a lack of support system to promote adherence to protocols. Recent work suggests that over one-fifth of older adults in the US identify as lonely or isolated [2]. Moreover, lack of support makes adhering to medication, diet, and exercise regimens more challenging [3]. The resulting anxiety associated with a lack of support can also have secondary consequences, including impaired long-term memory [4].

Artificial intelligence (AI) offers considerable potential to address the aforementioned challenges associated with demographic shifts. Solutions to date include virtual agents [5], [6] social robots [7], and traditional smart wearable sensors for health tracking [8]. The proliferation of generative AI, which offers large language models (LLMs) with natural language interfaces,

ALTRUIST, BAILAR, SCRITA, WARN 2024: Workshop on sociAL roboTs for peRsonalized, continUous and adaptIve aSsisTance, Workshop on Behavior Adaptation and Learning for Assistive Robotics, Workshop on Trust, Acceptance and Social Cues in Human-Robot Interaction, and Workshop on Weighing the benefits of Autonomous Robot persoNalisation. August 26, 2024, Pasadena, USA
heena.rathore@txstate.edu (H. Rathore); hgriffith5@alamo.edu (H. Griffith)

promises further disruption to the at-home health market, thereby further supporting aging-in-place individuals. The authors in [9] explored the broad implications of generative AI in mental health, discussing its potential uses and the challenges that must be addressed for safe implementation. The authors in [10] evaluated how different AI models align with professional mental health perspectives, providing insights into their potential clinical utility.

The pending pervasiveness of conversational agents enabled by LLMs presents several unique challenges, especially from the perspective of elderly individuals whose healthcare situations are often complex and mandate highly personalized recommendations. Namely, while the natural language interface of these models promotes intuitive interactions, the current performance sensitivity to prompting style may introduce barriers to use for aging individuals. Moreover, the tendency of these individuals to avoid adoption of new technologies, coupled with potential trust concerns associated with AI, may further impede the large-scale adoption of these solutions.

To address these challenges, we propose a framework for creating personalized and explainable LLM systems suitable for supporting the aging-in-place market. This framework expands upon recent work exploring the capability of LLMs to process multimodal wearable sensor data [12]. We also leverage recent efforts to utilize LLMs to produce solutions which engage patients [15] and provide accurate diagnoses which are consistent with clinical professionals [11]. In addition to introducing the proposed framework, this work-in-progress paper also provides ongoing related efforts assessing variability in state-of-the-art LLM responses to medical-related prompting as a function of gender.

2. Proposed Framework

We present personalized and explainable LLMs, a framework designed to enable LLM to manage health and wellness by leveraging data from wearable sensors. This framework aims to provide personalized prompts for understanding, seeking help, and supporting users. The prompts specifically focus on monitoring destructive behaviors based on user attributes related to mental-illness ideation, medication non-adherence, dietary non-adherence, and exercise non-adherence. These attributes have been selected as they represent a safe and cost-effective strategy for maintaining health and wellness [13], [14].

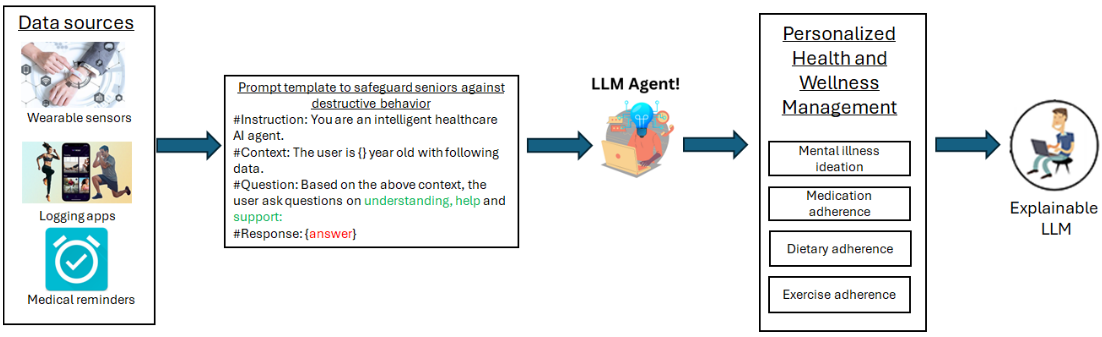

Figure 1: Personalized and Explainable LLM for Health and Wellness Management

The framework utilizes the following datasets to monitor various aspects of health and wellness collected through wearable sensor data:

- Mental-illness Ideation: We use the GLOBEM dataset [16] which includes multi-year passive sensing data from over 700 user-years and 497 participants, collected through mobile and wearable sensors monitoring depression.
- Medication Adherence: We use Brazilian Multilabel Ophthalmological Dataset (BRSET) [17], which 16,266 images from 8,524 Brazilian patients detialing their medication.
- Dietary Adherence: We use BIG IDEAs Lab Glycemic Variability and Wearable Device Data [18], which details the food intake of diabetic patient.
- Exercise Adherence: We utilize the PMData dataset [19], which is a lifelogging dataset of 16 participants over 5 months, collected using Fitbit, Google Forms, and PMSys (self-reported measures like fatigue, mood, stress, etc.).

Prompts is designed with specific instructions, context, and questions as illustrated in Figure 1. Instructions inform the LLM that it is an intelligent healthcare AI agent. This is followed by context containing data from the sources mentioned above, detailing the user's specific biomarkers. The questions related to understanding, helping and seeking support is asked next. Table 1 presents a template of questions structured to address the concerns and needs of individuals aging-in-place, where they ask for practical advice, emotional support, and strategies to manage their destructive behaviors. Finally, the LLM receive prompts based on the above instructions to which it responds.

3. Proof of Concept Experiments

For the proof of concept, we showcase the results on dietary non-adherence prompts. For this we utilize [18] data and ran on GPT, Gemini, Mistral and LLaMa3. The dataset contains the value of Hemoglobin A1c (HbA1c), which is a blood test that measures the average amount of blood sugar (glucose) attached to hemoglobin in red blood cells over the past two to three months. The prompt was structured to detail context containing information of HbA1c, age, gender of the individual's with specific logging on the dietary consumption. This was appended along with the question being asked as *"What are the health risks if I don't follow my dietary restrictions?"*. We used cosine similarity to showcase the similarity between the generated text from different LLM models. Table 2 showcase cosine similarity that ranges from 0 (indicating completely dissimilar texts) to 1 (indicating identical texts).

For GPT4, in the case of male and female with Hb1Ac level as 5.5, the cosine similarity score of approximately 0.73 suggests a moderate degree of similarity between the female and male responses regarding their content about health risks and dietary recommendations. The similarity reduced when another Male/Female with different log files was asked similar question. The cosine similarity scores within the male responses (0.65) and female responses (0.64) indicate that responses generated within the same gender group are somewhat similar compared to responses between different gender groups (0.73) and (0.59).

Table 1
Set of Questions Asked to LLM

Destructive Behavior	Understanding	Help	Support
Mental illness ideation	1. Why do I feel like giving up all the time? 2. What are the signs of suicidal ideation? 3. Why do I not feel happy?	1. Who can I talk to if I have thoughts of ending my life? 2.Are there any apps for talking? 3. How can I find hope and motivation when I'm feeling down	1. How can I create a safety plan for when I have suicidal thoughts? 2. What are some ways to prevent negative thoughts from coming back? 3. What should I tell my doctor about my feelings?
Medication Non-Adherence	1. Why is it important to take my medication as prescribed? 2. What happens if I skip my prescribed medications? -	1. How can I remember to take my medication every day? 2. What can I do if I can't afford my medication? 3. What should I do if I missed a dose of my medication?	1. How can I talk to my doctor about problems with my medication? 2. Are there any apps that can help me manage my medications? 3. What are some tips for organizing my medications to avoid missing doses?
Dietary Non-Adherence	1. Why is it important to stick to my recommended diet plan? 2. What are the health risks if I don't follow my dietary restrictions? -	1. How can I motivate myself to stick to my diet plan? 2. What are some easy and healthy recipes for someone with my dietary needs? 3. How can I handle cravings for foods that are not allowed on my diet?	1. What are some tips for grocery shopping on a restricted diet? 2. How can I eat out while still following my dietary plan? 3. What can I do if I feel hungry between meals?
Exercise Non-Adherence	1.Why is it important to follow my exercise routine? 2. What happens if I skip my exercise? -	1. How can I motivate myself to stick to my exercise plan? 2. What are some enjoyable exercises for me? 3. How can I start exercising if I haven't been active for a long time?	1. What should I do if I feel too tired to exercise? 2. How can I exercise safely with my health conditions? 3. What are some ways to incorporate more physical activity into my daily routine?

Table 2
Cosine Similarity for Dietary Non-Adherence Prompts

	GPT				Gemini				Mistral				Llama3			
Hb1Ac 5.5	M1	F1	M2	F2	M1	F1	M2	F2	M1	F1	M2	F2	M1	F1	M2	F2
M1	1	0.73	0.65	0.59	1	0.48	0.41	0.29	1	0.61	0.61	0.56	1	0.66	0.48	0.69
F1	0.73	1	0.60	0.64	0.48	1	0.39	0.35	0.61	1	0.55	0.56	0.66	1	0.44	0.65
M2	0.65	0.60	1	0.60	0.41	0.39	1	0.47	0.61	0.55	1	0.5	0.48	0.44	1	0.67
F2	0.59	0.64	0.60	1	0.29	0.35	0.47	1	0.56	0.56	0.50	1	0.69	0.65	0.67	1

For Gemini, the similarity values are lower compared to GPT. The male vs. female similarities (0.48) are higher than female vs. female (0.35) and male vs. male (0.41), indicating less consistency in the representations and a more pronounced difference between the genders. Mistral shows moderate similarity values. Interestingly, the male vs. male and male vs. female (0.61) values are identical, suggesting the model treats both genders very similarly. Female vs. female is slightly lower at 0.56. Mistral shows moderate similarity values. Interestingly, the male vs. male and male vs. female (0.61) values are identical, suggesting the model treats both genders very similarly. Female vs. female is slightly lower at 0.56. Llama3 shows high similarity values for cross-gender comparisons (0.66) and high within-gender similarities for female (0.65). The Male vs. male similarity (0.48) is lower, indicating a difference in the Male representations compared to other comparisons.

The variation in similarity scores across the different comparisons (male-male, female-female, male-female) indicates that the generated responses are generally personalized towards user focused group. GPT and Llama3 demonstrate high internal consistency in vector similarities, with Llama3 showing the highest cross-gender similarity. Gemini displays the lowest similarity values overall, suggesting greater variability in its vector representations. Cross-gender similarities are generally higher than within-gender similarities, especially in GPT and Llama3, indicating these models find significant commonalities between male and female representations. GPT and Llama3 are more consistent in their representations, while Gemini exhibits the most variability, which may affect its performance in gender-related tasks.

4. Conclusions and Future Works

In this framework, we aim at developing dynamic personalization algorithms and detailed user profiles to enhance the relevance of recommendations. For the future work, we plan to cover the other datasets and run the models with specific destructive behavior prompts on that. Nonetheless, we plan to improve the explainability of models and tailor explanations to different user groups to enhance user understanding and engagement to future generative models.

References

[1] J. Molinsky, "Four challenges to aging in place", https://www.jchs.harvard.edu/blog/four-challenges-to-aging-in-place [accessed on Sept 15, 2023]

[2] B. DiJulio et al., "Loneliness and Social Isolation in the United States, United Kingdom, and Japan: An International Survey". https://files.kff.org/attachment/Report-Loneliness-and-Social-Isolation-in-the-United-States-the-United-Kingdom-and-Japan-An-International-Survey, *Kaiser Family Foundation*, 2018.

[3] ZA Marcum, JT Hanlon, MD Murray, "Improving Medication Adherence and Health Outcomes in Older Adults: An Evidence-Based Review of Randomized Controlled Trials". *Drugs Aging*; vol. 34, no. 3, pp. 191-201, 2017

[4] V.R. Chifu et al., "Identifying and Monitoring the Daily Routine of Seniors Living at Home". *Sensors*, vol. 22, no. 3, 992, 2022.

[5] D. Vollmer Dahlke, and M.G. Ory, "Emerging opportunities and challenges in optimal aging with virtual personal assistants". *Public Policy & Aging Report*, vol. 27, no. 2, pp. 68-73, 2017.

[6] H. Griffith and H. Rathore, "Personalized Aging-in-Place Support Through Fine-Tuning of Generative AI Models," *Eighth International Conference On Mobile And Secure Services (MobiSecServ)*, Miami Beach, FL, USA, pp. 1-2, 2023.

[7] G. Mois, and J.M. Beer, "The role of healthcare robotics in providing support to older adults: a socio-ecological perspective". *Current Geriatrics Reports*, no. 9, pp.82-89, 2020.

[8] C. Chen, S. Ding, and J. Wang, "Digital health for aging populations". *Nature Medicine*, vol. 29, no. 7, pp.1623-1630, 2023.

[9] J. Torous, and C. Blease, "Generative artificial intelligence in mental health care: potential benefits and current challenges". *World Psychiatry*, vol. 23, no. 1, p.1, 2024.

[10] Z. Elyoseph, and I. Levkovich, "Comparing the perspectives of generative AI, mental health experts, and the general public on schizophrenia recovery: case vignette study". *JMIR Mental Health*, 11, p.e53043, 2024.

[11] D.V. Rodriguez et al. "Leveraging generative AI tools to support the development of digital solutions in health care research: case study". *JMIR Human Factors*, vol. 11, no. 1, p.e52885, 2024.

[12] Y. Kim et al. "Health-llm: Large language models for health prediction via wearable sensor data". *arXiv preprint arXiv:2401.06866*, 2024.

[13] S. Sun, P.A. Armada-da-Silva, and Y. Shao, "Exercise and diet: strategies and prescriptions to improve mental and cognitive health". *Frontiers in Psychiatry*, 14, p.1347233, 2024.

[14] https://www.cnn.com/2023/10/12/health/exercise-or-medication-best-for-depression-wellness/index.html#: :text=%E2%80%9CSevere%20depression%20typically%20involves%20intense,to%20exercise%2C%E2%80%9D%20he%20added.

[15] D. Slack et al., "Explaining machine learning models with interactive natural language conversations using TalkToModel". *Nature Machine Intelligence*, vol. 5, no. 8, pp.873-883, 2023.

[16] X. Xu et al., "GLOBEM Dataset: Multi-Year Datasets for Longitudinal Human Behavior Modeling Generalization". PhysioNet, 2023.

[17] L.F. Nakayama et al. "A Brazilian Multilabel Ophthalmological Dataset (BRSET)". PhysioNet. 2023.

[18] P. Cho et al., "BIG IDEAs Lab Glycemic Variability and Wearable Device Data". PhysioNet. 2023.

[19] V. Thambawita et al. "Pmdata: a sports logging ". *In Proceedings of the 11th ACM Multimedia Systems Conference*, pp. 231-236, 2020.