

Research Letter

Leveraging Multimodal Large Language Models for Fall Risk Reduction in Older Adults in the Home: Proposed Model Design

Justin Do¹, BS; Vivaswat Suresh², BS; Lily Zhang¹, BA; Bharvi M Chavre¹, BA; Jeremy Cha¹, BA; Robert Pugliese^{1,2}, PharmD

¹Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, United States

²Health Design Lab, Thomas Jefferson University, Philadelphia, PA, United States

Corresponding Author:

Justin Do, BS
Sidney Kimmel Medical College, Thomas Jefferson University
925 Chestnut St., Basement Vault
Philadelphia, PA 19107
United States
Phone: 1 4084718003
Fax: 1 2154841949
Email: jhd012@students.jefferson.edu

Abstract

This research letter proposes a novel model design leveraging natively multimodal large language models to identify fall risks and generate visualizations of recommended home environmental modifications, aiming to improve the accessibility and impact of personalized fall prevention advice for older adults. Through a pilot rating study, this work demonstrates that multimodal large language models can generate safe and actionable advice to reduce fall risk in lived spaces of older adults, and also generate realistic edits based on original images. While this concept needs further testing and clinical comparison, it highlights a promising avenue for further innovation of fall prevention tactics.

JMIR Aging 2026;9:e77591; doi: [10.2196/77591](https://doi.org/10.2196/77591)

Keywords: LLM; multimodal; image generation; fall risk; older adults; large language model

Introduction

Falls among older adults cause significant mortality and increased health care costs [1]. Current literature has identified combined behavioral and exercise interventions as effective preventions for fall risks, improving balance performance, and reducing fear of falls [2], while limited evidence exists regarding medication-induced falls [3]. Home environmental intervention is also effective: safety assessments have been shown to reduce fall rate by 23%-36% [1], with applied home modifications contributing a 7% risk reduction [4]. However, both external (insurance) and self-imposed (ie, the perception that safety assessments are invasive) barriers impede widespread implementation [5]. While research on frailty assessments is robust, gaps remain in technology-enabled interventions [6]. Prior studies have shown acceptability by older adults to embrace digital and electronic tools [5]. Existing remote home assessment protocols rely on caregiver camera operation,

written instructions comprehension, and professional review of footage [7], while telehealth occupational therapy (OT) assessments may require insurance authorization, creating both obstacles and delays. Multimodal large language models (LLM) can fuse visual and text information, offering a scalable alternative while preventing encroachment on user values. This study aims to evaluate the ability of LLMs to produce safe, clinically useful, and actionable outputs that identify fall risks from user-provided home imagery and uniquely generate visualizations of the recommended environmental modifications.

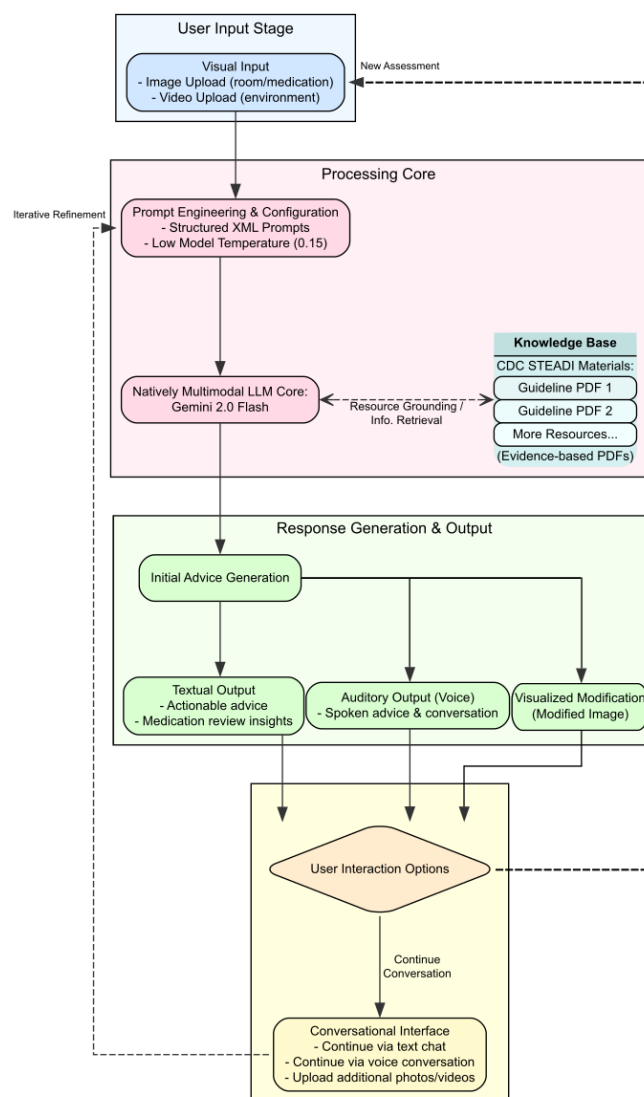
Methods

We selected Google's Gemini family due to its strong visual reasoning performance supported by validated benchmarks [8]. We designed our framework to focus on providing reliable output by employing a low model temperature (0.15), in-context learning through grounding responses

in evidence-based CDC STEADI patient materials (Figure 1), and structured XML prompts iterated using artificial intelligence (AI)-driven prompt engineering. The core innovation of this study used the gemini-2.0-flash-exp-image-generation model to directly modify the uploaded images with the model’s suggested changes. The model leverages a two-shot prompting system (Multimedia Appendix 1, Multimedia Appendix 2), in which the primary LLM generates textual recommendations based on an input image/video, then directs the image generation model to visually render these changes (eg, adding grab bars, removing

hazards) onto the original image, iterating until the generated image reflects the proposed modifications. We conducted a formative, blinded, paired comparison of outputs generated from 27 publicly licensed “lived-in” home interior images. We compared a non-optimized baseline prompt with an enhanced multimodal pipeline (“Steady”). Text and image outputs were compared based on clinical usefulness, safety, image fidelity/plausibility, and preference between baseline prompt output and our enhanced multimodal pipeline output. Detailed methods and output can be found in Multimedia Appendices 3 and 4.

Figure 1. Enhanced model architecture for advice generation and multimodal interaction. CDC: Centers for Disease Control and Prevention; LLM: large language model; STEADI: Stopping Elderly Accidents, Deaths, & Injuries.



Results

Initial Advice Generation, Multimodal Communication, and Modification Visualization

The model takes an uploaded image or video and provides specific, actionable advice supported by evidence-based

resources. The proposed architecture successfully applies both additive and subtractive modifications to images, providing users with a concrete visual representation of a safer environment.

Model Comparisons

We find that overall, our raters preferred the “Steady” system output of both image and text (40/54 times (74.1%), Table 1).

We demonstrate that contemporary LLMs produce relatively safe recommendations, regardless of the prompting system, with only one set of recommendations rated as unsafe in the baseline prompting system and none in the enhanced. We find that when given specific issues to visualize, image editing LLMs produce edits with good visualization fidelity

(46/54 times, 85.2% for baseline and 43/54 times, 79.6% for enhanced; [Table 1](#)), and low rates of implausible/hazard-producing edits (6/54 times, 11.1% for both systems; [Table 1](#)). Text recommendations and visualized outputs ranged from generally “somewhat actionable” for the baseline system to highly actionable for our enhanced system.

Table 1. Results of rating study.

Outcome	Baseline	Enhanced	Comparison ^a
Q1 Overall clinical usefulness (preference)	Preferred: 10/54 (18.5%)	Preferred: 40/54 (74.1%)	Win rate: 40/50 (80.0%, 95% CI 67.0-88.8); sign test $P \leq .001$; ties=4
Q2 Unsafe/inappropriate recommendation rate (Yes)	1/54 (1.9%, 95% CI 0.3-9.8)	0/54 (0.0%, 95% CI 0.0-6.6)	Risk difference (Enhanced-Baseline): -1.9 pp
Q3 Visualization fidelity rate (Yes)	46/54 (85.2%, 95% CI 73.4-92.3)	43/54 (79.6%, 95% CI 67.1-88.2)	Risk difference (Enhanced-Baseline): -5.6 pp
Q4 Hazard-introducing/implausible edit rate (Yes)	6/54 (11.1%, 95% CI 5.2-22.2)	6/54 (11.1%, 95% CI 5.2-22.2)	Risk difference (Enhanced-Baseline): +0.0 pp
Q5 Actionability (1-5 Likert)	median 3.0 (IQR 3.0-4.0)	median 5.0 (IQR 4.0-5.0)	Δ median (Enhanced-Baseline): +1.0 (IQR 0.0-2.0)

^an=rater case evaluations; outcomes are descriptive; Q1 sign test is exploratory.

Discussion

Principal Findings

This study introduces a novel application of multimodal LLMs, leveraging their image-generation capabilities for visualizing personalized home safety recommendations. We demonstrate that enhanced frameworks, such as structured prompting and grounding using trusted resources, produce safe, clinically useful, and actionable outputs that categorically rate better than outputs from baseline LLMs. The inherent flexibility of LLMs supports diverse interaction methods, uniquely enabling users to interact with their “consultant” in their preferred mode. LLMs may mitigate delays caused by insurance authorizations and restore autonomy to users.

The visual output capability is also key: generating suggestions directly onto uploaded images offers more intuitive, actionable guidance than abstract text instructions alone. The drive to protect the familiarity of their home from change was identified to be a major motive for older adults rejecting modification advice from OT [5]; direct visualization of user-fed images may help overcome this hurdle and increase acceptance. There are still limitations to the technology, namely outputs may be illogical such as the recommended soap placement, and movement of furniture and door in [Multimedia Appendix 1](#). However, overall, this study demonstrates that LLMs generally produce visual outputs with high fidelity, low hazard introduction rates, and high actionability.

To adhere to HIPAA (Health Insurance Portability and Accountability Act) compliance, future work should consider

working with LLM providers to sign a HIPAA Business Associate Amendment or other HIPAA-compliant program. Ethical considerations, such as disclosure of privacy and data protection, should be implemented in accordance with WHO guidance on AI in health [9].

Limitations

Further testing must be conducted against the current standard for in-home assessments to discover if the proposed model provides comparable advice to professionals. Implementation trials will be needed to mitigate concerns such as the digital divide and ensure accessibility among varying cognitive/visual functions. Implementations must comply with FDA digital health guidance, and characterization and limitation of unsafe output generation must be explored. This model is designed as a supplemental service to be integrated with OT rather than a replacement.

Conclusions

Multimodal LLMs that integrate image generation offer a novel, innovative approach to increasing end users’ accessibility to personalized home environment recommendations for fall prevention. This capability represents a potential supplement to current care services that may enhance patient understanding, motivation, and adherence, serving as a valuable resource to patients who defer or cannot access in-home safety assessments. Rigorous validation of clinical efficacy and user acceptance is essential to translate this technological potential into improved patient outcomes.

Acknowledgments

We would like to extend our gratitude to Robert Pugliese and MaryEllen Daley for all of their generous support throughout this project. We would also like to thank Dr. Bracken Babula, MD, Dr. Zhe Chen, MD, Dr. Ryan Tomlinson, PhD, Dr. Deanna Gray-Miceli, PhD, CRNP, Dr. Christine Hsieh, MD, and Dr. Brooke Salzman, MD, for their feedback and guidance on this project. No generative AI was used in any portion of the manuscript text generation. We used the generative AI Tool “Gemini 1.5 Pro” made by Google to draft the system prompt found in [Multimedia Appendix 2](#), with review and editing from the study group. Image portions of [Multimedia Appendices 1, 4 and 5](#) were generated using “Gemini 2.0 Flash Image Preview” as described in the manuscript text as part of the model design. Image portions of [Multimedia Appendices 1, 4 and 5](#) were generated using “Gemini 2.5 Flash Image” as described in the manuscript text as part of the model design.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Authors' Contributions

Conceptualization: JD, LZ, BC, JC. Methodology: JD, VS. Resources: JD, LZ, BC, JC. Supervision: RP. Writing – Original draft: JD. Writing – Revising and editing: JD, VS, LZ, BC, JC.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Two-shot image modification architecture.

[\[PNG File \(Portable Network Graphics File\), 942 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

XML Prompt and model parameters.

[\[DOCX File \(Microsoft Word File\), 11 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Blinded comparison methods.

[\[PDF File \(Adobe File\), 94 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Rating packet.

[\[PDF File \(Adobe File\), 15079 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Usage of model for recommendations for post-stroke patients.

[\[PNG File \(Portable Network Graphics File\), 1063 KB-Multimedia Appendix 5\]](#)

References

1. Niedermann K, Meichtry A, Zindel B, et al. Effectiveness and cost-effectiveness of a single home-based fall prevention program: a prospective observational study based on questionnaires and claims data. *BMC Geriatr*. Dec 28, 2024;24(1):1044. [doi: [10.1186/s12877-024-05586-x](https://doi.org/10.1186/s12877-024-05586-x)] [Medline: [39732675](https://pubmed.ncbi.nlm.nih.gov/39732675/)]
2. Azizan A, Justine M. Elders' exercise and behavioral program: effects on balance and fear of falls. *Phys Occup Ther Geriatr*. Oct 2, 2015;33(4):346-362. [doi: [10.3109/02703181.2015.1093060](https://doi.org/10.3109/02703181.2015.1093060)]
3. Gillespie LD, Robertson MC, Gillespie WJ, et al. Interventions for preventing falls in older people living in the community. *Cochrane Database Syst Rev*. Sep 12, 2012;2012(9):CD007146. [doi: [10.1002/14651858.CD007146.pub3](https://doi.org/10.1002/14651858.CD007146.pub3)] [Medline: [22972103](https://pubmed.ncbi.nlm.nih.gov/22972103/)]
4. Lektip C, Chaovalit S, Wattanapisit A, Lapmanee S, Nawarat J, Yaemrattanakul W. Home hazard modification programs for reducing falls in older adults: a systematic review and meta-analysis. *PeerJ*. 2023;11:e15699. [doi: [10.7717/peerj.15699](https://doi.org/10.7717/peerj.15699)] [Medline: [37489124](https://pubmed.ncbi.nlm.nih.gov/37489124/)]
5. Lee JJ, Patel D, Gadgil M, Langness S, von Hippel CD, Sammann A. Understanding barriers to home safety assessment adoption in older adults: qualitative human-centered design study. *JMIR Hum Factors*. Jun 24, 2025;12:e66854. [doi: [10.2196/66854](https://doi.org/10.2196/66854)] [Medline: [40554799](https://pubmed.ncbi.nlm.nih.gov/40554799/)]
6. Azizan A. Exercise and frailty in later life: a systematic review and bibliometric analysis of research themes and scientific collaborations. *IJPS*. 2024;11(1):1. [doi: [10.36922/ijps.3282](https://doi.org/10.36922/ijps.3282)]
7. Romero S, Lee MJ, Simic I, Levy C, Sanford J. Development and validation of a remote home safety protocol. *Disabil Rehabil Assist Technol*. Feb 2018;13(2):166-172. [doi: [10.1080/17483107.2017.1300345](https://doi.org/10.1080/17483107.2017.1300345)] [Medline: [28326967](https://pubmed.ncbi.nlm.nih.gov/28326967/)]

8. Yue X, Ni Y, Zheng T, et al. MMMU: a massive multi-discipline multimodal understanding and reasoning benchmark for expert AGI. arXiv. Preprint posted online on Jun 13, 2024. [doi: [10.48550/arXiv.2311.16502](https://doi.org/10.48550/arXiv.2311.16502)]
9. Regulatory considerations on artificial intelligence for health. World Health Organization; 2023. URL: <https://www.who.int/publications/i/item/9789240078871> [Accessed 2025-11-01]

Abbreviations

AI: artificial intelligence

CDC: Centers for Disease Control and Prevention

HIPAA: Health Insurance Portability and Accountability Act

LLM: large language model

MMMU: Massive Multi-discipline Multimodal Understanding and Reasoning

NICE: National Institute for Health and Care Excellence

OT: occupational therapy

STEADI: Stopping Elderly Accidents, Deaths, & Injuries

XML: Extensible Markup Language

Edited by Jinjiao Wang; peer-reviewed by Azliyana Azizan, Dimitrios Menychtas; submitted 21.May.2025; final revised version received 01.Mar.2026; accepted 16.Mar.2026; published 13.May.2026

Please cite as:

Do J, Suresh V, Zhang L, Chavre BM, Cha J, Pugliese R

Leveraging Multimodal Large Language Models for Fall Risk Reduction in Older Adults in the Home: Proposed Model Design

JMIR Aging 2026;9:e77591

URL: <https://aging.jmir.org/2026/1/e77591>

doi: [10.2196/77591](https://doi.org/10.2196/77591)

© Justin Do, Vivaswat Suresh, Lily Zhang, Bharvi M Chavre, Jeremy Cha, Robert Pugliese. Originally published in JMIR Aging (<https://aging.jmir.org>), 13.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Aging, is properly cited. The complete bibliographic information, a link to the original publication on <https://aging.jmir.org>, as well as this copyright and license information must be included.