

Bridging The Gap: An Audit Of Artificial Intelligence Simplified Dementia Care Resources

Alaa Mohamed¹ · Asmaa Mohamed¹ · Areej Emam¹ · Ayah Gharaibeh¹ · Azza Elkamali¹ · Hadeel Ali, M.B.B.S² · Lujane Sobair¹ · Marwa Mohamed¹ · Marwa Mohamed¹ · Raghad Ghazafi Mohamed¹ · Rim ben Khaled Bessassi¹ · Youmna Ismail¹ · Razan Abdelrahman³ · Nadir Abdelrahman, MD¹

1. Tele–Geriatric Research Fellowship, Okemos, MI, USA · 2. Thumbay University Hospital, Ajman, United Arab Emirates · 3. Michigan State University, East Lansing, MI, USA

3140580
POSTER



KEY TAKEAWAYS

74% Readability Gain

Mean reading grade level fell from 15.9 to 4.2 after AI simplification of post–diagnostic dementia materials

High Clinician–Rated Fidelity

Simplified outputs scored a mean of 4.5/5.0 for clinical utility and accuracy across three independent reviewers

A Compassion Gap Remains

Tone was occasionally rated as overly clinical or detached — reinforcing the need for Human-in-the-Loop oversight

BACKGROUND

Alzheimer’s disease or related dementia (ADRD) diagnoses often involve complex clinical documentation that can overwhelm families.

Large Language Models (LLMs), specific types of Artificial Intelligence (AI), offer a way to simplify medical jargon; however, their accuracy and empathy in a clinical context remain under–evaluated.

WHY IT MATTERS

- Post–diagnostic materials overwhelm patients and caregivers
- Professional clinical text averages a 15.9 grade reading level
- AI accuracy and empathy in clinical contexts remain under–evaluated
- Risk of "hallucination" in high–stakes health information

METHODS

An audit study evaluating 3 LLMs available to the public [Gemini 1.5, GPT–4o, and DeepSeek V3.2] in translating professional post–diagnostic advice sheets into patient–friendly language.

EVALUATION CRITERIA

- Clinical Accuracy** — fidelity to the original clinical meaning
 - Readability** — Flesch–Kincaid Grade Level formula
 - Compassion** — empathetic, supportive tone of simplified text
- All outputs independently rated by three clinicians on a 5–point Likert scale

MODELS EVALUATED

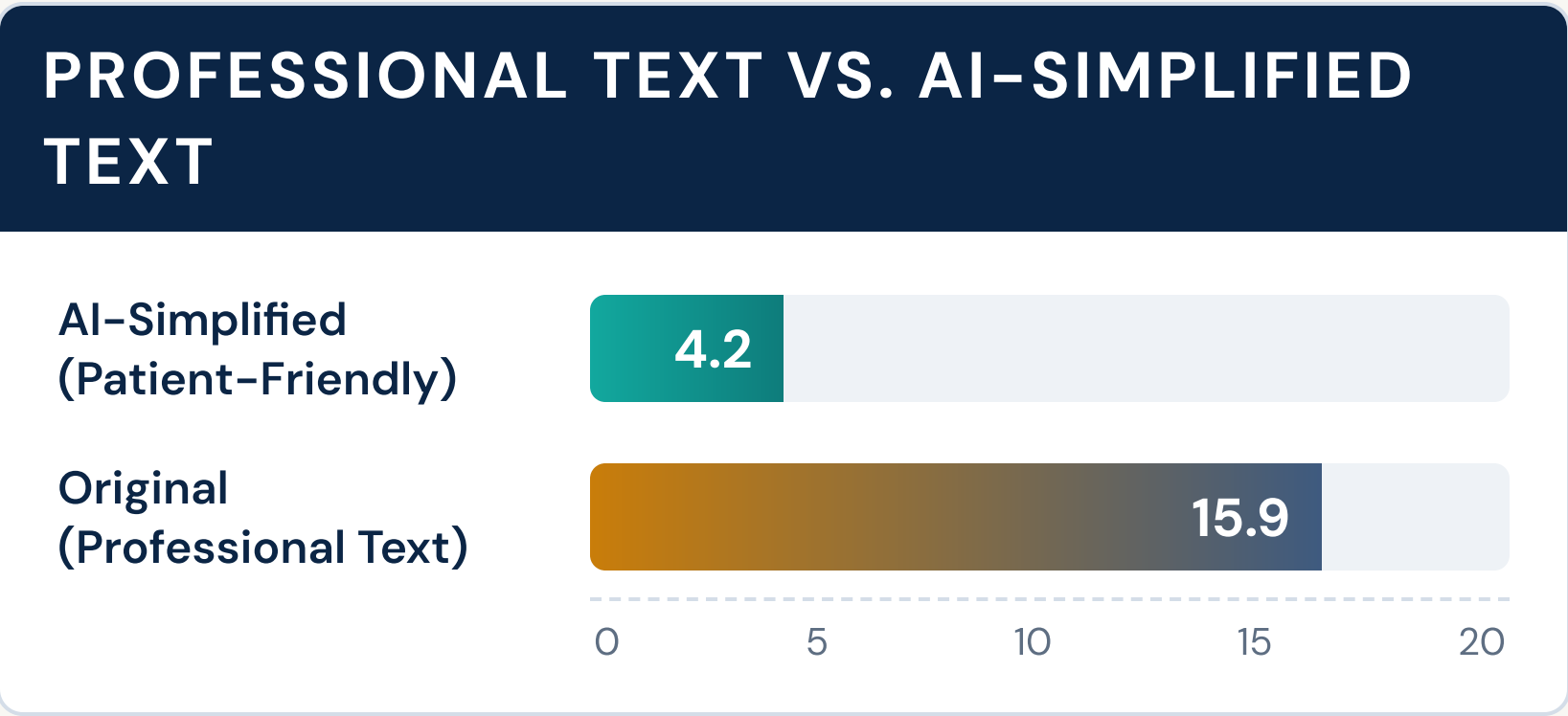
Gemini 1.5 Google multimodal LLM	GPT–4o OpenAI multimodal LLM
DeepSeek V3.2 Open–weight LLM	Fixed Prompt 6th–grade level, accuracy & empathy preserved

TABLE 1 · EVALUATION DOMAINS

Domain	What Was Assessed
Clinical Accuracy	Fidelity to original clinical meaning
Readability	Flesch–Kincaid Grade Level formula
Compassion	Empathetic, supportive tone

Rated by 3 independent clinicians on a 5–point Likert scale

FIGURE 1 · MEAN READING GRADE LEVEL



RESULTS

- 15.9→4.2** Mean Reading Grade: Before → after simplification
- 74%** Readability Improvement: Across 10 information sheets
- 4.5/5.0** Mean Clinician Score: Utility & fidelity rating
- 3** LLMs: Gemini 1.5 · GPT–4o · DeepSeek V3.2

- Mean reading grade level decreased from 15.9 to 4.2, a 74% improvement in readability
- Clinician review demonstrated high utility and fidelity scores (mean 4.5/5.0)
- Qualitative error analysis: infrequent deficits, primarily within the "Compassion" domain
- Tone was occasionally perceived as overly clinical or detached

FUTURE DIRECTIONS

- NEAR–TERM**
Evaluate AI–generated materials directly with patients and caregivers in real–world settings
- MID–TERM**
Develop strategies to improve empathy and compassionate communication in AI outputs; enhance safeguards to reduce hallucinations and improve reliability
- STRATEGIC**
Explore multilingual adaptation and integration of AI tools within Human-in-the-Loop (HITL) clinical workflows

CONCLUSIONS

- LLMs are great tools that can bridge the health literacy gap by transforming complex clinical data into patient–centered materials.
- Integrating AI into post–diagnostic workflows can enhance caregiver support and healthcare equity.
- The potential for hallucinations in high–stakes scenarios supports the use of AI as a clinical "co–pilot" within a Human-in-the-Loop (HITL) framework.

Contact Information

TGRF info

LEARNING OBJECTIVES

- Evaluate AI–integrated protocols for standardizing nonpharmacologic de–escalation and EHR documentation
- Analyze behavioral analytics for reducing psychotropic medication reliance
- Identify accessibility design considerations for diverse nursing workforces

DESIGN PRINCIPLES

- Device–agnostic · No installation required
- Grade 4 reading level target
- Fixed output structure for consistency
- Photo policy: Green — sharing permitted

INFORMATION

- CareResponse / Crisis Card AI
- The Tele–Geriatric Research Fellowship
- <https://aicrisiscards.com/>
- Educational tool · Not diagnostic
- The authors declare no competing financial interests related to this work.

TRY THE MOBILE APP NOW

