

Reliability Analysis of AI-Enhanced Multimodal Assistive Systems for In-Home Elderly Care

Sinan Chen *

*Center of Mathematical and Data Sciences
Graduate School of Engineering
Kobe University
1-1 Rokkodai-cho, Nada-ku, Kobe, Japan
chensinan@gold.kobe-u.ac.jp*

Masahide Nakamura

*Center of Mathematical and Data Sciences
Graduate School of Engineering
Kobe University
1-1 Rokkodai-cho, Nada-ku, Kobe, Japan
masa-n@cs.kobe-u.ac.jp*

Abstract—This paper addresses the reliability challenges of AI-assisted elderly care systems within smart home environments. By integrating multimodal inputs, including automatic speech recognition from instructional videos and optical character recognition from caregiving materials with context-aware processing, we develop a novel question answering system tailored for in-home care scenarios. A sliding context window mechanism prioritizes temporally relevant information, enhancing response accuracy. We propose three reliability metrics, Answer Accuracy Rate (AAR), Context Consistency Rate (CCR), and Modal Agreement Rate (MAR), to systematically assess system performance across four levels of care complexity, where MAR is reported separately as an auxiliary indicator of cross-modal consistency. Experimental results demonstrate optimal reliability at intermediate complexity levels, where procedures are well-structured and queries are clearly articulated. The findings contribute to the development of reliable AI-enhanced assistive systems, bridging advances in electrical engineering, automation technologies, and artificial intelligence for practical healthcare applications.

Index Terms—elderly care AI, assistive technology, system reliability, multimodal integration, context enhancement, smart healthcare

I. INTRODUCTION

The integration of electrical engineering, automation, and artificial intelligence has catalyzed significant advancements in smart home healthcare, particularly in supporting aging populations. AI-enhanced assistive systems, leveraging multimodal sensing and intelligent processing, offer promising solutions for in-home elderly care by providing timely information and decision support [1], [2]. These systems not only alleviate caregiver burden but also promote independent living through automated monitoring and context-aware assistance [3], [4].

However, ensuring the reliability of such systems remains a critical challenge, especially when processing complex, context-dependent queries in dynamic home environments. Traditional approaches often rely on static knowledge bases or generic language models, lacking the contextual awareness necessary for accurate, situation-specific responses in caregiving scenarios [5], [6]. This gap highlights the need for systems that can effectively integrate multimodal data while maintaining contextual coherence throughout extended interactions.

* Corresponding author

Our study addresses this need by developing and evaluating a context-enhanced multimodal question answering system specifically designed for elderly home care applications. By combining real-time speech and text processing with adaptive context management, we aim to create a reliable interface between care recipients and automated assistance systems. This research aligns with the conference’s focus on intelligent system integration, offering insights into the practical implementation of AI and automation technologies in healthcare domains.

II. RELATED WORK

A. Evolution of AI-Based Question Answering Systems in Elderly Care

AI-powered question answering (QA) systems are increasingly used in elderly care to support caregivers and elderly users by providing immediate and reliable health or procedural information. They enhance caregiving experiences and reduce the cognitive load on human caregivers.

Traditional systems often rely on medical knowledge bases or pre-trained language models. While they can answer general health questions, they struggle with procedure-specific or context-sensitive queries related to daily care activities.

A major limitation is their inability to understand the immediate caregiving context. Responses may not match the current instructional content or the specific needs of the situation. To address this, newer systems use large language models and retrieval techniques to provide more relevant answers [7], [8].

B. Applications of Multimodal Data in Assistive AI Systems

Multimodal data such as video subtitles (ASR) from caregiving tutorials, slide content (OCR) from health instruction materials, and visual text enhance AI performance in assistive settings [9]. ASR transcribes spoken instructions, while OCR extracts text from images or slides, enabling systems to understand both audio and visual materials [10], [11].

Fields like Document Visual Question Answering (DocVQA) and VideoQA show how combining text and visual data improves answer accuracy. These models process complex layouts and video content to generate more relevant

responses, which is crucial for accurately interpreting care procedure guides [12], [13].

C. Context Management for Coherent Dialogue Generation in Assistive Systems

This is the “sliding context window” tactic used in typical question-answering systems, which helps frame past conversation history in order to formulate acceptable or context-relevant answers. This approach dynamically recognizes which segments from previous interactions are relevant for generating the subsequent output, ensuring an optimal trade-off between a comprehensive understanding and computational efficiency [14], [15]. The context window in Transformer-based architectures determines the effective receptive field of the model, with significant implications for the flow and coherence of text generation. Nevertheless, restricting the window size results in a loss of important long-range dependencies which are significant in the context of multi-turn conversations about ongoing care [16].

Adding historical question-answer pairs makes the dialogue coherent and consistent. In addition, referencing what was said previously helps the system align with the user, keep track of what is being discussed, and tailor its responses according to the specific care context [17]. However, current models are faced with the task of selectively incorporating useful historical information without bringing any redundancy and contradictions [18]. Therefore, improving the way systems manage conversational context remains a significant area of exploration for assistive AI to behave more naturally and reliably.

D. Contributions and Novelty

While existing assistive QA systems have made progress, many still rely on unimodal inputs and struggle to maintain contextual relevance across diverse home care scenarios [19]. Few systems are validated across multiple levels of care complexity, limiting their generalizability in real-world settings.

This study introduces an integrated framework that combines multimodal inputs, ASR-derived video subtitles from care instructions, OCR-extracted slide text, and historical QA records, to enhance understanding of caregiving content [20]. A sliding context window further improves response coherence by dynamically retrieving recent, relevant content segments [21].

While most studies propose novel architectures for improving QA performance in healthcare, few systematically evaluate the reliability of such systems across different real-world care contexts. This gap motivates our current work.

III. METHODOLOGY

A. System Architecture

1) *Overview*: This study presents a multimodal context-enhanced question-answering system designed to improve AI-assisted elderly home care. Unlike conventional health information systems that rely solely on pre-trained knowledge, this system integrates real-time instructional video subtitles,

OCR-extracted slide text from care procedures, and historical question-answer records to enhance response accuracy and situational relevance.

The system consists of four main components: input module, data processing module, answer generation module, and output module. Each module processes information to ensure that AI responses align with the specific care instruction content.

2) *System Workflow*: The system follows a structured process, as shown in Fig. 1, to retrieve, process, and integrate care instruction content before generating a response. The workflow includes four steps:

Caregiver Query Input: The caregiver or user submits a question through text or speech. If speech is used, automatic speech recognition (ASR) converts it into text.

Context Retrieval and Processing: The system extracts instructional video subtitles using ASR, retrieves slide text using OCR recognition, and accesses historical question-answer records to ensure contextual relevance.

Answer Generation: A large language model (LLM) processes the query along with multimodal inputs. A sliding context window is applied to focus on relevant information, and multimodal sources are merged to ensure the response is accurate and instruction-specific.

Response Output: The system presents a text response to the caregiver or user, with an optional spoken response generated using text-to-speech (TTS).

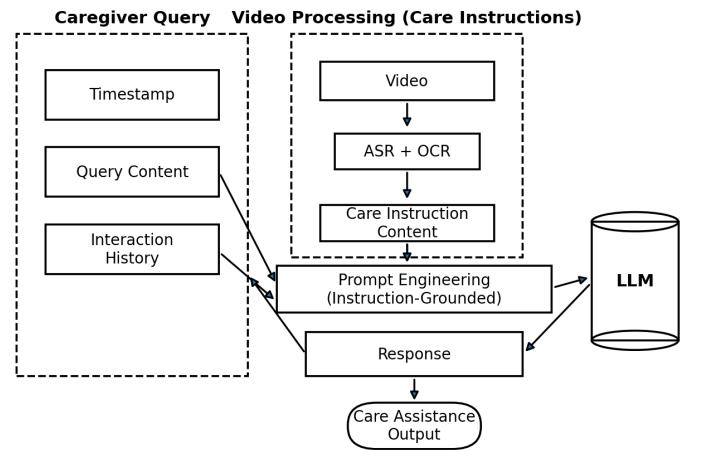


Fig. 1. The proposed system flow chart for in-home elderly care support.

To improve accuracy and coherence, the system integrates multiple key techniques. Table I summarizes the main methods used to enhance AI-generated responses.

B. ASR and OCR Model Configuration

For automatic speech recognition (ASR), the system adopts a pre-trained Transformer-based ASR model operating in an offline inference mode to reduce privacy risks in home care environments. Similarly, OCR processing is performed locally to avoid transmitting raw audio or image data outside the system. According to the official benchmark reported by the

TABLE I
METHODS FOR ENHANCING AI-GENERATED RESPONSES IN HOME CARE
CONTEXT.

| Method | Description |
|------------------------|--|
| Multimodal Integration | The system combines video subtitles from care tutorials, OCR-extracted text from procedure slides, and historical Q&A to provide a complete understanding of the caregiving context. |
| Sliding Context Window | A time-based window selects the most relevant n seconds of instructional content to maintain coherence with the current care activity. |
| Prompt Optimization | The system applies hierarchical prompting to ensure AI-generated answers are driven by the provided care instructions rather than generic health knowledge. |

TABLE II
SUMMARY OF MULTIMODAL EXTRACTION MODELS USED IN THE SYSTEM

| Modality | Model Type | Accuracy and Role |
|----------|---|--|
| ASR | Transformer-based speech-to-text model | WER \approx 8–12% (clean speech); used to extract instructional subtitles from care videos |
| OCR | Deep learning-based document OCR engine | Character accuracy $>$ 95% (printed slides); used to extract textual content from care instruction materials |

model provider, the word error rate (WER) on clean instructional speech is approximately 8–12%, while performance may degrade moderately under background noise conditions.

For optical character recognition (OCR), a deep learning-based OCR engine designed for document and slide text extraction is employed. The model achieves over 95% character-level accuracy on high-resolution printed materials and approximately 90% accuracy on slide images captured from instructional videos. These accuracy characteristics indicate that the noise level introduced by multimodal input extraction remains limited and does not dominate downstream QA performance.

The extracted ASR transcripts and OCR texts are treated as probabilistic inputs rather than ground-truth knowledge, and subsequent context filtering and prompt prioritization are applied to mitigate the impact of transcription errors.

The ASR and OCR configurations adopted in the proposed system are summarized in Table II.

C. Context Enhancement

1) *Sliding Context Window*: The system applies a sliding context window to capture relevant instructional content from recent n seconds of video subtitles. This helps AI focus on the most recent part of the care tutorial instead of generating responses based only on general medical knowledge. The system also retrieves historical Q&A records to improve coherence. If a user’s question is similar to a previous one, the AI can refer to past responses to avoid contradictions.

To manage context efficiently, the system dynamically adjusts the window size based on the complexity of the question.

For factual questions about a step, a shorter window may be enough, while conceptual or multi-step questions about a procedure may require a longer context [22].

2) *Historical Q&A Retrieval Strategy*: To retrieve historical question–answer (Q&A) records, the system adopts a lightweight semantic retrieval strategy based on sentence-level embeddings. Each past question is encoded into a fixed-length vector representation using a pre-trained embedding model. When a new query is submitted, cosine similarity is computed between the query embedding and stored historical embeddings.

Only the top- k most semantically similar Q&A pairs are retrieved, where k is empirically set to a small value to avoid context overload. This semantic retrieval mechanism enables the system to capture paraphrased or implicitly related questions beyond simple keyword matching.

The retrieval scope is restricted to the current session and the same care complexity level, ensuring both contextual relevance and privacy. Retrieved historical answers are not directly reused; instead, they are provided to the language model as auxiliary context to improve response consistency and reduce contradictions.

3) *Multimodal Integration*: The system integrates video subtitles from ASR and slide text extracted using OCR to improve the relevance of AI responses to care procedures. These inputs allow AI to access both spoken instructions and visual content, reducing reliance on pre-trained knowledge. When processing a question, the system assigns different weights to each modality. If the question refers to something just mentioned in the video, ASR-extracted subtitles are prioritized. If it involves a term from a procedure slide, the system focuses more on OCR-recognized text.

To prevent conflicting information, the system applies content filtering before merging different sources. If the same term appears in both the subtitles and slides but with slight differences, the AI gives more priority to the clearer or more structured version [23], [24].

4) *Prompt Engineering*: The system applies hierarchical prompting to ensure AI-generated responses prioritize the provided care instruction materials. The prompt structure first instructs the AI to reference video subtitles and OCR-extracted text before using general knowledge. If no relevant information is found in the extracted content, the AI is allowed to generate a response based on its pre-trained data.

D. Evaluation Methodology

To evaluate the reliability of the proposed home care QA system, we designed a structured human evaluation framework that focuses on three core reliability dimensions.

It is important to note that the Modal Agreement Rate (MAR) measures cross-modal semantic agreement between ASR- and OCR-derived content, rather than the recognition accuracy of individual extraction models.

Answer Accuracy Rate (AAR): Measures the percentage of responses that correctly reflect the care instruction content without factual errors.

TABLE III
SURVEY QUESTIONNAIRE FOR AI RESPONSE EVALUATION IN HOME CARE
CONTEXT.

| ID | Question Item | Metric |
|-----|--|--------|
| Q1 | Reflects the care instruction content. | AAR |
| Q2 | Avoids factual errors in care procedures. | AAR |
| Q3 | Provides sufficient detail for the query. | AAR |
| Q4 | Uses recently presented instruction content. | CCR |
| Q5 | Matches information in slides or subtitles. | MAR |
| Q6 | Avoids irrelevant or unrelated details. | CCR |
| Q7 | Remains consistent with prior answers in the same session. | CCR |
| Q8 | Produces similar answers for related questions. | CCR |
| Q9 | Is clear and well structured. | AAR |
| Q10 | Is more accurate than the baseline model. | AAR |
| Q11 | Is more relevant than the baseline model. | CCR |
| Q12 | Avoids over-generalization compared to the baseline. | MAR |

Context Consistency Rate (CCR): Assesses whether similar or repeated questions receive logically consistent answers within the same care scenario.

Modal Agreement Rate (MAR): Evaluates the consistency between information extracted from different input sources (e.g., ASR from video, OCR from slides).

These reliability metrics were derived through human rating of AI-generated responses using a 12-item structured questionnaire. We used an AI system without any prompts as the baseline model for comparison, where the baseline disables instruction-grounded prompting, historical content retrieval, and sliding context window mechanisms. Each response was rated by evaluators (simulating caregivers or trained assistants) on a 5-point Likert scale, with higher scores indicating better reliability. The evaluation covered four levels of care complexity, Basic Daily Care, Intermediate Home Care, Advanced Health Monitoring, and Complex Medical Care. Evaluators for the basic level were supported by caregiving staff. The system’s answers were rated based on the following grouped dimensions (Table III):

E. System Control and Fault Tolerance Integration

From a system design perspective, the proposed context-enhanced question-answering framework can be viewed as a closed-loop assistive system, where multimodal sensing serves as input acquisition, context reasoning functions as decision logic, and response generation acts as output delivery. Future work may explore the integration of fault-tolerant control concepts and adaptive monitoring mechanisms to improve system robustness under sensor noise, input inconsistencies, or network delays in real-world home care environments.

IV. EXPERIMENTAL EVALUATION

A. Experimental Setup

To evaluate the system’s performance, we selected four instructional videos, each corresponding to a different level of elderly home care complexity: basic daily care (e.g., mobility assistance), intermediate home care (e.g., medication

management), advanced health monitoring (e.g., vital signs), and complex medical care procedures (e.g., post-operative care guidance). These videos cover a range of topics relevant to home-based support.

For each video, we recruited five evaluators familiar with caregiving at the respective complexity level. Each evaluator was instructed to watch the instructional video and pose one question at any point during the video, simulating natural inquiries from a caregiver or an informed elderly user. This method ensured a natural distribution of questions based on individual understanding and situational needs.

The system used DeepSeek R1 to generate answers, integrating video subtitles, OCR-extracted slide text, and historical Q&A. The large language model component can be deployed in a local or private computing environment, or operated on de-identified textual inputs, ensuring that no raw audiovisual data are exposed during inference.

B. Instructional Data Sources

The instructional videos and nursing-related materials used in the experiment were self-collected from publicly available educational resources and institutional training materials for caregiving education. No private patient data were included. All materials were used under appropriate educational or research permissions and contain no personal or identifiable data.

In total, four instructional videos were selected, each representing a distinct level of care complexity. The videos range from 6 to 12 minutes in duration and cover approximately 15–25 key procedural statements per video after ASR transcription and OCR extraction.

The nursing topics include mobility assistance, medication management, basic health monitoring, and post-operative home care guidance. Although the dataset size is limited, it was intentionally designed to reflect realistic caregiver learning scenarios and to evaluate system reliability under controlled but representative conditions [25].

The goal of the dataset construction is not large-scale generalization, but reliability comparison across different care complexity levels under consistent multimodal input settings.

C. Example Responses

To evaluate the system’s ability to generate instruction-based responses, we used a structured prompt format that ensures the AI references the provided care content. Below is an example of how the model processes a question from a caregiver based on a basic care video about fall prevention.

1) *Prompt Format Sent to LLM:* The system sends a structured message to the language model, ensuring it focuses on relevant care instruction content while generating responses.

```
{
  "role": "system",
  "content": "You are a virtual care assistant specialized in answering questions based on provided elderly care instruction videos. Use the provided slide content, video explanations, and key safety points to ensure accuracy. Prioritize information from the care instruction content before using general medical knowledge. If the question is outside the scope of the
```

```

provided content, acknowledge that and provide relevant
guidance."
}

```

Additionally, the system includes control instructions to enhance response consistency.

Reference Priority: The AI must use care procedure slides, video script, and key safety concepts first.

Context Awareness: If the question relates to a recent instruction section, the AI should prioritize nearby content.

Clarity & Simplicity: Responses should be clear, reassuring, and easy to understand for caregivers.

2) *Instructional Content Provided to the AI:*

```

{
"title": "Home Safety and Fall Prevention",
"content": [
"A cluttered floor is a major tripping hazard for the
elderly.",
"Ensure good lighting, especially in hallways and bathrooms
.",
"Install grab bars in the bathroom near the toilet and
shower.",
"Encourage the use of non-slip footwear indoors."
],
"image": "Diagram of a safe bathroom layout",
"slide": "Key Home Safety Tips",
"script": "Let's talk about making your home safer. First,
look at the floor.@ Is it clear of cords, rugs, or
clutter? These are trip hazards. Next, lighting. Is the
path to the bathroom well-lit at night?@"
}

```

3) *Question from User (Simulated Caregiver):* **Timestamp:** The locations marked with "@" in "script". **User's Question:** "If good lighting is so important, what type of bulb is best for a nighttime hallway light?"

4) *AI Response (Context-Enhanced):*

```

{
"scenario": "basic_care",
"question": "If good lighting is so important, what type of
bulb is best for a nighttime hallway light?",
"response": "That's an important practical question. Based
on the current safety instructions, we know that
ensuring good lighting, especially in hallways, is a
key point for preventing falls. However, the specific
type of bulb (like LED, wattage, color temperature) is
not discussed in this particular video segment. For
nighttime hallway lighting, general safety knowledge
suggests using nightlights or low-wattage, warm-
temperature LEDs that provide sufficient illumination
without being too harsh or causing glare when waking up
. It's also a good idea to place them so they light the
path without creating shadows. You may want to consult
a home safety checklist for more detailed product
recommendations."
}

```

This response follows the structured approach by: - Referencing the provided instruction content first (confirming the importance of good lighting). - Acknowledging missing information (noting that bulb type wasn't specified). - Providing additional practical insight based on general safety principles.

5) *Baseline Response (Without Context):* For the same question, the response from a baseline model without context enhancement is as follows:

```

{
"question": "If good lighting is so important, what type of
bulb is best for a nighttime hallway light?",
"response": "For nighttime hallway lighting, warm white LED
bulbs are often recommended. They provide enough light
to see clearly without being too bright or disrupting
sleep cycles. A bulb with a color temperature around
2700K is suitable. Also consider using bulbs with lower
lumens or dedicated plug-in nightlights."
}

```

D. Performance Analysis

We evaluated the reliability of the proposed QA system based on 20 user-generated questions across four levels of care complexity: Basic Daily Care, Intermediate Home Care, Advanced Health Monitoring, and Complex Medical Care. The results focus on two core dimensions, Answer Accuracy Rate (AAR) and Context Consistency Rate (CCR), while Modal Agreement Rate (MAR) is reported separately due to its consistently high performance across all conditions.

1) *Overall Reliability Metrics:* Table IV summarizes the reliability metrics of the context-enhanced system across different care complexity levels.

TABLE IV
AAR AND CCR ACROSS LEVELS OF CARE COMPLEXITY.

| Metric | Basic Care | Inter. Care | Adv. Monitor. | Complex Medical | Avg. |
|--------|------------|-------------|---------------|-----------------|-------|
| AAR | 85.0% | 92.0% | 88.0% | 81.0% | 86.5% |
| CCR | 82.0% | 91.0% | 84.0% | 76.0% | 83.3% |

The intermediate home care level achieved the highest scores in both AAR and CCR, indicating that the system performs best when care procedures are well-structured and user queries are clearly formulated. In contrast, the complex medical care level performance was the lowest, with an AAR of 81.0% and a CCR of 76.0%.

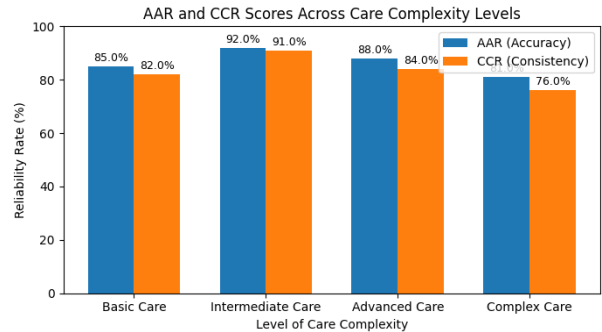


Fig. 2. AAR and CCR Scores Across Levels of Care Complexity.

Fig. 2 visually contrasts the system's reliability across care complexity levels. A clear gap emerges between intermediate and complex medical care levels in both AAR and CCR, highlighting how performance varies with the abstraction and specificity of care content. The convergence of AAR and CCR values in the intermediate care segment suggests a stable interaction pattern at this level, while the divergence seen in the complex medical group reveals challenges in maintaining contextual coherence under sophisticated medical discourse.

2) *Modal Agreement Rate (MAR):* The system demonstrated high modal agreement across all levels, with average MAR exceeding 95%. Most subtitle text produced via ASR was clear and accurately aligned with slide content extracted via OCR. This result indicates that current multimodal extraction techniques are sufficiently robust for care instruction

materials and did not significantly impact system reliability. Therefore, MAR is not included in the comparative analysis with the baseline system.

3) *Comparison with Baseline System:* To measure the effect of contextual and multimodal enhancements, we compared our system with a baseline version lacking context windowing and content retrieval. Table V presents the improvement in AAR and CCR.

TABLE V
IMPROVEMENT OVER BASELINE SYSTEM.

| Care Complexity Level | AAR Gain | CCR Gain |
|----------------------------|----------|----------|
| Basic Daily Care | +6.0% | +5.0% |
| Intermediate Home Care | +9.0% | +8.0% |
| Advanced Health Monitoring | +5.0% | +4.0% |
| Complex Medical Care | +4.0% | +3.0% |

The largest gains were observed in the intermediate home care group, which benefited the most from the system’s enhancements. Gains were more modest at the complex medical care level, again suggesting that higher-level reasoning and abstract medical questioning reduce the relative advantage of contextual integration based on fixed video segments.

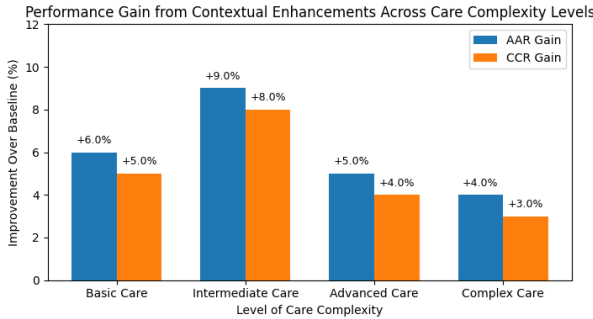


Fig. 3. Performance Gain (AAR & CCR) from Contextual Enhancements Across Care Levels.

Fig. 3 illustrates how the benefits of contextual integration are not uniformly distributed. While all groups showed improvements, the magnitude of gain decreases progressively from intermediate to complex medical care. This suggests diminishing returns of short-window context strategies as content shifts from structured procedural knowledge to open-ended conceptual or medical reasoning, calling for more adaptive context modeling techniques for advanced care support.

V. DISCUSSION

The experimental results confirm that the proposed context-enhanced QA system for elderly home care demonstrates improved reliability compared to a non-contextual baseline. Notably, the most substantial gains in both answer accuracy and contextual consistency were observed at the intermediate home care level. This indicates that the system is particularly effective when dealing with moderately complex care procedures and clearly formulated queries. In contrast, the

complex medical care level responses showed the lowest reliability, suggesting that the system struggles with higher-level discourse involving abstract medical reasoning and more ambiguous question phrasing.

Context consistency (CCR) exhibited a declining trend with increasing care complexity. This pattern highlights a critical limitation of the current system design: while the sliding context window improves local relevance, it may not fully capture long-range dependencies or conceptual continuity required for complex medical discussions or multi-stage care procedures. As query sophistication increases, even subtle gaps in contextual linkage can reduce response coherence.

Another limitation stems from the fixed-length sliding context window, which restricts the scope of available instructional material during answer generation. When users refer to concepts or precautions discussed much earlier in the session, the system may lack access to that information, resulting in partial or incomplete responses. Additionally, OCR-extracted content often emphasizes the most recent slide, potentially omitting foundational context that could improve answer accuracy for cumulative topics.

These findings suggest several directions for future work aligned with intelligent systems development. Dynamic context windowing strategies, based on semantic salience rather than strict temporal windows, could help maintain coherence across longer care instruction sessions. Incorporating user-specific interaction history or adaptive memory mechanisms may also enhance consistency, particularly in complex care settings where medical discourse is nuanced and user phrasing is varied. Integration with smart home sensor data (IoT) could provide an additional, real-time modality to ground responses in the immediate physical context of the elderly user, further improving relevance and reliability.

A. Limitation of Sliding Context Windows and Long-Term Memory

A key limitation of the current system lies in its reliance on a fixed-length sliding context window. When users ask referential questions such as “What are the side effects of the antihypertensive drug mentioned earlier?”, relevant information may fall outside the active window if the discussion occurred significantly earlier in the session.

In such cases, the system prioritizes acknowledging missing contextual information rather than hallucinating unsupported answers. While this design choice improves safety, it also reveals the inherent trade-off between computational efficiency and long-term contextual coverage.

This limitation motivates future integration of long-term memory mechanisms, such as persistent vector-based memory stores or domain knowledge graphs. By linking entities (e.g., medications, procedures, risks) across temporal boundaries, the system could support referential queries beyond the immediate context window while maintaining consistency and traceability.

Incorporating such memory structures would transform the system from a purely window-based conversational model

into a hybrid architecture combining short-term contextual reasoning with long-term semantic memory.

B. Automation and Control System Perspective

The proposed context-enhanced question-answering framework can be viewed as a closed-loop control system: multimodal sensing serves as input acquisition, context reasoning functions as control logic, and response generation acts as actuator output. Future work may explore reinforcement learning-based adaptive controllers and nonlinear system models to dynamically adjust context window parameters based on system state and input signal variation. This aligns with the conference topics in *nonlinear control theory, complex systems management, and intelligent automation*.

VI. CONCLUSION

This study evaluated the reliability of a context-enhanced QA system for AI-assisted elderly home care using a structured, multi-metric framework. By integrating video subtitles from care instructions, OCR-extracted slide text, and historical QA records, the system enhances both answer accuracy and contextual consistency. Among the four levels of care complexity tested, the intermediate home care group achieved the highest reliability scores, indicating that the system is particularly effective when procedural complexity is moderate and user queries are well-structured. In contrast, performance was lowest at the complex medical care level, where abstract concepts and varied question phrasing posed greater challenges for contextual coherence.

These results reinforce the importance of designing assistive AI systems that are not only capable of accessing multimodal inputs but also able to manage context dynamically to suit the specific domain. For the field of electrical engineering, automation, and AI, this work demonstrates a concrete application of multimodal integration and intelligent control strategies in the critical domain of smart health and elderly care.

Future research should explore adaptive context windowing based on semantic importance, as well as personalized memory mechanisms and integration with IoT-based ambient data, to support reliable and trustworthy AI assistance in complex home care and health support scenarios.

ACKNOWLEDGMENT

This research was partially supported by JSPS KAKENHI Grant Numbers JP25H01167, JP25K02946, JP25K24389, JP24K02765, JP24K02774, JP23K17006, JP23K28091, JP23K28383, and JST SICORP Grant Number JPMJKB2312.

REFERENCES

- [1] J. Zhang and W. Leong, "The personalised emotion-driven ai model for elderly care: Optimising emotion recognition and adaptive response mechanisms," *ASM Science Journal*, vol. 20, no. 1, pp. 1–7, 2025.
- [2] H. Abdollahi and M. Mahoor, "Artificial emotional intelligence in socially assistive robots for older adults: A pilot study," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 2098–2109, 2022.
- [3] R. Pita and M. Fantinato, "An integrated social robot and virtual assistant solution to support medical management for older adults," *Expert Systems*, vol. 42, no. 3, p. e70112, 2025.
- [4] O. Brown, R. Davison, and J. Gore, "Theory-driven perspectives on generative artificial intelligence in business and management," *British Journal of Management*, vol. 35, no. 1, pp. 1–24, 2024.
- [5] R. Sharma and K. Mehta, "Emotion recognition from text and voice: A multimodal ai approach to understanding human feelings," *International Journal for Research in Applied Science and Engineering Technology*, vol. 13, no. 7, pp. 73 790–73 797, 2025.
- [6] P. Goyal and S. Gupta, "Iot based stress detection using cognitive assistance for the elderly," *Journal for ReAttach Therapy and Developmental Diversities*, vol. 8, no. 1, pp. 108–117, 2025.
- [7] N. Veigas and T. Navya, "Emotion recognition of elderly people using deep learning," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 13, no. 4, pp. 134–202, 2024.
- [8] L. Wang and C. Ye, "A study on interaction optimization of home service robots based on generative ai: Focusing on the aging population," *Journal of Humanities and Arts Perspectives*, vol. 1, no. 1, pp. 1–10, 2025.
- [9] Z. Yang *et al.*, "Cc-ocr: A comprehensive and challenging ocr benchmark for evaluating large vision-language models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025, pp. 1–13.
- [10] X. Yu and L. Ge, "Multimodal sensing-enabled large language models for automated emotional regulation: A review of current technologies, opportunities, and challenges," *Sensors*, vol. 25, no. 15, p. 4763, 2025.
- [11] D. Chen, J. Chen, Z. Shao, and M. Gong, "Rcvqa: Visual question answering model based on reading comprehension," *Neural Networks*, vol. 184, p. 108365, 2025.
- [12] Y. Liu and Z. Wang, "Investigating elderly individuals' acceptance of artificial intelligence (ai)-powered companion robots: The influence of individual characteristics," *Behavioral Sciences*, vol. 15, no. 5, p. 697, 2025.
- [13] B. S. Kim *et al.*, "Visual question answering: A survey of methods, datasets and challenges," *ACM Computing Surveys*, vol. 58, no. 4, pp. 1–40, 2025.
- [14] S. Khaksar and D. Saplacan, "Robotics in elderly healthcare: A qualitative analysis of 20 recent european research projects," *ACM Transactions on Human-Robot Interaction*, vol. 12, no. 4, pp. 1–28, 2023.
- [15] S. Chen and M. Nakamura, "Generating personalized dialogues based on conversation log summarization and sentiment analysis," in *Proceedings of the 23rd International Conference on Information Integration and Web Intelligence (iiWAS 2021)*, 2021, pp. 217–222.
- [16] Z. Feng and J. Weng, "Artificial intelligence in chronic disease management for aging populations: A systematic review of machine learning and nlp applications," *International Journal of General Medicine*, vol. 18, pp. 1–28, 2025.
- [17] T. Laban and M. Spitale, "Critical insights about robots for mental wellbeing," *arXiv*, vol. abs/2506.13739, 2025.
- [18] S. Chen and M. Nakamura, "Developing a facial identification system using pre-trained model and spoken dialogue agent," in *2022 International Balkan Conference on Communications and Networking (BalkanCom)*. IEEE, 2022, pp. 62–67.
- [19] C. Liu and J. Cao, "From screens to scenes: A survey of embodied ai in healthcare," *Information Fusion*, vol. 91, p. 103033, 2025.
- [20] H. Qiu and Y. Chen, "Embodied artificial intelligence in ophthalmology," *npj Digital Medicine*, vol. 8, no. 1, p. 54, 2025.
- [21] P. Ihämäki and K. Heljakka, "Robot pets as "serious toys" – activating social and emotional experiences of elderly people," *Information Systems Frontiers*, vol. 23, no. 5, pp. 1267–1287, 2021.
- [22] J. Zhang *et al.*, "Hierarchical modeling for medical visual question answering with cross-attention fusion," *Applied Sciences*, vol. 15, no. 9, p. 4712, 2025.
- [23] N. H. Nguyen *et al.*, "Openvivqa: Task, dataset, and multimodal fusion models for visual question answering," *Information Fusion*, vol. 101, p. 101868, 2023.
- [24] Q. Chen, M. Bian, and H. Xu, "Mmq: Multi-question learning for medical visual question answering," in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2024*, ser. Lecture Notes in Computer Science, vol. 15005. Springer Nature Switzerland, 2024, pp. 480–489.
- [25] Anonymous, "3d-rad: A comprehensive 3d radiology medical visual question answering benchmark," *arXiv preprint*, vol. arXiv:2506.11147, 2025.