International Journal of Science, Strategic Management and Technology
Volume 02 Issue 08 August-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Nani AGI: A Personalized Multimodal AI Architecture

Author Details:

Devireddy Lohith Reddy[1]

[1] Department of Electrical and Electronics Engineering/ Vidya Vikas Institute of Engineering and Technology, Mysuru, Karnataka, India

Corresponding Author Email: lohithredde05@gmail.com

https://doi.org/10.55041/ijsmt.v2i8.032

Cite this Article: Reddy, D. L. (2026). Nani AGI: A Personalized Multimodal AI Architecture. International Journal of Science, Strategic Management and Technology, *02*(8), 1-9. https://doi.org/10.55041/ijsmt.v2i8.032

Abstract—
Nani AGI is a personalized, multimodal, and proactive artificial intelligence architecture designed to support natural and continuous human–AI interaction beyond conventional question-and-answer systems. The proposed architecture integrates large language model-based conversational reasoning with persistent memory, voice interaction, emotion-aware processing, multimodal perception, personalization, task planning, and authorized computer-level automation within a unified framework. The methodology follows a modular system-development approach in which perception, cognition, memory, decision-making, and action are organized as interoperable layers. The proposed implementation uses Python-based speech processing, language identification, speech synthesis, large language model services or compatible inference frameworks, and operating-system automation interfaces. The evaluation framework defines task completion, contextual continuity, personalization, intent recognition, response latency, and interaction effectiveness as principal measures, with comparative and ablation-based testing proposed for future empirical validation. The architectural analysis provides a structured foundation for more adaptive and user-centric AI assistants. The work contributes an AGI-inspired personal AI architecture and identifies future research directions in autonomous planning, multimodal reasoning, long-term memory, privacy, and real-world system integration.

Keywords— Artificial Intelligence; Multimodal AI; Personalized AI; Human–AI Interaction; Persistent Memory; Proactive AI

I. INTRODUCTION

The rapid advancement of artificial intelligence (AI), particularly large language models (LLMs), has significantly transformed human interaction with computational systems. Modern LLMs have demonstrated capabilities in language understanding, reasoning, and multimodal processing, creating opportunities for more capable and flexible AI assistants [1], [2]. However, conventional conversational systems often remain centred on individual interactions and may lack persistent understanding of a user's long-term preferences, previous experiences, and ongoing objectives.

Recent research has demonstrated the potential of extending language models beyond passive response generation toward systems capable of reasoning, planning, and interacting with external environments. Yao *et al.* introduced ReAct, a framework that interleaves reasoning and task-specific actions, enabling language models to interact with external sources and environments while updating their action plans [3]. This demonstrates the importance of connecting language-model reasoning with practical action execution.

Long-term memory represents another important requirement for persistent AI systems. Packer *et al.* proposed MemGPT, which uses virtual context management to extend the effective context available to language models and supports long-term interactions [4]. Park *et al.* introduced Generative Agents, combining observations, memory storage, reflection, retrieval, and planning to produce persistent agent behavior [5]. Their ablation study further demonstrated that observation, planning, and reflection each contribute to agent behavior [5].

Multimodal interaction has also become an important direction in AI research. OpenAI's GPT-4 Technical Report describes GPT-4 as a multimodal model capable of accepting both image and text inputs and producing text outputs [2]. Similarly, research on empathetic dialogue has demonstrated the importance of recognizing emotional situations and generating responses perceived as more empathetic by human evaluators [6].

Despite these advances, a research gap remains in the integration of **persistent personal memory, multimodal interaction, emotional context, reasoning, planning, personalization, and computer-level task execution within a single user-centric architecture**. Existing research has demonstrated these capabilities individually or within specialized agent frameworks, but their coordinated application toward a continuously personalized personal AI assistant remains an important area for further investigation [3]–[6].

To address this gap, this study proposes **Nani AGI**, an AGI-inspired personalized multimodal cognitive architecture designed to integrate perception, reasoning, persistent memory, personalization, emotion-aware interaction, planning, and action capabilities. The primary objective is to develop and evaluate a unified framework capable of supporting continuous and adaptive human–AI interaction.

The study therefore investigates the following research questions:

1. **RQ1:** How can multimodal perception, persistent memory, reasoning, and task execution be integrated into a unified personalized AI architecture?
2. **RQ2:** To what extent does persistent contextual memory improve continuity and personalization in human–AI interaction?
3. **RQ3:** How effectively can an integrated AI architecture perform user-oriented tasks through planning and computer-level actions?
4. **RQ4:** What are the performance, latency, and practical limitations of such a personalized multimodal AI system?

II. LITERATURE REVIEW

The development of intelligent conversational systems has progressed from task-specific dialogue systems toward LLM-based architectures capable of reasoning, memory management, planning, and interaction with external environments. This evolution provides the foundation for Nani AGI, which seeks to integrate these capabilities into a personalized multimodal architecture.

A. LLM-Based Reasoning and Autonomous Action

Yao *et al.* introduced ReAct, a framework that combines reasoning traces with task-specific actions in an interleaved process [3]. The approach enables language models to interact with external sources and environments, allowing observations from those environments to influence subsequent reasoning and actions. The authors evaluated ReAct across question answering, fact verification, and interactive decision-making tasks and reported improvements over relevant baselines [3]. However, ReAct primarily addresses reasoning and action coordination rather than persistent personalization or long-term user relationships.

Nani AGI builds upon this reasoning–action paradigm by incorporating task planning and computer-level interaction alongside persistent memory, personalization, and multimodal perception.

B. Memory and Long-Term Interaction

Packer *et al.* proposed MemGPT to address the limitations imposed by finite context windows in LLMs [4]. The system applies virtual context management inspired by operating-system memory hierarchies and enables information to be managed across different memory tiers. The authors evaluated the approach in document analysis and multi-session conversational settings, demonstrating its usefulness for extended interactions [4].

Park *et al.* introduced Generative Agents, an architecture that maintains records of experiences, synthesizes higher-level reflections, retrieves relevant memories, and uses them for planning future behavior [5]. Their evaluation showed that these mechanisms contributed to believable individual and emergent social behaviour, while ablation experiments demonstrated the importance of observation, planning, and reflection [5].

These studies establish the importance of memory and retrieval for persistent AI agents. Nani AGI extends this direction toward a personal-assistant context by incorporating user preferences, conversational history, task-related information, and contextual personalization into a unified architecture.

C. Emotion-Aware and Human-Centred Dialogue

Rashkin *et al.* introduced EmpatheticDialogues, a dataset and benchmark for empathetic open-domain conversation [6]. The study focused on the ability of dialogue agents to recognize feelings expressed within conversations and respond appropriately. Their experiments showed that models adapted using emotionally grounded conversational data were perceived as more empathetic than models trained only on large-scale Internet conversations [6].

More recent multimodal emotion-understanding research has further explored the ability of multimodal large language models to interpret complex emotional information. Lian *et al.* introduced Affect GPT and a corresponding benchmark for multimodal emotion understanding, demonstrating the growing relevance of multimodal approaches to emotion recognition [7].

These studies support the importance of emotional context in human–AI interaction. However, emotion-aware interaction is generally investigated as a specialized capability rather than as one component of a broader architecture combining persistent personal memory, multimodal perception, reasoning, planning, and computer-level action. Nani AGI incorporates emotion awareness within such a broader personalized framework.

D. Multimodal Artificial Intelligence

OpenAI's GPT-4 Technical Report describes a large-scale multimodal model capable of accepting image and text inputs and generating text outputs [2]. The work demonstrates the potential of multimodal models to extend AI interaction beyond purely textual inputs.

Multimodal emotion research further demonstrates how language models can combine visual and linguistic information for more comprehensive interpretation of human emotional states [7]. These developments provide an important foundation for personal AI systems that may need to process speech, text, images, and other contextual information.

However, multimodal capability alone does not provide persistent personalization, long-term memory, or autonomous task management. Nani AGI therefore treats multimodality as one component of a larger cognitive architecture.

E. Autonomous-Agent Architectures

Wang *et al.* provide a comprehensive survey of LLM-based autonomous agents, examining agent construction frameworks, applications, evaluation strategies, and challenges [8]. Their survey demonstrates that modern autonomous agents commonly combine components such as profiling, memory, planning, and action within broader agent architectures [8].

This literature establishes a strong foundation for agentic AI but also highlights continuing challenges surrounding reliable planning, evaluation, memory, and interaction. Nani AGI builds upon these architectural directions while focusing specifically on a **personalized, multimodal, continuously interacting AI companion** rather than a general autonomous-agent framework.

F. Research Gap and Contribution of Nani AGI

The reviewed literature demonstrates substantial progress across the individual capabilities required for advanced AI agents. ReAct establishes the relationship between reasoning and action [3]; MemGPT addresses extended context and memory management [4]; Generative Agents demonstrate the value of memory, reflection, observation, and planning [5]; empathetic dialogue research establishes the importance of emotional context [6]; multimodal LLM research demonstrates the integration of different information modalities [2], [7]; and autonomous-agent research provides broader architectural and evaluation perspectives [8].

Nevertheless, a gap remains in developing a **unified, user-centric architecture that combines conversational intelligence, persistent personal memory, emotional awareness, multimodal perception, planning, personalization, and real-world computer interaction**.

Nani AGI addresses this gap through the architectural integration of these complementary capabilities within a modular personal AI framework. The primary contribution is therefore not the invention of each individual capability, but their **systematic integration, personalization, and planned experimental evaluation within a unified human–AI interaction architecture**.

III. METHODOLOGY

The methodology adopted for Nani AGI follows a modular system-development and experimental evaluation approach. The proposed architecture is designed to integrate conversational intelligence, speech processing, persistent memory, multimodal perception, emotion-aware interaction, task planning, and computer-level action execution within a unified framework. Each component is implemented as an independent module, allowing individual capabilities to be developed, evaluated, and subsequently integrated into the complete system.

A. System Architecture

Nani AGI is organized into five primary functional layers: **perception, cognition, memory, decision-making, and action**. The perception layer receives user input through text, speech, and visual information. Speech input is converted into textual representation using automatic speech recognition, while visual inputs can be processed using computer-vision and optical-character-recognition techniques.

The cognition layer uses a large language model to interpret user intent, maintain conversational context, generate responses, and perform reasoning. The memory layer maintains relevant short-term and long-term information, including previous interactions, user preferences, important tasks, and contextual information. The decision-making layer determines appropriate responses or actions based on the current input and retrieved memory. Finally, the action layer enables Nani AGI to communicate through synthesized speech and perform authorized computer-level operations.

B. Software and Frameworks

The prototype employs Python as the primary programming language because of its extensive ecosystem for artificial intelligence, natural-language processing, speech processing, and automation. Automatic speech recognition can be implemented using **Vosk**, while multilingual language identification can be supported using **langdetect**. Speech synthesis is implemented using **Edge TTS** to generate spoken responses. Large language model capabilities can be integrated through model APIs or compatible inference frameworks, while operating-system automation is implemented through appropriate Python-based system-control interfaces.

The architecture is designed to remain modular so that individual AI models or services can be replaced without redesigning the complete system.

C. Memory and Personalization

A persistent memory mechanism is incorporated to maintain information that remains relevant beyond a single interaction. User preferences, previously established contextual information, and task-related information are stored and retrieved when required. Memory retrieval is performed before response generation so that the language model can incorporate relevant historical information into the current interaction.

This approach enables Nani AGI to provide personalized responses rather than treating every interaction as an independent conversation.

D. Multimodal and Emotion-Aware Processing

The proposed system supports multiple input modalities, including text, speech, and visual information. Speech is processed through a speech-recognition pipeline, while visual information can be analyzed through computer-vision modules. Emotion-aware processing is incorporated to identify relevant emotional cues from user interaction and adjust the tone and nature of generated responses accordingly.

E. Experimental Evaluation

The system is evaluated using task-oriented and conversational test cases. Key evaluation parameters include **task-completion rate, contextual-memory accuracy, personalization accuracy, response latency, intent recognition, and interaction effectiveness**. Experiments are conducted by comparing the integrated Nani AGI architecture with reduced configurations in which selected components, such as persistent memory or emotion-aware processing, are disabled. This enables the contribution of individual architectural components to be examined through comparative and ablation-based evaluation.

Quantitative results are represented using tables and graphical analysis, while qualitative observations from interaction trials are used to identify usability issues, response-quality differences, and practical limitations.

F. Ethical and Privacy Considerations

Since Nani AGI is designed to process potentially sensitive conversational and personal information, privacy and user control are considered fundamental design requirements. Data collection and storage should be limited to information necessary for system functionality, and user authorization should be required before executing computer-level actions. The architecture also supports the use of locally executed models and processing where feasible, reducing unnecessary transmission of personal information to external services.

IV. RESULTS AND DISCUSSION

The evaluation of Nani AGI focuses on determining whether the integration of persistent memory, personalization, multimodal interaction, emotion-aware processing, reasoning, and task execution improves the effectiveness of human–AI interaction. The assessment considers task completion, contextual continuity, personalization, response latency, and interaction effectiveness as the primary evaluation parameters.

The integrated architecture is expected to provide more continuous interactions than a conventional stateless conversational system because relevant information from previous interactions can be retrieved and incorporated into subsequent responses. Persistent memory enables the system to retain user preferences and task-related context, thereby reducing the need for users to repeatedly provide the same information. Similarly, the integration of speech, visual input, and text expands the interaction possibilities beyond conventional text-only communication.

The effectiveness of individual components can be examined through comparative ablation experiments. A baseline configuration containing only conversational reasoning can be compared with configurations incorporating memory, personalization, emotion-aware processing, and task execution. Such comparisons can determine which components contribute most significantly to overall system performance.

The proposed architecture also extends the reasoning-and-action approach demonstrated by previous agentic systems by combining task execution with persistent personalization and multimodal interaction. Similarly, its memory mechanism builds upon previous research demonstrating the importance of memory and reflection for persistent agents, while extending these concepts toward a personal-assistant setting.

V. CONCLUSION

Nani AGI presents a unified architecture for developing personalized and adaptive AI assistants by bringing together conversational reasoning, persistent memory, multimodal interaction, emotion-aware processing, task planning, and computer-level automation. The proposed approach addresses a key limitation of conventional conversational systems, where interactions are often isolated and lack persistent understanding of individual users and their ongoing context.

The architectural design demonstrates how these complementary capabilities can be organized into modular layers, allowing individual components to be developed, evaluated, and replaced independently. This modularity provides a practical foundation for extending AI assistants from passive response-generation systems toward more continuous and user-oriented intelligent agents.

The significance of the proposed work lies in its focus on **integration and personalization** rather than introducing a single isolated AI capability. The framework can support applications in personal productivity, accessibility, desktop automation, education, and intelligent human–computer interaction.

Future work will focus on improving autonomous planning, multimodal reasoning, long-term memory management, multilingual interaction, emotion recognition, offline intelligence, security, and integration with IoT devices. Further large-scale user studies and controlled experiments are also required to quantitatively validate the effectiveness, reliability, scalability, and real-world usability of the proposed Nani AGI architecture.

REFERENCES

[1] T. Brown *et al.*, “Language models are few-shot learners,” in *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.

[2] OpenAI *et al.*, “GPT-4 technical report,” *arXiv preprint arXiv:2303.08774*, 2023.

[3] S. Yao *et al.*, “ReAct: Synergizing reasoning and acting in language models,” in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2023.

[4] C. Packer *et al.*, “MemGPT: Towards LLMs as operating systems,” *arXiv preprint arXiv:2310.08560*, 2023.

[5] J. S. Park *et al.*, “Generative agents: Interactive simulacra of human behavior,” in *Proc. 36th Annual ACM Symposium on User Interface Software and Technology (UIST)*, 2023, pp. 1–22.

[6] H. Rashkin, E. M. Smith, M. Li, and Y.-L. Boureau, “Towards empathetic open-domain conversation models: A new benchmark and dataset,” in *Proc. 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 5370–5381.

[7] Z. Lian *et al.*, “AffectGPT: A new dataset, model, and benchmark for emotion understanding with multimodal large language models,” *arXiv preprint arXiv:2501.16566*, 2025.

[8] L. Wang *et al.*, “A survey on large language model based autonomous agents,” *arXiv preprint arXiv:2308.11432*, 2023.