International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5218 - 5225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0683

Generative AI: Beyond ChatGPT

Mr Syed Jamesha S N[1], Vasuki M[2], Sadhana sri S[3], Nathira M[4], Rithanya S[5], Senthil Kumar S[6]
[1]Assistant Professor, Department of Electronics and Communication Engineering, Al-Ameen Engineering College (Autonomous),
[2]Assistant Professor, Department of Computer Science Engineering, Al-Ameen Engineering College (Autonomous),
[3,4,5,6]Student, Department of Computer Science Engineering, Al-Ameen Engineering College (Autonomous),
Emails: *aecsyedjamesha@gmail.com[1], vasukicse95@gmail.com[2], sadhanasri477@gmail.com[3], mnathira786@gmail.com[4], rithanvarshan2024@gmail.com[5], senthilkumarselvakumarer@gmail.com[6]*

Abstract
Generative Artificial Intelligence (AI) has become an important area of computer science, changing the way people create, process, and interact with digital content. Although ChatGPT has made Generative AI widely known, its capabilities extend far beyond conversational systems. Generative AI includes several technologies, such as Large Language Models (LLMs), Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Diffusion Models, which can generate text, images, audio, video, software code, and other forms of content. This paper examines the development of Generative AI and its applications in different fields. In education, it can support personalized learning and content generation, while in healthcare it can assist with medical imaging and drug research. In software development, Generative AI can support code generation and debugging. It is also being used in entertainment, business automation, and scientific research. However, the rapid growth of this technology has introduced several challenges, including inaccurate or misleading outputs, privacy concerns, copyright issues, bias in generated content, and security risks. These challenges highlight the importance of responsible development and use of Generative AI. This paper discusses how Generative AI is evolving beyond ChatGPT and explores its potential to support human creativity and problem-solving. It also emphasizes the need for reliable, transparent, secure, and responsible AI systems for future applications.

Keywords*: Generative AI, Large Language Models, GANs, VAEs, Diffusion Models, Artificial Intelligence, Responsible AI*

1. Introduction
Artificial Intelligence (AI) is the branch of computer science concerned with building systems that perform tasks normally associated with human intelligence, such as perception, reasoning, language understanding, and decision-making. Within AI, Machine Learning (ML) and, more specifically, Deep Learning (DL) have enabled models to learn complex patterns directly from data. Generative AI is a subfield of Deep Learning focused on models that produce new content — text, images, audio, video, or code — rather than only classifying or predicting labels for existing data. The evolution of Generative AI can be traced from early rule-based and statistical language systems, through deep neural sequence models, to the modern era defined by three parallel developments: Generative Adversarial Networks (GANs) [1], Variational Autoencoders (VAEs) [2], and the Transformer architecture [3], which underpins today's Large Language Models (LLMs). Transformer-based LLMs such as GPT-3 [4] demonstrated that scaling model size and training data yields strong few-shot generalization, and the subsequent application of instruction tuning and Reinforcement Learning from Human Feedback (RLHF) [5], [6] produced conversational assistants capable of following natural-language instructions.
ChatGPT became popular largely because it packaged a capable LLM behind a simple, free-form conversational interface, with RLHF-based alignment [5], [6] making responses feel cooperative

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5218 - 5225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0683

and safe enough for mainstream use. However, Generative AI extends well beyond conversational text. GANs [1] and VAEs [2] enabled learned image synthesis; diffusion models [7], [8] subsequently became the dominant approach for high-fidelity image and, increasingly, video generation [9]; and multimodal foundation models now combine several of these techniques to generate or reason across text, images, and video within a single system. Despite this progress, current Generative AI systems — including chat-based assistants and multimodal foundation models — still exhibit several recurring problems: hallucination, i.e., confident generation of unsupported or incorrect content [10]; limited persistent personalization across sessions; inconsistent reliability, since outputs are rarely automatically checked before being shown to the user; privacy risk associated with storing or reusing conversational data; bias inherited from training data; copyright and intellectual-property exposure in generated media; and security concerns such as prompt injection and model misuse. These problems are usually addressed in isolation — a personalization feature here, a safety filter there — rather than as parts of one coordinated system. This observation motivates the framework proposed in this paper. Rather than proposing a new base model, we propose a system-level architecture that combines consent-governed personal memory, a coordinated multi-agent structure for multimodal generation, and an explicit human verification stage, so that personalization, coordination, and reliability are treated as first-class, interacting concerns rather than afterthoughts. Research Gap: The literature contains substantial work on LLMs [3], [4], GANs [1], VAEs [2], diffusion models [7], [8], multi-agent LLM systems [11], [12], and human-in-the-loop alignment [5], [6], [13], but these strands are largely studied separately. There is limited published work on an integrated framework that combines persistent, user-consented personalization; a multi-agent division of labor across text, image, and video generation; and a confidence-based routing mechanism to human reviewers, evaluated as a single coordinated system rather than as independent components. Contribution: This paper's contribution is (1) a system-level architecture integrating Personal Memory, Multi-Agent generation, and Human Verification; (2) a description of agent responsibilities and communication for text, image, and video generation; (3) a proposed evaluation methodology and metric set for comparing this framework against single-agent and verification-only baselines; and (4) an implementation roadmap toward a working prototype. We do not claim that any individual component (memory, multi-agent orchestration, or human-in-the-loop verification) is novel in isolation; the contribution is their integrated design and evaluation.

2. Existing System

Most deployed Generative AI systems today are built around a single model — or a single model augmented with tools — that handles a user's request end-to-end.

2.1.Conventional LLM-Based and Chatbot Systems

Systems such as ChatGPT, Claude, and Gemini present a conversational interface in which a single LLM, trained with instruction tuning and RLHF [5], [6], produces a response for each turn. Session context is typically limited to the current conversation window; persistent memory across sessions, where offered, is usually a lightweight, largely opaque feature rather than a structured, user-governed component.

2.2.Single-Agent AI Systems

In most production systems, one model (optionally calling external tools or plugins) is responsible for interpreting the request, generating the content, and returning it. There is no explicit division of labor among specialized agents, and no dedicated agent whose role is to verify another agent's output before it reaches the user.

2.3.Existing Text Generation

Transformer-based LLMs [3], [4] generate fluent, contextually relevant text and underlie most existing text-generation systems. Their well-documented weakness is hallucination — fluent but unsupported or factually incorrect output [10].

2.4.Existing Image Generation

GANs [1] and VAEs [2] were the first widely used deep generative approaches for images; diffusion models [7], including latent diffusion approaches such as those underlying Stable Diffusion [8], now

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5218 - 5225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0683

produce high-fidelity images from text prompts and are used in commercial tools such as DALL-E and Midjourney.

2.5.Existing Video Generation

Text-to-video systems extend diffusion and transformer techniques to the temporal domain [9]; commercial systems (e.g., Sora-class models) demonstrate short, coherent video clips from text prompts, but as with image models, outputs are not automatically verified for factual or safety concerns before being returned.

2.6.Existing Multimodal Systems

Multimodal foundation models (e.g., GPT-4-class vision-enabled models, Gemini) accept and/or produce more than one modality within a single integrated model. This improves cross-modal consistency compared to stitching together separate tools, but coordination remains internal to one model rather than an explicit, inspectable multi-agent process.

2.7.Existing Memory and Personalization Approaches

Retrieval-Augmented Generation (RAG) allows a model to condition on external documents at inference time, and some commercial assistants offer opt-in memory of user facts. However, published, systematic treatments of consent management, memory scope control, and privacy-by-design for conversational memory remain limited relative to the maturity of the underlying generative models [14].

2.8.Existing Human-in-the-Loop Approaches

RLHF [5], [6] and related preference-learning methods incorporate human feedback primarily during training/fine-tuning, shaping model behavior in aggregate. This differs from a runtime, per-output human verification step that reviews a specific generated item before it is delivered, which is the mechanism this paper proposes. As shown in Table 1 Summary of Limitations in Existing Generative AI Systems. This paper does not claim that existing systems entirely lack these features — several commercial systems offer partial memory, moderation, or safety filtering. Rather, the contribution of this work is the integrated framework described in Sections III–VI and its proposed evaluation, not the invention of any single mechanism in isolation.

Table 1 Summary of Limitations in Existing Generative AI Systems

Aspect	Limitation
Personalization	Session-bound or opaque; limited persistent, user-governed memory
Agent structure	Single-agent/monolithic; no explicit division of labor
Reliability	Hallucination [10]; outputs largely unverified before delivery
Human oversight	RLHF shapes training [5],[6]; rarely applied per-output at runtime
Multimodal coordination	Internal to one model; not an inspectable multi-agent process
Privacy	Data retention/consent practices inconsistently documented [14]
Bias / Copyright / Security	Inherited from training data; prompt-injection and misuse risks remain

3. Proposed System

We propose the Personalized, Multi-Agent and Human-Verified Multimodal Generative AI Framework, which integrates three innovations — Personal Memory, Multi-Agent Generative AI, and Human Verification — into a single coordinated pipeline supporting text, image, and video generation.

3.1.Personal Memory

The Personal Memory module stores user-approved contextual information — stated preferences, preferred language/style, prior project context, learning preferences, and summarized relevant conversation history — for reuse in future interactions. Memory is opt-in: information is written to memory only after explicit user consent, is visible and editable by the user, can be deleted on request, and is stored in an access-controlled, encrypted store consistent with data-protection principles such as those in the GDPR [15] and general Responsible AI guidance [16]. Sensitive categories of information are excluded from automatic storage. The memory module supplies the Manager Agent with a compact,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5218 - 5225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0683

relevant context snippet rather than raw conversation logs, limiting exposure of stored data.

3.2.Multi-Agent Generative AI

Instead of one model handling every request, the framework defines a Manager/Orchestrator Agent that classifies the incoming task and routes it to one or more specialized agents: a Text Agent, an Image Agent, and a Video Agent, drawing on established patterns for coordinating multiple LLM-driven agents [11], [12]. Agents communicate through a shared, structured context object (task identifier, modality, prompt, retrieved memory, intermediate output, confidence score, status) rather than free-form text, which keeps coordination inspectable and auditable. A dedicated Verification Agent is a first-class participant in this pipeline rather than an optional add-on.

3.3.Human Verification

Generated content is not delivered directly to the user. The Verification Agent automatically evaluates each output and produces a confidence/quality score. Outputs above a defined confidence threshold, and not flagged as sensitive, proceed to Final Output. Outputs that are low-confidence, high-risk (e.g., medical, legal, safety-related), or ambiguous are routed to a Human Verification interface, consistent with human-AI interaction guidance [13], before being finalized. This provides a runtime safeguard that complements, rather than replaces, training-time alignment such as RLHF [5], [6]. End-to-end, the proposed workflow is: User Input → Memory Retrieval → Personalization → Manager Agent → Task Classification → Specialized Agent Selection → Content Generation → Verification → Confidence/Quality Evaluation → Human Verification (if required) → Final Output → Optional, consented Memory Update. The final memory-update step is deliberately gated by explicit user consent and the same privacy policy that governs memory retrieval, so that the system's personalization improves over time without silently expanding what is stored about the user.

4. Architecture

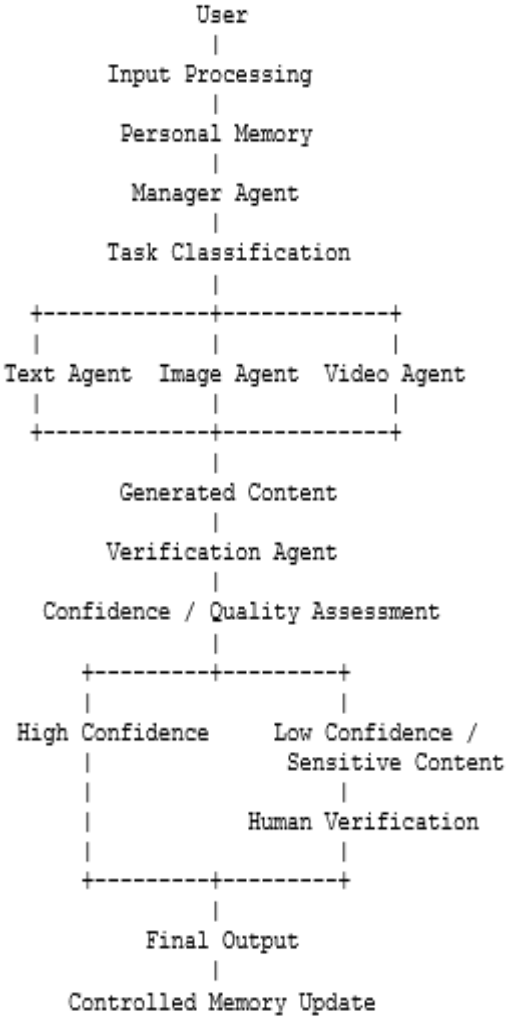
The proposed architecture consists of fourteen components spanning the interface, memory, orchestration, generation, verification, and safety layers, summarized in Table 2.

Table 2 Architecture Components and Responsibilities

Component	Responsibility
User Interface	Captures user input and displays final output
Authentication / Profile	Identifies user; links to their memory store
Personal Memory Module	Stores consented user context long-term
Memory Retrieval Module	Retrieves relevant memory for the current task
Manager / Orchestrator Agent	Classifies task; routes to specialized agents
Text Generation Agent	Produces personalized text content
Image Generation Agent	Produces personalized image content
Video Generation Agent	Produces personalized video content
Verification Agent	Automatically evaluates generated output
Confidence/Quality Module	Scores output; decides routing
Human Verification Interface	Presents flagged output to reviewers
Responsible AI / Safety Layer	Applies policy, bias, and safety checks
Final Output Layer	Delivers the approved output to the user
Controlled Memory Update	Writes new context back only with consent

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5218 - 5225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0683

Figure 1 Proposed system architecture



Control flows top-down from the user through personalization and orchestration to modality-specific generation, then back up through verification and, where required, human review, before the response is delivered and (optionally) folded back into memory. The Responsible AI/Safety Layer operates across the pipeline rather than at a single point, applying policy checks at generation, verification, and memory-update stages. As shown in Figure 1 Proposed system architecture.

5. Ai Interaction

The following scenarios illustrate the proposed interaction flow for each modality.

5.1.Scenario 1. Text (Personalized Educational Content)

A student who previously indicated (with consent) a preference for beginner-level explanations and Python examples asks for an explanation of recursion. Flow: User Input → Memory (retrieves prior level/language preference) → Manager Agent classifies this as a text task → Text Agent generates a beginner-level, Python-based explanation → Verification Agent checks technical correctness and clarity → high confidence → Final Output. No human review is required in this low-risk, high-confidence case.

5.2.Scenario 2. Image (Style-Consistent Diagram)

The same student asks for "a diagram like before" for a new topic. Flow: User Input → Memory (retrieves stored visual-style preference, e.g., minimalist infographic style) → Manager Agent routes to the Image Agent → Image Agent generates a diagram consistent with the stored style → Verification Agent checks content relevance and screens for unsafe or copyrighted elements → if confidence is high, Final Output; if the request touched a sensitive or ambiguous depiction, it is routed to Human Verification first.

5.3.Scenario 3. Video (Short Educational Video)

The user requests a 60-second explainer video on a science topic. Flow: User Input → Memory → Manager Agent decomposes the task, invoking the Text Agent (script) and the Video Agent (visuals/narration) in sequence, coordinated through the shared context object → Verification Agent evaluates factual alignment between script and narration and screens visual content → because video generation carries higher risk of subtle factual or visual errors, this scenario is more likely to be routed to Human Verification before Final Output. Across all three scenarios, agents communicate via a shared context object rather than ad hoc messages: the Manager Agent writes the task specification and retrieved memory into this object, each specialized agent reads its relevant fields and writes back its output and self-reported confidence, and the Verification Agent reads the completed object to perform its assessment. This keeps inter-agent communication structured, logged, and auditable.

6. Methodology

6.1.Dataset / Test Cases

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5218 - 5225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0683

We propose evaluating the framework on a combination of (i) established text, image, and video generation benchmark prompts drawn from the existing literature, and (ii) a purpose-built set of personalization scenarios (e.g., paired requests before and after a stored preference) to test memory-driven personalization specifically. Concrete dataset selection is left to the implementation phase and is not fabricated here.

6.2.System Architecture

As described in Section IV, comprising the memory, orchestration, generation, verification, and human-review layers.

6.3.Agent Design

Each agent (Manager, Text, Image, Video, Verification) is implemented as an independently invocable module with a defined input/output schema, allowing any agent's underlying model to be replaced without altering the overall pipeline (a model-agnostic design).

6.4.Memory Mechanism

structured profile store (explicit preference fields) combined with a vector-based semantic retrieval store (for relevant prior context) [14], with consent logged at write time and honored at retrieval time.

6.5.Multimodal Generation

Text generation via an instruction-tuned LLM [4]; image generation via a diffusion-based model [7], [8]; video generation via a diffusion- or transformer-based text-to-video pipeline [9]. The framework is architecture-agnostic with respect to which specific underlying models are used.

6.6.Verification Mechanism

Automatic checks combining self-consistency/factuality prompting for text, artifact and policy screening for images, and frame-level consistency and factual-alignment checks for video, feeding a single confidence/quality score per output.

6.7.Human Evaluation

A defined reviewer protocol: stratified sampling of outputs (including all outputs below the confidence threshold plus a random sample of high-confidence outputs for audit), a rating rubric covering accuracy, relevance, personalization quality, and safety, and inter-rater agreement reporting.

6.8.Experimental Setup

Three configurations are proposed for comparison, described in Table 3.

Table 3 Proposed System Comparison

System	Personalization	Agents	Verification
System 1: Standard GenAI	None / session-only	Single agent	None
System 2: GenAI + Verification	None / session-only	Single agent	Automatic only
System 3: Proposed Framework	Consented, persistent memory	Multi-agent	Automatic + human-in-the-loop

6.9.Evaluation Metrics

Table 4 lists the proposed evaluation metrics.

7. Results (Expected / Proposed Evaluation)

No experiments have been run for this proposal; the results below are expected outcomes to be confirmed by future implementation, not experimental findings.

7.1.Expected Performance Comparison

We expect System 3 (proposed framework) to show higher Personalization Quality and Human Acceptance Rate than Systems 1 and 2, and a lower Hallucination Rate than System 1, at the cost of higher Response Time and Computational Cost due to multi-agent coordination and human review. Verification Accuracy is expected to improve from System 2 to System 3 because human review directly checks the automatic verifier's low-confidence decisions.

7.2.What Should Be Measured

Post-implementation work should measure all metrics in Table IV across a held-out test set for each of the three system configurations, report statistical significance of differences, and separately report failure cases where the human reviewer overturned the automatic verifier's decision, to characterize where the framework's added complexity is actually earning its cost.

8. Limitations

- Increased computational requirements from

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5218 - 5225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0683

running multiple specialized agents rather than one model.

- Memory storage introduces privacy and data-protection obligations that must be actively managed [15].
- Performance depends on the quality of the underlying models chosen for each agent.
- Agent-to-agent communication adds coordination overhead and potential latency.
- Verification and human review add latency compared to a single-pass response.
- Human reviewer workload can become a bottleneck at scale.
- Automatic evaluation of image and video quality/correctness remains technically harder than text evaluation.
- Hallucination is expected to be reduced, not eliminated, by verification and human review.
- Bias present in underlying training data may still surface in outputs.
- Copyright and intellectual-property risk in generated media is not fully resolved by this framework.
- Security risks such as prompt injection and adversarial inputs require dedicated mitigation beyond this proposal's scope.
- Data poisoning of the memory store is a risk that requires input validation and monitoring.
- The framework depends on external AI models/APIs whose availability, cost, and policies may change.

9. Future Scope

- Real-time multimodal agents capable of low-latency generation and verification.
- Better long-term memory mechanisms, including forgetting/decay policies.
- Privacy-preserving memory, e.g., on-device or encrypted memory stores.
- Federated learning approaches to personalization without centralizing raw user data.
- More autonomous agent collaboration protocols, building on multi-agent LLM research [11], [12].
- Improved automatic verification models, particularly for image and video correctness.
- More advanced, controllable video generation techniques.
- Domain-specific agents for fields such as healthcare, law, or education.
- Explainable Generative AI to make agent and verification decisions more interpretable.
- Edge AI deployment to reduce latency and data-transmission privacy exposure.
- Energy-efficient Generative AI to reduce the computational and environmental cost of multi-agent pipelines.
- Continued development of Responsible AI mechanisms addressing bias, safety, and accountability [16].

Conclusion

Generative AI has grown well beyond single-purpose chat interfaces, but current systems still generally rely on a single model, limited persistent personalization, and little runtime verification of what they produce. This paper proposed a framework that integrates consent-governed Personal Memory, a Multi-Agent architecture for coordinated text, image, and video generation, and a Human Verification layer that routes low-confidence or sensitive outputs to human reviewers. We described the proposed architecture, interaction scenarios, and an evaluation methodology comparing the framework to single-agent and verification-only baselines, without fabricating experimental results. This framework does not claim to fully solve hallucination or reliability problems in Generative AI; rather, it proposes a system-level, human-centered structure intended to reduce their impact while keeping personalization and generation decisions auditable and consent-governed, in line with Responsible AI principles [16].

References

[1]. Z. Ji et al., “Survey of Hallucination in Natural Language Generation,” ACM Computing Surveys, vol. 55, no. 12, 2023.
[2]. T. Chakraborty et al., “Ten Years of Generative Adversarial Nets: A Survey of the State-of-the-Art,” 2023.
[3]. “A Survey on the Memory Mechanism of

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5218 - 5225
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0683

Large Language Model-Based Agents," ACM Transactions on Information Systems, 2025.

[4]. D. Caffagni et al., "The Revolution of Multimodal Large Language Models: A Survey," Findings of ACL, 2024.

[5]. Z. Lin et al., "Towards Trustworthy LLMs: A Review on Debiasing and Dehallucinating in Large Language Models," Artificial Intelligence Review, vol. 57, article 243, 2024.