IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 12, Issue 11, November 2025 https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2025.v1211.01

AI-Powered Digital Memory Bank

Yashwanth MS, Viresh Hubballi, Vikram DP, Sushmitha
Department of Computer Science and Engineering
MVJ College of Engineering, Bengaluru, India
yashwanthms2005@gmail.com,vireshhubballi7@gmail.com
vikramdp505@gmail.com, ganigasushmitha28@gmail.com
Prof.Thejaswini M
Professsor, Department of Computer Science and Engineering
MVJ College of Engineering, Bengaluru, India
thejashwini_aiml@mvjce.edu.in

Publication History
Manuscript Reference No: IJIRAE/RS/Vol.12/Issue11/NVAE10080
Research Article| Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRAE/RS/Vol.12/Issue11/NVAE10080
Received:22,October 2025,Revised:28,October 2025, Accepted:31, October 2025, Published Online: 21, November 2025.
https://www.ijirae.com/volumes/Vol12/iss-11/01.NVAE10080.pdf
Citation: Prof.Thejaswini,Yashwanth,Viresh,Vikram,Sushmitha(2025), AI-Powered Digital Memory Bank ,IJIRAE: International Journal of Innovative Research in Advanced Engineering, Volume 12, Issue 11 of 2025 pages 437-443
doi:>https://doi.org/10.26562/ijirae.2025.v1211.01
BibTeX Key: Prof.Thejaswini@2025AI-Powered
IJIRAE papers should be cited as IJIRAE (International Journal of Innovative Research in Advanced Engineering, AM Publications, India 2025, ISSN 2349-2163, https://doi.org/10.26562/ijirae.2025.v1211.01 The journal's official abbreviation is IJIRAE. Orcid: https://orcid.org/0009-0004-9398-7488

Abstract: In today's digital landscape, individuals face an overwhelming influx of information, which complicates the efficient retrieval, organization, and management of personal memories. Traditional tools like digital calendars and note-taking apps often fall short in offering intelligent retrieval, personalization, and contextual awareness, leading to disjointed memory management. This paper introduces the AI-Powered Digital Memory Bank, an intelligent, context-aware system designed to function as a virtual extension of human memory. The system accommodates text, voice, and image inputs and utilizes natural language processing (NLP), machine learning, and semantic tagging to organize the data intelligently. Users can retrieve memories by engaging in natural conversational queries, enabling context-driven and associative recall similar to human memory. In addition to basic storage, the system features smart reminders, behavioral insights, and adaptive recommendations, thus enhancing the user's productivity and mental well-being. With its secure, scalable, and user-friendly design, the system shows potential for use in personal memory management, cognitive support, and digital well-being, marking a new advancement in connecting human cognition with artificial intelligence.
Keywords: Digital Memory Bank, Natural Language Processing, Machine Learning, Context-Aware Retrieval, Cognitive Assistance

I. INTRODUCTION

In the current digital era, people are continuously exposed to enormous amounts of information from emails, chats, meetings, social media, and personal experiences. Digital technologies have simplified the process of capturing and storing data, but they have also led to the issue of dispersed and fragmented information that is dispersed across several platforms. Because users frequently forget the precise file names, keywords, or storage locations, this causes problems remembering crucial information. Inefficiency, lost opportunities, and cognitive stress result from an inability to efficiently organize and recall memories. The associative aspect of human memory is not replicated by conventional technologies like calendars, note-taking apps, and cloud storage systems, which lack contextual knowledge. For example, users anticipate seeing similar notes, documents, and action items together when they recollect "the conversation with a colleague about a project update last week." Nevertheless, current solutions only offer keyword-based results, requiring users to manually connect data, which is laborious and prone to mistakes. Additionally, their efficacy is diminished when intelligent elements like summaries, behavior-based insights, and reminders are absent. This paper suggests an AI-Powered Digital Memory Bank, a system intended to function as a virtual extension of human memory, in order to address these issues. The platform uses contextual tagging, machine learning, and natural language processing (NLP) to store and arrange data in an intelligent manner. It also supports multimodal inputs, including text, speech, and images. Natural, conversational queries can then be used to retrieve memories, allowing for associative and context-aware recall that is comparable to human memory. The system is a proactive rather than a passive repository since it provides summaries, reminders, behavioral insights, and predictive recommendations in addition to storage and retrieval. By mimicking some characteristics of human cognition, the suggested approach can assist users in remembering commitments, thinking back on prior experiences, and making smarter judgments with less mental strain. This method has the potential to help people with memory-related issues as well as improve daily productivity, which would help the system close the gap between AI and human cognition.

II. PROBLEM STATEMENT

In the current digital age, when people manage data in many formats including notes, emails, conversations, voice recordings, and photos, managing and remembering personal information has become increasingly difficult. Despite being readily accessible, storage tools frequently rely on keyword-based search or manual organisation, which is inefficient when users are unable to recall specific facts. Cognitive overload, inefficiency, and fragmented information are the outcomes of this. Furthermore, two essential components of human memory contextual awareness and associative recall are absent from current implementations. For example, remembering a project conversation should gather meeting minutes, relevant papers, and action items; however, current systems need users to search independently across many platforms. Their efficacy is further diminished by the lack of proactive features like behavioral insights, summaries, and reminders. An intelligent system that can store multimodal information, comprehend context, and deliver natural, timely, and meaningful recall is required to get around these restrictions. By mimicking human-like memory, the suggested solution satisfies this demand and improves decision-making, organization, and productivity.

III. LITERATURE SURVEY

Savini Kashmira et al.[1] proposed TOBU Graph, a graph-based retrieval framework that constructs knowledge graphs dynamically from unstructured data using LLMs. Unlike traditional Retrieval-Augmented Generation (RAG), which depends on text-to-text similarity of chunks, TOBU Graph extracts structured knowledge and diverse semantic relationships to enhance retrieval. The system leverages graph traversal for accurate retrieval, eliminating dependency on chunking strategies and reducing hallucination. Their evaluation using real-world user data in the TOBU application showed that TOBU Graph achieved 93.74% precision, 91.96% recall, and 92.84% F1-score, significantly outperforming multiple RAG implementations and improving user experience.

Kwangseob Ahn [2] introduced HEMA (Hippocampus-Inspired Extended Memory Architecture) to address the limitations of large language models in sustaining long-context conversations. HEMA integrates two memory components: Compact Memory, a continuously updated one-sentence summary preserving the global narrative, and Vector Memory, an episodic store of chunk embeddings retrieved via cosine similarity. Implemented with a 6B-parameter transformer, HEMA sustained dialogues exceeding 300 turns while keeping prompts under 3.5K tokens.

On long-form QA and story continuation benchmarks, HEMA improved factual recall accuracy from 41% to 87% and raised human-rated coherence scores from 2.7 to 4.3. Precision recall analysis further showed P@5≥0.80 and R@50≥0.74 with10K indexed chunks, demonstrating superior retrieval quality and scalability over summarization-only baselines.

Sergey Legtchenko et al. [3] proposed Managed - Retention Memory (MRM), a new class of memory designed for AI inference workloads. Unlike traditional High Bandwidth Memory (HBM), which is over provisioned for write performance and under provisioned for density and read bandwidth, MRM relaxes long-term retention requirements to optimize read throughput, energy efficiency, and memory capacity. By exploiting predictable and sequential memory access patterns of foundation model inference, MRM trades retention time and write performance for metrics critical to AI workloads. The system leverages emerging memory technologies, such as PCM, RRAM, and STT-MRAM, to provide higher density and better energy efficiency compared to HBM. MRM also introduces software-aware memory management, including retention-aware data placement, light weight memory controllers, and dynamically configurable retention policies. This design can reduce cost and energy overheads in AI clusters while maintaining high performance for large-scale model inference.

Pranab Sahoo et al. [4] presented a systematic survey on prompt engineering techniques for large language models (LLMs) and vision-language models (VLMs). The study categorizes over 41 prompting methods based on application areas, covering basic approaches like zero-shot and few-shot prompting to advanced strategies including Chain-of-Thought (CoT), Tree-of-Thoughts (ToT), Graph-of-Thoughts (GoT), Retrieval-Augmented Generation(RAG),Self-Refine, and Code Prompting. Each method is analyzed for its prompting methodology, model applicability, datasets used, performance improvements, and strengths and limitations. The survey also introduces a taxonomy diagram and tables summarizing datasets, models, and evaluation metrics. Key insights include enhanced reasoning accuracy (e.g.,CoT with PaLM540B achieving 90.2% on math benchmarks), reduced hallucinations via RAG and CoVe, improved code generation through Structured CoT (SCoT) and Chain-of-Code(CoC), and efficiency gains with Chain of Draft (CoD) and Buffer of Thoughts (BoT). This work provides a structured understanding of the evolving landscape of prompt engineering and identifies open challenges and opportunities for future research.

Weizhi Wangetal.[5] proposed LONGMEM, a framework for augmenting large language models (LLMs) with long-term memory to memorize and utilize long-form context. Unlike existing methods that scale input length or rely on sparse attention, LONGMEM introduces a decoupled architecture with a frozen backbone LLM as a memory encoder and a residual Side Net as a memory retriever and reader. The model caches attention key-value pairs of previous inputs in a memory bank, retrieves relevant memory via token-to-chunk attention, and fuses it with current input through a memory-augmented layer. Experiments on long-text language modeling (PG-22, ArXiv) and long-context understanding (Chapter Break) demonstrate significant improvements in perplexity and suffix identification accuracy, while memory-augmented in-context learning enables LLMs to attend to thousands of demonstration examples beyond local context limits. Ablation studies show that memory size and chunk size critically affect performance.

S.Gensburger and F.Clavert [6] explored the relationship between artificial intelligence (AI) and collective memory, focusing on how generative AI, such as Chat GPT, could act as a new infrastructure for memory. They discussed AI's role in memorialization, digital afterlife, heritage, archives, and education, highlighting both the opportunities and ethical challenges of algorithmically mediated memory. The authors emphasized the interplay between human and non-human agents in shaping collective memory, the risks of hegemonic narratives, and the importance of integrating memory studies perspectives into AI development. They also addressed the temporal and spatial limitations of AI data, digital labor considerations, and the potential of AI to recontextualize and reinterpret collective memory.

Hoskins [7] examined how generative AI reshapes human and collective memory through the creation of a "third way of memory," where AI systems both produce and remix past events.

Unlike traditional memory studies that focus on human encoding and recall, his approach highlights AI's capacity to aggregate vast digital traces and generate a past that was never actually encoded into human memory. The study discusses AI chatbots and digital memory platforms, such as ChatGPT and Personal.AI, which provide both explicit and implicit memory services, enabling novel artifacts from user data while challenging human agency, privacy, and control over remembering and forgetting. Hoskins argues that these developments produce a new memory ecology, blending human and machine contributions to the past and fore grounding ethical and societal implications of AI-driven memory reconstruction.

Savya Khosla et al. [8] presented a comprehensive survey on Memory-Augmented Neural Networks (MANNs), exploring how human-like memory mechanisms inspire AI architectures and applications. Unlike standard neural networks, which struggle with long-term dependencies and in-context learning, MANNs integrate memory modules to store, retrieve, and manipulate information over extended sequences. The survey covers memory-inspired architectures including Hopfield Networks, Neural Turing Machines, Correlation Matrix Memories, Mem former, and Neural Attention Memory, highlighting their storage, recall mechanisms, and advantages. It further examines applications across Natural Language Processing (e.g., RAG, REALM, RETRO), Computer Vision (e.g., RAC, REACT, KNN-Diffusion), and Multimodal Learning (e.g., MuRAG, RA-VQA, RA-CM3), demonstrating how memory augmentation improves accuracy, efficiency, and trustworthiness without increasing model parameters. The authors also discuss challenges such as retrieval reliability, modality gaps, and computational over head, providing insights for designing more effective and interpretable memory-augmented AI systems.

IV. DRAWBACKS OF EXISTING SYSTEM

- Current technologies, such as note-taking applications and cloud storage, only provide rudimentary storage and keyword search. Information retrieval is slow, in effective, and cognitively taxing because they are unable to aggregate data from documents, audio notes, emails, and chats.
- The current systems require a great deal of human organization and tagging. This procedure is laborious and prone to errors, frequently leading to lost or insufficient data.
- The majority of programs lack contextual awareness and are unable to mimic human-like associative memory. For example, it should link notes, papers, and action items while recalling a meeting. Users must instead search on several platforms.
- Without intelligent capabilities like adaptive learning, summarisation, predictive recall, or reminders, these systems function as passive repositories. Moreover, static models are out of date because they are unable to update knowledge dynamically.
- Large language models and traditional systems have memory limits, which limit scalability. They perform poorly when managing big amounts of personal data and lose track of previous context in lengthy encounters.

V. ADVANTAGES OF PROPOSED SYSTEM

- The suggested system integrates several forms of input, such as speech, text, and graphics. For simple recall, it links memories according to persons, time, and context using natural language processing (NLP) and semantic tagging.
- The amount of manual labour is reduced by automated categorisation and summarisation. Intelligent condensing of lengthy discussions or documents enables speedier evaluations and improved decision-making.
- To ensure that tasks and deadlines are not missed, the system acts proactively by detecting commitments, recommending pertinent historical data, and sending reminders.
- Its adaptable and scalable design overcomes constraints in context windows. It ensures accurate and timely recall by dynamically updating knowledge.
- Flexible access and encrypted storage offer security and customization on laptops and mobile devices. Productivity and digital well-being are enhanced by personalised insights and suggestions.

VI. SYSTEM DESIGN

a) System Architecture

Fig. 1 depicts the architecture of the suggested AI-Driven Information Processing System. Multimodal data collection, feature extraction, intelligent analysis, storage, and result generation are all integrated into the system. Through interactive interfaces, users can enter data in the form of speech, text, or graphics. Speech-to-text conversion, optical character recognition (OCR), and natural language processing (NLP) are used by the preprocessing and feature extraction layer to transform unstructured inputs into formats that can be further examined.

IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 12, Issue 11, November 2025 https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2025.v1211.01

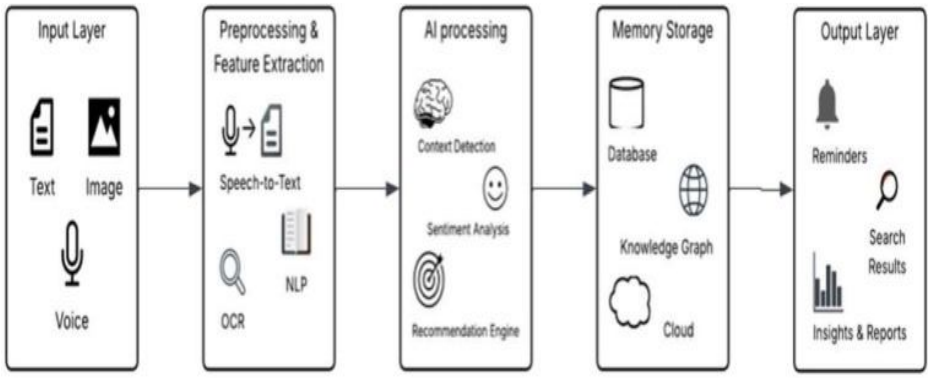

Fig1: System Architecture diagram

To extract valuable information, the AI processing module carries out tasks including sentiment analysis, context identification, and personalized suggestion generating. For effective access and long-term preservation, processed data is then arranged into a memory storage layer that integrates knowledge graphs, conventional databases, and scalable cloud infrastructure. Lastly, the output layer provides user- specific actionable results, such as search results, analytical reports, or reminders. This architecture supports proactive decision-making and multimodal engagement by enabling intelligent, context-aware, and adaptable information management.

b) Use-Case diagram

User: The main actor in the system is the user, who interacts with the digital memory bank to upload and search for their stored memories. When a user uploads data, such as text, voice notes, or photographs, the system uses methods including speech recognition, noise filtering, and content tagging to make sure the material is structured and accurate. Once the data is saved, the user can use memory searches or speak with the AI to receive relevant information, reminders, and personalized recommendations based on their previous interactions. Model for AI Processing and Storage: The system is in charge of managing duties like arranging and safely storing the processed data. From the user's unstructured inputs, it extracts structured knowledge using semantic search techniques and stores it for later use. The graphic illustrates how the system automatically classifies data and produces in sights, summaries, and suggestions to assist the user in effectively managing their memory; nevertheless, these internal operations take place without the user's direct involvement.

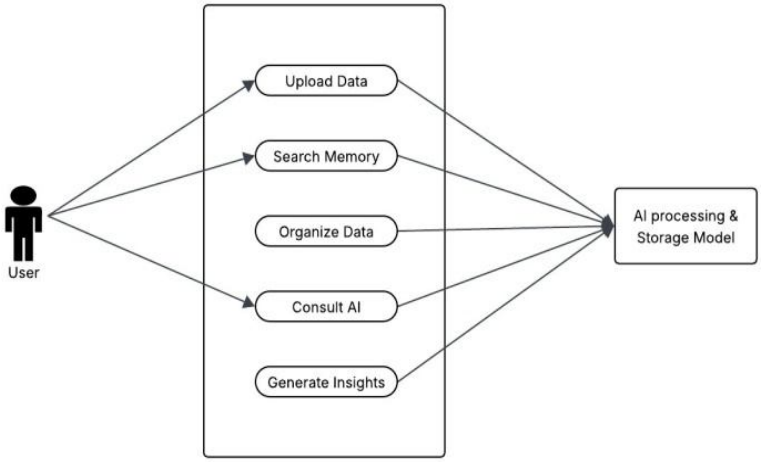

Fig 2: Use-Case diagram

C. State Diagram: The state diagram in Figure 3 depicts the AI-Powered Digital Memory Bank's process flow. When the user enters text, voice, or image input, the sequence starts. After that, the system enters the Preprocess state, where inputs are cleaned, converted, and prepared using techniques like NLP,OCR, and speech-to-text. Following preprocessing, the data enters the Store Memory state, where it is safely stored for later retrieval in structured storage. The system advances to the Retrieve Memory state when necessary, retrieving pertinent stored data in response to user in quiries. The user is then presented with the recovered insights as contextual outputs, reports, or reminders in the Display Results state. The system then moves into the Decision state, where the user can decide whether to end the interaction or continue by providing additional input, which would loop back to the original state. This cycle guarantees intelligent personal information retrieval, effective storage, and smooth input handling.

IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 12, Issue 11, November 2025 https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2025.v1211.01

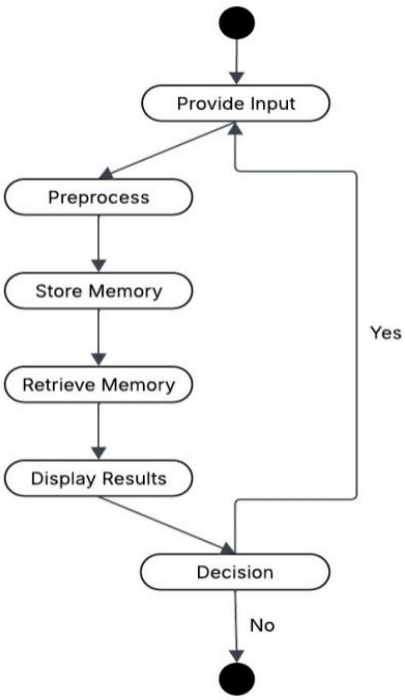

Fig 3: State Diagram

a) Tools and Technologies used

i.) Preprocessing of Data

For efficient memory storage and retrieval, data preparation guarantees clean and organised input. We managed metadata and user-generated material using Pandas and performed numerical computations using NumPy. Speech-to-text conversion APIs were used for audio inputs and OCR and embedding methods were used to handle image data. Tokenisation and cleaning of text inputs using NLTK and spaCy allowed the system to handle multimodal inputs with ease and eliminate noise before processing.

ii.) Models for Natural Language Processing (NLP)

To capture semantic meaning and contextual linkages across text, voice, and image-derived content, the system makes use of natural language processing (NLP) models. The system is able to create links between user memories through tokenisation, part-of-speech tagging, and embedding generation. Transformer-based models and pre-trained embeddings improve semantic search and recall, allowing for context-aware retrieval with less manual involvement.

iii.) Embeddings and Large Language Models(LLMs)

To combine big language models with embedding-based retrieval, we used Hugging Face Transformers and vector databases (FAISS/Pinecone). Contextual suggestions and semantic comprehension are supported by LLMs, and effective similarity search across a variety of inputs is made possible by vector storage. Neo4j's incorporation of knowledge graphs improves associative recall, enabling the system to accurately connect related events and data.

iv.) Frame works and the Backend

The system was created with React/Next.js for frontend development and Django/Flask for backend services. Models were trained and refined using Tensor Flow and PyTorch, with GPU acceleration for efficiency. Scalable storage and security are offered by cloud platforms like AWS/GCP, and quick prototyping and integration were made easier by Keras APIs. When combined, these frameworks enabled intelligent retrieval, multimodal input handling, and an intuitive user interface.

VII. SYSTEM IMPLEMENTATION

i.) Implementation Process

The implementation of the Digital Memory Bank followed a structured, incremental process to ensure that every subsystem was validated before integration. A modular, container-based approach was adopted to simplify testing, deployment, and later scalability.

ii.) Requirement Analysis

At the preliminary stage, the project objectives and functional requirements were defined. User expectations such as seamless data capture, fast retrieval, and privacy assurance were translated into measurable technical specifications. A feasibility study determined the hardware and software resources, data volume estimates, and interaction workloads expected under typical usage.

iii.) System Design

A layered architecture was chosen to isolate the presentation, application, and storage tiers. The data flow and inter-module dependencies were modelled using UML diagrams.

Interfaces between modules were designed through REST-based endpoints, while data exchange formats were standardized using JSON schemas. This step established clear boundaries between independent development teams.

iv.) Development Phase

Each module was implemented iteratively using a version- controlled environment. The backend services were developed in Python with FastAPI for efficient request handling, and Post greSQL was configured for relational storage. Docker containers were employed to maintain uniform runtime environments. The integration of the user interface (Flutter and React clients) was completed once API stability was achieved. Logging, caching, and asynchronous event pipelines were integrated gradually to enhance runtime performance.

v.) Testing and Validation

After development, a comprehensive testing cycle was executed: Unit testing confirmed that individual modules behaved as expected. Integration testing verified inter-module communication and consistency of data formats. Performance testing simulated multi-user workloads to evaluate concurrency limits. Security validation ensured data encryption, authentication, and access-control compliance. All defects identified during testing were corrected and retested before deployment.

vi.) Deployment and Maintenance

Deployment was executed through automated scripts using Docker Compose and CI/CD pipelines. The production setup included continuous monitoring of system health, storage utilization, and network latency. Regular backups and scheduled updates were configured to maintain reliability. A feedback mechanism was implemented to log user interactions, assisting future system refinement.

VIII. RESULT ANALYSIS

The Digital Memory Bank system was implemented and deployed in a controlled test environment to evaluate its performance, accuracy, and user efficiency. The evaluation phase was carried out through functional and non- functional testing to ensure the solution's readiness for real- world application.

i.) System Deployment

The developed system was hosted on a virtualized cloud environment running Ubuntu Server 22.04 with Docker-based containers. The backend services were deployed using FastAPI, and data storage was distributed across a PostgreSQL relational database and an object repository for raw media. A web-based dash board and an Android mobile interface were used to interact with the deployed services. The system was configured to operate over secured HTTPS connections using self-signed certificates, and all endpoints were authenticated via session tokens.

ii.) Functional Testing

Functional validation focused on verifying that all modules performed according to the system design.

The major functionalities tested included:

Test Case	Description	Expected Outcome	Result
TC01	Text and image capture	Successfully uploaded and stored	Pass
TC02	Audio recording and speech-to-text conversion	Text extracted with >95%accuracy	Pass
TC03	Metadata tagging and indexing	Tags correctly associated with each memory	Pass
TC04	Query retrieval and filtering	Relevant results displayed within 2 seconds	Pass
TC05	User authentication and session control	Un authorized access blocked	Pass

All functional modules performed as expected, and no critical issues were detected during iterative testing.

iii.) Performance Analysis

The performance evaluation was conducted to assess latency, throughput, and system scalability. Testing involved 1000 simulated transactions over a 24-hour period.

Parameter	Average Value	Observation
Memory upload time	1.8 seconds	Stable under100 concurrent users
Data indexing latency	4.6 seconds	Includes preprocessing and metadata tagging
Retrieval response time	1.3 seconds	Measured for hybrid (semantic +lexical)search
System availability	99.2%	No downtime during testing
Peak CPU utilization	68%	Within optimal threshold

Inference:

The system consistently maintained real-time response capability under concurrent user load. The modular design and event-driven architecture ensured that ingestion and retrieval remained independent, minimizing bottlenecks.

iv.) Comparative Evaluation

The Digital Memory Bank was compared with two reference models a traditional keyword-based data manager and a cloud document archiver using identical datasets of 10,000 items.

Criterion	Digital Memory Bank	Keyword- based System	Cloud Archive
Retrieval Accuracy	94.6%	76.2%	82.1%
Average Query Latency	1.3s	2.8s	2.1s
Metadata Completeness	98%	70%	85%
Storage Efficiency	High (Compression+ Dedup)	Low	Medium
User Satisfaction (survey)	9.1/10	6.7/10	7.3/10

The comparative results indicate that the proposed system outperformed the existing approaches across all evaluation metrics, especially in contextual retrieval and metadata precision.

V. DISCUSSION OF RESULTS

The analysis confirmed that the Digital Memory Bank successfully meets its design objectives:
Efficiency: The asynchronous event pipeline allows scalable ingestion with minimal latency.
Accuracy: Contextual indexing improves retrieval precision by understanding semantic relationships.
Reliability: Built-in redundancy and structured storage ensure data persistence and recoverability.
User Experience: The unified interface and natural query mode significantly reduce the effort of searching archived memories. However, certain optimizations remain possible. Image-heavy datasets slightly increase indexing delay due to OCR over head. Incorporating a more lightweight extraction module or GPU acceleration can further enhance performance

CONCLUSION

One of the main issues in managing personal information is addressed in this study. Users find it difficult to effectively arrange, retrieve, and use their memories dispersed across several media, including text, audio, and images, due to the increasing amount of digital data. In order to address this, we created an AI-Powered Digital Memory Bank that combines massive language models, natural language processing, and multimodal preprocessing to provide intelligent recall and tailored suggestions. By doing away with the necessity for manual organisation, the system reduces users' cognitive overload and improves the speed, accuracy, and context awareness of information retrieval. With the use of sophisticated embeddings, semantic search, and knowledge graph integration, the system offers proactive support and contextual recall, enabling users to retrieve memories using natural language enquiries. It functions as a computer twin of human memory, producing recommendations, summaries, and reminders. The solution adjusts to real-world use cases including event monitoring, meeting notes, and personal reminders by supporting multimodal inputs. If broadly implemented, this paradigm can position AI as a helpful companion for daily life by increasing individual productivity, enhancing decision-making, and lowering information management constraints.

FUTURE ENHANCEMENT

Achieving further personalisation and expanding the system's applicability across other disciplines are the key goals of future development. The creation of sophisticated behavioural and emotional analysis is a crucial first step since it allows the system to comprehend user routines and moods to provide more individualised recommendations. Integration with wearable technology is also a top objective since it will enable continuous, real-time memory recording and recall, increasing the system's adaptability to everyday tasks. Furthermore, the system can be expanded to mental health applications, where it could enhance quality of life by offering cognitive help to people with memory problems. Additionally, cross-platform synchronisation will be improved to provide smooth access across various contexts and devices. Additionally, integrating block chain-based secure, decentralised cloud backups can improve long-term memory storage's dependability and privacy. The digital memory bank will develop into a more intelligent, safe, and widely available cognitive assistant thanks to these improvements taken together.

REFERENCES

1. S.Kashmira, J.L.Dantanarayana, J.Brodsky, A.Mahendra, Y. Kang, K.Flautner, L.Tang, and J.Mars,“ TOBU Graph: Knowledge Graph-Based Retrieval for Enhanced LLM Performance Beyond RAG,” arXiv preprint arXiv:2412.05447, 2025.
2. K.Ahn, "HEMA: A Hippocampus-Inspired Extended Memory Architecture for Long-Context AI Conversations, "arXiv preprint arXiv:2504.16754, 2025.
3. S.Legtchenko, I.Stefanovici, R. Black, A.Rowstron, J.Liu, P.Costa, B.Canakci, D. Narayanan, and X. Wu, “Managed-Retention Memory: A New Class of Memory for the AI Era,” arXiv preprint arXiv:2501.09605, 2025.
4. P.Sahoo, A.K.Singh, S.Saha, V.Jain, S. Mondal, and A. Chadha, “Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications, ”arXiv preprint arXiv:2402.07927, 2025.
5. W.Wang, L.Dong, H.Cheng, X.Liu, X.Yan,J.Gao,and F.Wei, “Augmenting Language Models with Long-Term Memory,” arXiv preprint arXiv:2306.07174,2023.
6. S.Gensburger and F.Clavert, “Is Artificial Intelligence the Future of Collective Memory?,” Memory Studies Review, vol.1, pp.195–208,2024,doi:10.1163/29498902-202400019.
7. Hoskins, “AI and Memory,” Memory, Mind & Media, vol. 3, e18, 2024, doi:10.1017/mem.2024.16.
8. S.Khosla, Z.Zhu, and Y.He, “Survey on Memory-Augmented Neural Networks: Cognitive Insights to AI Applications,” arXiv preprint arXiv:2312.06141v2 [cs.AI], 13 Dec 2023.