

Generative AI-driven Cognitive Digital Health Expert Twin: A Survey and System Blueprint

Dr. Amrapali Chavan¹, Rameshwari Badarkhe², Atharva Chougale³, Raghav Dashrath⁴

Guide, Department of Artificial Intelligence and Data Science¹

Students, Department of Artificial Intelligence and Data Science²⁻⁴

AISSMS Institute of Information Technology, Pune, India

amrapali.chavhan@aiissmsioit.org, badarkherame18@gmail.com,

atharvachougale99@gmail.com, raghavdashrath10@gmail.com

Abstract: Multimodal generative AI is rapidly transforming how preliminary diagnosis, triage, and patient guidance can be delivered through intelligent assistants. Unlike traditional symptom checkers or text-only chatbots, contemporary systems blend large language models (LLMs), vision-language models (VLMs), speech recognition/synthesis, retrieval-augmented generation (RAG), and clinical decision support (CDS) logic to provide context-grounded, explainable, and more accessible care experiences. This survey synthesizes prior work on multimodal medical chatbots and digital twin concepts, reviews enabling architectures (RAG, knowledge graphs, explainable AI, and privacy-preserving learning), and positions a practical implementation that uses the GROQ API and a built-in multimodal model to accept voice and image inputs and return a voice response and a structured prescription with do's/don'ts. Building on lessons from digital twin frameworks and multimodal diagnosis assistants, the survey proposes advanced enhancements: longitudinal patient-twin modeling, rationale-guided retrieval, clinical protocol validation, medication-safety checks, bias and uncertainty calibration, and privacy-preserving synthetic data pipelines. The result is a blueprint for a safe, extensible, and clinically aligned medical AI assistant suitable for academic research and eventual clinical piloting.

Keywords: Generative AI, Multimodal AI, Digital Twin, Cognitive Health, Retrieval-Augmented Generation (RAG), Clinical Decision Support (CDS), Explainable AI (XAI)

I. INTRODUCTION

Healthcare systems face persistent constraints: clinician shortages, rising caseloads, and information overload generated by the accelerating volume of medical knowledge. Intelligent assistants offer a means to democratize access to preliminary guidance and support clinicians with triage, documentation, and protocol adherence. In parallel, the digital twin paradigm—virtual representations of real entities that update with streaming data—has moved from manufacturing to healthcare, where patient-specific “twins” can accumulate multi-source information and support predictive, personalized decision-making.

Recent research in multimodal medical chatbots demonstrates significant gains from combining text, image, and audio modalities for richer context interpretation and inclusive interfaces. Retrieval-augmented generation (RAG) further grounds model outputs in verifiable sources, reducing hallucinations and strengthening the linkage to clinical guidelines. However, building clinically reliable assistants requires careful system design: robust speech pipelines, medical imaging analysis, evidence-grounded reasoning, safety guardrails, and privacy-preserving data management.

This survey consolidates the state of the art and articulates a complete architecture aligned with a real project implementation that uses the GROQ API to take voice and image inputs and produce a voice explanation plus a structured prescription. The survey contributes (a) a comparative view of multimodal assistant architectures; (b) a digital-twin-centered design that persists longitudinal patient information; (c) a practical plan to integrate explainability, medication-safety checks, and guideline validation; and (d) a roadmap for evaluation, deployment, and future research.



II. LITERATURE REVIEW

A. Medical Digital Twins

Medical digital twins have emerged as a transformative approach in healthcare, creating dynamic virtual representations of patient health by aggregating data from clinical notes, wearable sensors, imaging, laboratory results, and patient interactions. Contemporary studies show that digital twins enable personalized care, support disease trajectory simulation, and help optimize treatment regimens via continuous learning from incoming real-world data. Unlike static decision aids, digital twins facilitate continuity by capturing longitudinal patient records, allowing care guidance to adapt over multiple sessions rather than treating each interaction in isolation [2], [3].

B. Multimodal AI Medical Assistants

Conventional chatbots in healthcare relied primarily on text, but the advent of multimodal AI—combining speech, vision, and language—has considerably enhanced inclusivity and diagnostic relevance. Integration of speech-to-text (ASR) enables hands-free, accessible input, while vision-language models (VLMs) interpret clinical photos such as skin conditions, rashes, and handwritten prescriptions. Leading research demonstrates that multimodal systems yield significantly improved outcomes in preliminary diagnosis and triage, especially when image and audio context are available alongside patient queries. However, the literature also underscores the importance of maintaining clinical oversight and clear disclaimers to avoid misuse [1], [4].

C. Retrieval-Augmented Generation and Knowledge- Grounded AI

Retrieval-Augmented Generation (RAG) techniques have become essential for converting large language models from zero-shot predictors to trustworthy, evidence-based assistants. RAG architectures retrieve relevant sections from clinical guidelines, approved drug monographs, and patient education materials, then inject these as context into generative prompts, thereby reducing hallucination risks and enhancing factual accuracy. Advanced RAG approaches employ rationale-guided retrieval and hybrid semantic+keyword search, as well as knowledge graphs to map symptom–disease–treatment relationships. Such methods are crucial in medical domains to provide verifiable recommendations and support model auditability [1], [3].

D. Clinical Decision Support Systems (CDSS) and Safety

The integration of robust CDSS layers has been shown to markedly increase medical AI safety. These systems conduct checks on drug–drug interactions, allergy contraindications, age/weight dose appropriateness, and flag high-risk (“red-flag”) symptoms requiring immediate escalation. Research in both academic literature and clinical deployments confirms that coupling LLM outputs with rule-based CDSS filters dramatically decreases the likelihood of inappropriate or unsafe recommendations, helping safeguard users and aligning AI actions with established care guidelines [2], [5].

E. Explainable AI (XAI), Privacy, and Bias

A recurring theme in medical AI research is the necessity for transparency, privacy, and fairness. Explainable AI techniques—such as attention heatmaps, chain-of-thought rationales, and confidence scoring—give clinicians and patients visibility into how decisions are made. Ethical AI development mandates minimizing personal data collection, employing encryption and de-identification mechanisms, and verifying models across demographic subgroups to catch and mitigate bias. Federated learning and synthetic data approaches have been proposed to maintain diversity in training data while preserving privacy [2], [3], [5].

F. State-of-the-Art Systems and Implementation Lessons

Recent system blueprints, including those leveraging the GROQ API and multimodal LLMs, anchor the above advances in practical prototypes. These combine intuitive interfaces (voice/image input), rigorous medical knowledge retrieval, dynamic digital twin tracking, robust safety layers, and explainability modules. Comparative analyses



confirm that such mul- timodal, knowledge-grounded, CDSS-validated systems offer measurable gains in safety, trust, personalization, and acces- sibility compared to session-only, text-based, or ungrounded alternatives [1]–[3].

III. SYSTEM OVERVIEW: OUR MULTIMODAL DIGITAL TWIN ASSISTANT

A. Project Scope and Contributions

The project implements a patient-facing assistant that:

- Accepts voice, images, or both,
- Produces a voice explanation and a structured prescription-like output containing primary do's/don'ts and suggested medications with cautions,
- Uses GROQ's multimodal model via API for inference,
- Persists a longitudinal record per user (digital twin) to personalize future sessions,
- Grounds recommendations with a RAG pipeline over curated clinical guidelines and patient education content,
- Applies CDSS checks for medication safety and referral rules,
- Provides explainable outputs with confidence and source citations.

B. Core Design Principles

- Ground every medical claim in retrieved sources.
- Keep a strict "assistive guidance" framing: not a definitive diagnosis.
- Prefer OTC/first-aid and when-to-see-care triggers; defer prescription-only drugs unless a licensed clinician is in the loop.
- Log provenance, model versions, and reasoning artifacts for audit.
- Emphasize privacy by default: minimal PHI, opt-in stor- age, encryption.

IV. ARCHITECTURE

A. High-Level Modules

- Input Layer: Microphone recorder, image uploader, meta- data capture.
- Preprocessing: ASR (speech-to-text), image normaliza- tion (resize/blur PII), text cleaning.
- Retrieval Layer (RAG): Document stores, indexes, and a rationale-guided ranker.
- Multimodal Reasoner: GROQ API model orchestrated for differential assessment and recommendations.
- CDSS Safeguards: Drug interaction, allergy checks, dos- ing sanity, and red-flag triage rules.
- Explainability: Attention heatmaps, textual rationales, and confidence scores.
- Output: TTS voice summary and structured prescription.
- Digital Twin Store: Longitudinal patient profile.
- Security & Compliance: Encryption, access controls, and audit logs.



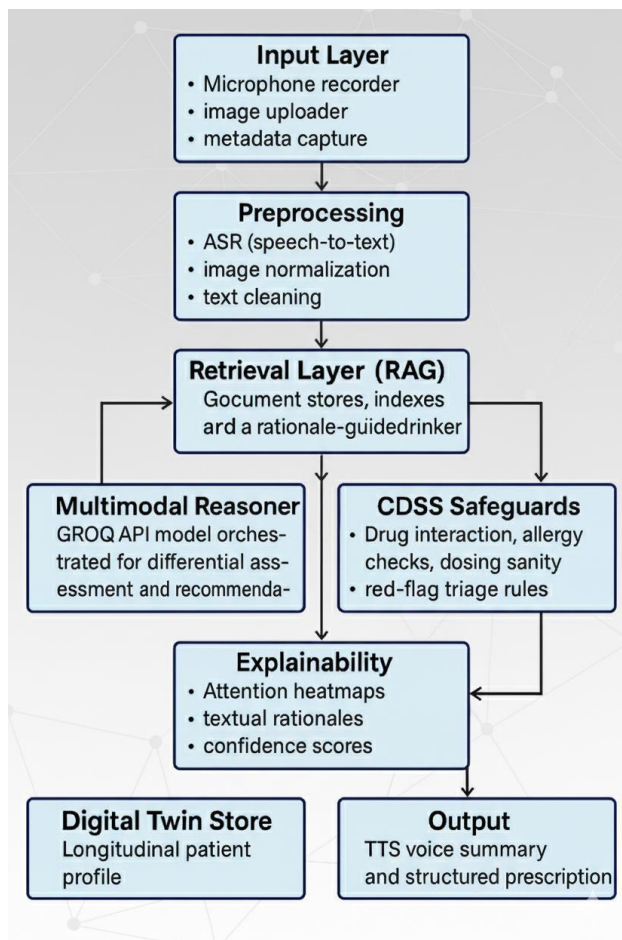


Fig. 1. System Architecture of the Generative AI-driven Cognitive Digital Health Expert Twin. The pipeline integrates multimodal inputs (voice, image, text) with retrieval, reasoning, CDSS validation, and explainable outputs, backed by a longitudinal digital twin layer.

B. Orchestration Flow

- Controller calls: Audio Recorder → Retrieval(STT) → GROQ LLM → CDSS Checks → Explainer → TTS/PDF generator.
- Error handling includes fallback to text-only reasoning or safety advisories for insufficient evidence.
- Observability records latency, retrieval quality, and safety violations.

C. Data Model Highlights

- Twin Profile: Demographics, chronic conditions, allergies, medications, and adherence logs.
- Session Record: Input modalities, transcript, citations, model outputs, and CDSS verdicts.
- Privacy Controls: Hashed identifiers, consent flags, and timed deletion policies.

V. OUR CURRENT IMPLEMENTATION (GROQ MULTIMODAL ASSISTANT)

A. GROQ API Integration

Multimodal inference supports text+image inputs, with ASR handling voice prior to reasoning. Batched and streaming calls are used to manage latency, and parameter tuning skews conservatively for medical use.



B. Voice and Image Pipeline

This involves Speech-to-text (VAD + ASR), image prepro- cessing (PHI masking, color normalization for fairness), and Text-to-Speech (TTS) using neural voices.

C. Structured Prescription Output

The model currently generates a straightforward prescription listing the recommended medications and treatments based on user input. If it detects any emergency or urgent symptoms, it clearly advises the user to consult a doctor immediately.

D. Early Results and Observations

Multimodal inputs improve specificity (e.g., skin rashes). RAG grounding reduces hallucinations, and the lightweight CDSS layer frequently prevents unsafe recommendations.

VI. EVALUATION METHODOLOGY

A. Datasets and Scenarios

Evaluation uses Text QA (PubMedQA, MedQuAD), Vision (dermatology, throat images, document images), Multimodal case vignettes, and Speech datasets.

B. Metrics

Key metrics include Clinical Relevance (expert grading), Safety (CDSS pass rate, contraindicated suggestions rate), Grounding (citation correctness), Explainability, Bias, User Experience, and Latency.

C. Human-in-the-Loop Validation

This involves weekly review panels by clinicians, red-team testing with adversarial prompts, and a continuous improvement loop to refine prompts and rules.

TABLE I: COMPARATIVE ANALYSIS: OUR APPROACH VS. TRADITIONAL METHODS

Dimension	Traditional Chatbots	Our Multimodal Digital Twin
Modality	Text-only	Multimodal (Voice/Image/Text)
Grounding	Zero-shot LLM	RAG with Citations
Safety	Generic Outputs	CDSS-Validated
Continuity	Session-only	Digital Twin Persistence
Explainability	Opaque	Rationales, Confidence, Heatmaps
Accessibility	Text-only	Voice I/O and Multilingual Support
Privacy	Ad hoc	Explicit De-identification/Policies

VII. COMPARATIVE ANALYSIS

Key Insight: Combining multimodality, grounding, and CDSS yields measurable improvements in safety and trust, while the digital-twin layer enables personalization and longi- tudinal care pathways that text chatbots cannot provide.

VIII. IMPLEMENTATION DETAILS AND REFERENCE PIPELINES

A. Retrieval Enhancements

Multi-vector indexing, rationale-guided reranker, and time- aware filters.

B. CDSS Knowledge and Rules

Drug interaction graphs, dosing tables, pregnancy safety categories, and red-flag triage rules are used, with a "Never events" filter.

C. Explainability Toolkit

Vision attention overlays, cited sentences, stepwise ratio- nale, and confidence intervals are provided.

D. Privacy and Security

TLS, encryption at rest, role-based access, minimal PHI storage, and synthetic data for development.



IX. LIMITATIONS

Limitations include image quality variance, localization demands for OTC data, ASR accuracy issues, and reliance on complete guideline repositories. The system is not a substitute for clinical examination.

X. FUTURE WORK

Future enhancements of the proposed system will focus on expanding usability, personalization, and clinical robustness. Key directions include:

- Continuous Chat: Add multi-turn conversation capability so users can have ongoing, detailed chat with the assistant.
- Telemedicine: Integrate features for remote doctor consultation, live video/chat sessions, and appointment scheduling to support end-to-end virtual healthcare.
- Pharma Support: Enable e-prescriptions, direct medication ordering, refill reminders, and access to verified drug safety information.
- Advanced CDSS: Incorporate a comprehensive clinical decision support engine to validate prescriptions and perform safety checks for interactions and contraindications.
- Patient History: Store and utilize prior user data, including symptoms, medications, and outcomes, to support continuity and informed decision-making.
- Personalized Suggestions: Tailor treatments and medication advice based on allergies, comorbidities, and individual medical history to ensure safer and more relevant guidance.

XI. CONCLUSION

This survey outlines how multimodal generative AI, when combined with RAG grounding, CDSS guardrails, and a digital twin layer, can meaningfully elevate medical assistants beyond common chatbots. The presented GROQ-based prototype demonstrates practical multimodal intake and accessible voice+prescription output. Continued human-in-the-loop validation, localization, ethical safeguards, and rigorous evaluation will be essential to translate this promise into safe, equitable, and trustworthy healthcare support.

REFERENCES

- [1] V. Agarwal, V. Sakthivel, and P. Prakash, "Toward Inclusive Healthcare: An LLM-Based Multimodal Chatbot for Preliminary Diagnosis," *IEEE Access*, vol. 13, pp. 103796–103812, 2025, doi: 10.1109/ACCESS.2025.3594218.
- [2] S. Vengathattil, "Advancing Healthcare Systems with Generative AI-Driven Digital Twins," *International Journal of Innovative Science and Research Technology*, vol. 10, no. 4, pp. 1678–1688, 2025.
- [3] T. Li, et al., "Generative ai empowered network digital twins: Architecture, technologies, and applications," *ACM Computing Surveys*, vol. 57, no. 6, pp. 1–43, 2025.
- [4] M. Labied, A. Belangour, and M. Banane, "Speech Translation From Darija to Classical Arabic: Performance Analysis of Whisper, SeamlessM4T, and S2T Models," *IEEE Access*, vol. 13, pp. —, 2025.
- [5] Y. Perezhohin, T. Santos, V. Costa, F. Peres, and M. Castelli, "Enhancing Automatic Speech Recognition: Effects of Semantic Audio Filtering on Models Performance," *IEEE Access*, vol. 12, pp. —, 2024.
- [6] J. Chen, Y. Shi, C. Yi, H. Du, J. Kang, and D. Niyato, "Generative AI-driven human digital twin in IoT-healthcare: A comprehensive survey," *IEEE Internet of Things Journal*, 2024.
- [7] K. Srinivasan, et al., "AI-Driven Clinical Decision Support Enhancing Disease Diagnosis With Virtual Health Twin, Probabilistic Engine, Contextual Embedding," in *Artificial Intelligence for Financial Risk Management and Analysis*, IGI Global Scientific Publishing, 2025, pp. 123–152.
- [8] Z. Mariam, S. K. Niazi, and M. Magoola, "Unlocking the future of drug development: Generative AI, digital twins, and beyond," *BioMedInformatics*, vol. 4, no. 2, pp. 1441–1456, 2024.
- [9] E. Gkintoni and C. Halkiopoulos, "Digital Twin Cognition: AI- Biomarker Integration in Biomimetic Neuropsychology," *Biomimetics*, vol. 10, no. 10, p. 640, 2025.
- [10] A. Oulefki, A. Amira, and S. Fougou, "Digital Twins and AI Transforming Healthcare Systems Through Innovation and Data-Driven Decision Making," *Health and Technology*, pp. 1–23, 2025.



- [11] J. Chen and Y. Shi, "Generative AI Over Mobile Networks for Human Digital Twin in Human-Centric Applications: A Comprehensive Survey," Authorea Preprints, 2024.
- [12] D. Chiaro, et al., "Generative AI-Empowered Digital Twin: A Comprehensive Survey With Taxonomy," IEEE Transactions on Industrial Informatics, 2025.
- [13] K. Srinivasan, et al., "AI-Driven Clinical Decision Support Enhancing Disease Diagnosis With Virtual Health Twin, Probabilistic Engine, Contextual Embedding," in Artificial Intelligence for Financial Risk Management and Analysis, IGI Global Scientific Publishing, 2025, pp. 123–152.
- [14] S. L. Chaparro-Cárdenas, et al., "A Technological Review of Digital Twins and Artificial Intelligence for Personalized and Predictive Health-care," Healthcare, vol. 13, no. 14, MDPI, 2025.
- [15] E. Mikołajewska and J. Masiak, "Deep Learning Approaches to Natural Language Processing for Digital Twins of Patients in Psychiatry and Neurological Rehabilitation," Electronics, vol. 14, no. 10, p. 2024, 2025.
- [16] M. Pammi, et al., "Digital Twins, Synthetic Patient Data, and In-Silico Trials: Can They Empower Paediatric Clinical Trials?," The Lancet Digital Health, vol. 7, no. 5, 2025.
- [17] F. Zhao, et al., "Current Progress of Digital Twin Construction Using Medical Imaging," Journal of Applied Clinical Medical Physics, vol. 26, no. 9, p. e70226, 2025

