MediMate AI: An AI Assistant for General Practitioners

Jan Saro[1], Jana Mazancová[2] and Helena Brožová[1]

[1]Department of Systems Engineering, Faculty of Economics and Management, Czech University of Life Sciences Prague, Czech Republic

[2]Department of Sustainable Technologies, Faculty of Tropical AgriSciences, Czech University of Life Sciences Prague, Czech Republic

saroj@pef.czu.cz (corresponding author)
mazan@ftz.czu.cz
brozova@pef.czu.cz

Abstract: Primary care remains one of the most demanding areas of medicine, with general practitioners (GPs) facing rising consultation volumes, complex patient presentations, and increasing administrative tasks. These pressures contribute to cognitive overload, diagnostic delays, and burnout. Large language models (LLMs), such as GPT-4, show potential to reduce documentation workload, enhance diagnostic reasoning, and improve overall workflows. However, their integration must be carefully managed to ensure data protection, patient safety, compliance with clinical guidelines (e.g., WHO, CDC), and ethical standards. This paper introduces MediMate AI, a GPT-4–powered prototype assistant designed to support GPs in real time. Implemented as a web-based, mobile-accessible platform, the system integrates multimodal inputs—including speech, text, and images—to transcribe consultations, extract key symptoms, and generate structured summaries. Beyond documentation, MediMate AI provides differential diagnoses with confirmatory test recommendations, evaluates geographic epidemiological risks, and produces tailored hospital routing plans. The prototype was tested in a digital innovation incubator using synthetic patient records and simulated consultations. This approach enabled safe experimentation without breaching patient confidentiality while providing early insights into feasibility, usability, and workflow integration. By combining transcripts, symptom extraction, dermatology image data, and standardized checklists, the prototype reflects the heterogeneous nature of real-world primary care. Results indicate that MediMate AI can reduce documentation workload by an estimated 25–30%, deliver clinically coherent summaries, and generate guideline-aligned differential diagnoses with improved clarity for decision-making. Physicians testing the prototype highlighted its ability to consolidate fragmented data streams, improve continuity of care, and enhance patient–clinician communication. While not intended to replace medical expertise, MediMate AI demonstrates the promise of generative AI to augment decision-making, improve efficiency, and support more patient-centred care. Future work will include prospective clinical validation, integration with electronic health record systems, and the introduction of explainability and bias detection modules. Addressing these aspects will be essential to ensure safe, ethical, and sustainable deployment in healthcare environments.

Keywords: Artificial intelligence, Healthcare, General practitioners, GPT-4, Diagnostic support, Patient triage

1. Introduction

Primary care is the cornerstone of modern healthcare but remains under constant pressure due to rising consultation volumes, complex patient needs, and increasing administrative tasks. General practitioners (GPs) must simultaneously take patient histories, evaluate risks, and maintain thorough documentation while ensuring patient-centred communication, often within a limited timeframe. This workload contributes significantly to cognitive overload and burnout (You *et al.*, 2025). Artificial intelligence (AI) offers a potential solution to these challenges. Ambient documentation systems have already been shown to reduce the burden of note-taking and improve clinician well-being (You *et al.*, 2025). More recently, large language models (LLMs), such as GPT-4, have been applied in clinical reasoning and workflow automation. Randomized controlled trials demonstrated that GPT-4 can improve physician performance in management reasoning tasks, especially among less experienced clinicians (Goh *et al.*, 2025). Similarly, GPT-4 has been shown to provide accurate diagnostic support across diverse clinical scenarios (Noda *et al.*, 2025). Despite these advances, ethical concerns remain. Research highlights that GPT-4 may produce outputs inconsistent with medical ethical standards, underscoring the need for human oversight and transparency (Xiong *et al.*, 2025). Scholars have also raised concerns that uncritical reliance on AI may erode clinical expertise or reinforce systemic biases (Pavuluri *et al.*, 2024). Within general practice, generative AI has been identified as a tool capable of integrating multimodal data and offering more personalized diagnostic and therapeutic reasoning (Geersing *et al.*, 2025). Evaluations of GPT-4 as a clinical support tool highlight its potential to augment performance, but also report non-negligible error rates of approximately 20% (Fabre *et al.*, 2024). Beyond direct diagnostic reasoning, AI-driven systems are transforming documentation and workflows, leading to reduced burnout and improved care efficiency (Bongurala *et al.*, 2024). Importantly, early adoption is already underway. A recent survey revealed that one in five GPs are using generative AI tools such as ChatGPT for documentation, diagnosis suggestions, and treatment recommendations (Kharko *et al.*, 2025). Decision support studies further indicate that AI assistance can improve diagnostic accuracy without amplifying demographic biases (Goh *et al.*, 2025). Taken together, these findings suggest that AI—and LLMs in particular—have significant potential to reduce workload, support decision-making, and enhance patient communication in primary care. However, their integration must be carefully managed to ensure transparency, safety, and clinical reliability. Against this background, we present MediMate AI, a GPT-4–powered assistant designed to transform patient voice, text, and image input into structured clinical outputs including summaries, diagnostic insights, urgency scores, and hospital routing plans.

2. Data and Methods

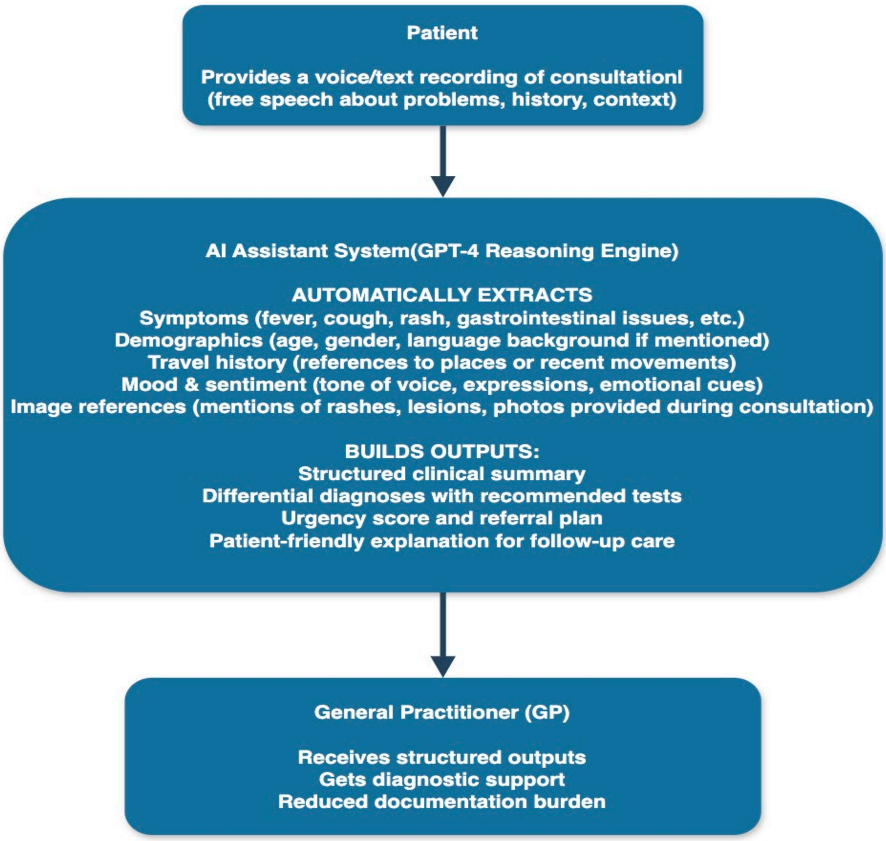

Figure 1: System Workflow: From Patient Input to GP Output

At this stage of development, no large-scale clinical datasets were available. Instead, the prototype was validated using synthetic patient records and mock consultation dialogues carefully designed to simulate realistic encounters in general practice. This approach enabled safe testing without breaching patient confidentiality and ensured early evaluation of technical feasibility and workflow integration. The records used for prototyping included:

- **Patient demographics:** age, gender, language background.
- **Symptom descriptions:** fever, cough, dermatological conditions, gastrointestinal complaints.
- **Audio recordings:** simulated consultations in English.
- **Images: publicly** available dermatological datasets (skin lesions, rashes) used for integration testing.
- **Clinical guidelines:** open-source WHO and CDC symptom checklists and differential diagnosis rules.

By combining these different modalities, the system could be tested on a diverse set of inputs, including voice, text, and image data, reflecting the heterogeneous nature of primary care practice. The emphasis of this stage was not on clinical validation, but on exploring technical feasibility, multimodal data integration, and workflow support for general practitioners.

2.1 Functional Requirements

Based on current literature, a set of **functional requirements** was defined as essential for any AI assistant in primary care. These requirements are summarized in Table 1.

Table 1: Functional requirements for MediMate AI prototype

Functionality	Rationale (literature)	Expected Benefit
Automatic transcription and summarization of consultations	Ambient documentation reduces burden and burnout	Saves time, reduces administrative workload
Symptom extraction and mapping to clinical terminology	Standardized records improve diagnostic accuracy (Geersing *et al.*, 2025)	Accelerates reasoning, ensures consistency
Differential diagnosis with test recommendations	LLMs support clinical decision-making (Noda *et al.*, 2025)	Reduces diagnostic errors, supports junior doctors
Multimodal integration (image + text)	AI improves performance in image-based specialties (Fabre *et al.*, 2024)	Early detection of dermatological/systemic conditions
Mood and sentiment tracking	Emotional monitoring is critical for continuity of care (Kharko *et al.*, 2025)	Early detection of anxiety/depression, improved communication
Geographic risk evaluation	Epidemiological context is essential for differential diagnosis (WHO/CDC)	Identification of imported infections, improved safety
Automated triage and routing	AI enhances triage consistency but requires oversight (Xiong *et al.*, 2025)	Timely referrals, improved patient safety
Structured EHR output	Seamless workflow requires EHR integration (Bongurala *et al.*, 2024)	Reduces duplicate work, ensures usability in practice

3. Results

3.1 Implemented Prototype Software Architecture

A fully functional prototype of MediMate AI was implemented and deployed for a digital innovation incubator in Croatia. The system was implemented as a web-based, mobile-accessible platform with a modular back-end powered by GPT-4. At this stage, the focus was primarily on evaluating technical feasibility, usability, and workflow integration, rather than conducting clinical trials on real patient populations. The prototype successfully integrated multiple modes of patient input, including text, speech, and images, into structured summaries and diagnostic suggestions. The patient history dashboard (Figure 1) was central to this implementation. It provided an overview of prior consultations, automatically generated summaries, AI-predicted diagnoses, and sentiment analyses. Physicians testing the prototype reported that this dashboard improved the continuity of care by consolidating previously fragmented information streams into a single interface. The underlying solution architecture (Figure 2) was implemented in Python 3.11 with a Streamlit UI for patient and physician interaction, connected to the OpenAI Web API. The architecture integrates multiple AI components: Whisper for multilingual speech-to-text transcription, GPT-3.5 for lightweight summarization tasks, and GPT-4 as the core reasoning engine for diagnosis, test recommendations, triage, and empathetic responses. Outputs are returned in structured JSON format and stored in a local encrypted SQLite database for prototype testing. This design enabled rapid prototyping, modularity, and safe experimentation while maintaining compatibility with potential future EHR integration.

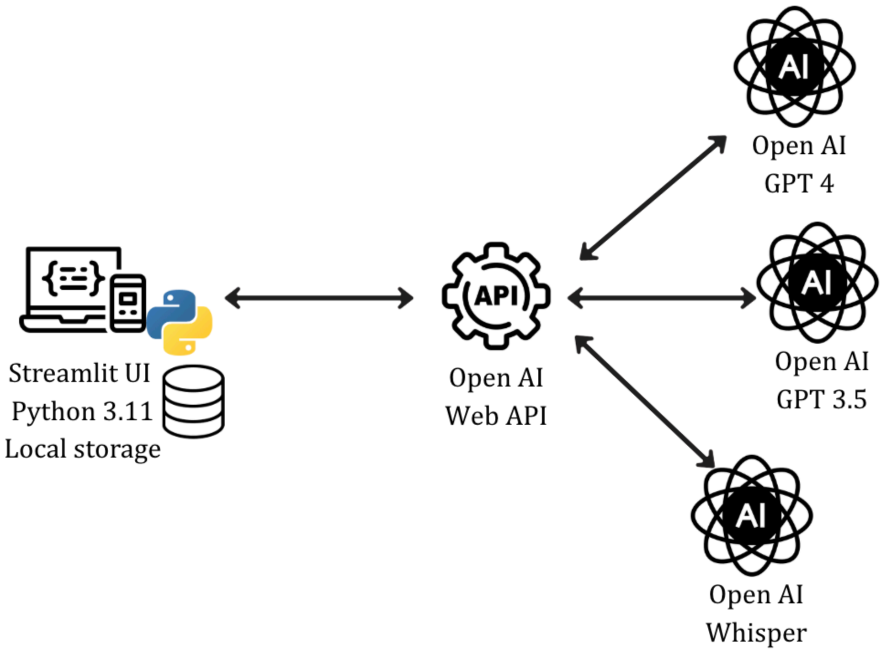

Figure 2: Application solution architecture

3.2 Prompt Design and Feature-Specific Outputs

A critical component of the MediMate AI prototype is the design of task-specific prompts, which guide the large language model to generate structured, clinically relevant outputs. Each system feature was implemented as a modular prompt template, enabling consistent and reproducible responses. Table 1 provides an overview of the prompt structures, their functional role in the application, and the corresponding outputs.

The design principle was to keep prompts concise, structured, and task-oriented, while ensuring alignment with expected clinical use cases. For example, transcription and summarization prompts were focused on reducing administrative burden, while diagnostic reasoning prompts explicitly requested probability scores, clinical guideline alignment, and confirmatory tests. This modular approach also facilitated multimodal integration, where text and image data were processed jointly through a combined reasoning prompt. Beyond medical reasoning, prompts were also designed to capture patient-centred aspects of care. Sentiment analysis prompts provided longitudinal sentiment tracking, while empathetic response prompts generated short, supportive messages in the patient's native language (Czech). These elements reflect the dual purpose of MediMate AI: both reducing clinician workload and enhancing the patient experience.

Table 2: Application Features and Prompt Structures

Feature	Prompt Structure (English)	Provided Output
Automatic transcription & summarization	"Here is the patient's spoken/audio transcript. Translate into English, summarize main points, and return JSON with structured transcript and summary."	English transcript + structured consultation summary
Symptom extraction & mapping	"From the patient transcript, extract symptoms and map them to clinical terminology. Return JSON with symptom list."	JSON list of standardized symptoms (e.g., fever, cough, rash)
Differential diagnosis + test recommendations	"Given these symptoms: […], suggest differential diagnoses, probability scores, and recommended confirmatory tests. Return JSON."	Possible diagnoses, reasoning, distinguishing features, and recommended tests

Feature	Prompt Structure (English)	Provided Output
Multimodal integration (image + text)	"Combine the provided dermatology image with the transcript symptoms. Suggest likely conditions and required tests. Return JSON."	Integrated diagnosis from text + image (e.g., rash + fever → viral exanthem)
Mood & sentiment tracking	"Analyze patient transcript for emotions (stress, anxiety, fatigue, fear). Return JSON with each emotion, score (0–1), and explanations."	Sentiment label (positive/negative), emotions with scores & explanations
Geographic risk evaluation	"From patient transcript, extract travel history and activities. Cross-reference with epidemiological risks. Return JSON with likely risks and alerts."	Regional risk factors, exposure window, disease alerts, recommendations
Automated triage & routing	"Based on symptoms, diagnoses, and urgency, generate referral plan: urgency score, department, routing steps. Return JSON."	Urgency score, routing (Emergency, Infectious Diseases, Dermatology), PDF plan
Structured EHR output	"Transform patient transcript, diagnoses, tests, and referral into structured EHR-compatible JSON object."	JSON record ready for integration with EHR (summaries, diagnoses, tests, risk factors, empathy message)
Empathetic patient response	"Generate empathetic message in Czech acknowledging patient's stress and symptoms, encouraging next steps."	Short empathetic text message in Czech

3.3 Prompt Design and Feature-Specific Outputs

To evaluate the capabilities of the proposed clinical decision support prototype, three representative case studies were developed. Each case demonstrates a distinct functionality of the system: differential diagnosis, integration of geographic epidemiological risk, and automated hospital routing.

3.3.1 Case 1 – differential diagnosis of respiratory symptoms

A patient presenting with fever (38,5 °C), severe fatigue, muscle aches, and swollen cervical lymph nodes was processed through the Differential Diagnosis Assistant. The system generated a structured summary of symptoms and proposed the following diagnoses:

- **COVID-19 (most likely)** – justified by fever, fatigue, and respiratory issues with typical progression.
- **Influenza** – considered due to abrupt onset and severe muscle aches.
- **Mononucleosis** – suggested given lymphadenopathy and prolonged fatigue, requiring blood test confirmation.

The assistant also provided distinguishing features and recommended diagnostic tests (COVID-19 PCR, rapid influenza test, CBC). This case demonstrates the system's ability to transform unstructured patient narratives into prioritized differential diagnoses.

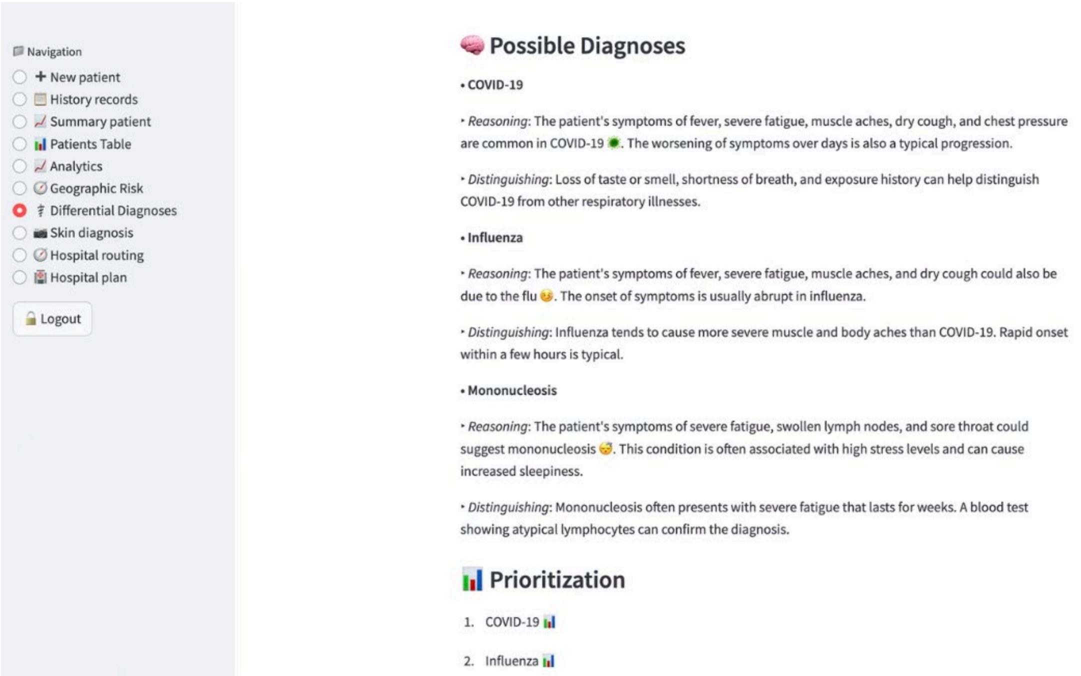

Figure 3: Differential Diagnosis Output

3.3.2 Case 2 – geographic risk and epidemiological alerting

The Geographic Risk Evaluation module was applied to the same patient. Given the high incidence of COVID-19 epidemiological data from Croatia, where a high incidence of COVID-19 was recorded at that time and flagged a "Top Alert – High Epidemiological Risk." The alert was based on:

- patient symptoms consistent with viral infection
- symptom progression aligning with the COVID-19 incubation period (2–14 days)
- regional epidemiological data.

The recommendation included immediate testing, patient isolation, and close monitoring. This case highlights the integration of personal health records with real-time geographic epidemiological risks to support context-aware decision-making.

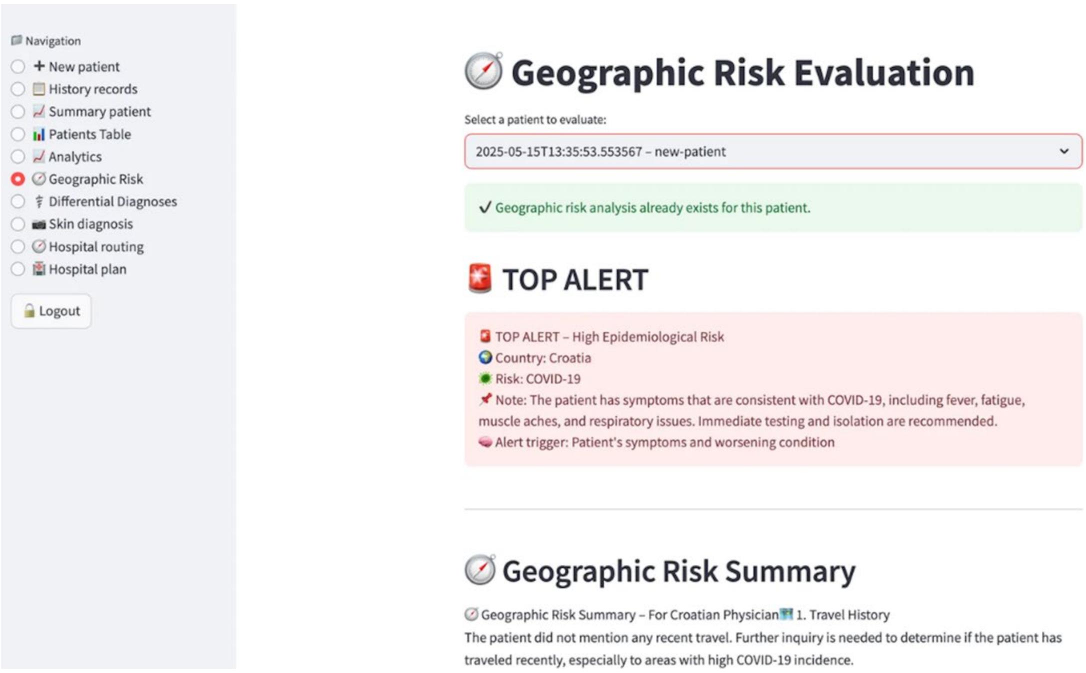

Figure 4: Geographic Risk Evaluation and Alerting

The assistant also provided distinguishing features and recommended diagnostic tests (COVID-19 PCR, rapid influenza test, CBC). This case demonstrates the system's ability to transform unstructured patient narratives

3.3.3 Case 3 – automated hospital plan

Based on the suspected influenza or COVID-19 diagnosis, the Hospital Plan module generated a tailored routing strategy:

- **Emergency Department** – for urgent evaluation or when other services are unavailable
- **Infectious Diseases Unit (Building C, Floor 1)** – if COVID-19 is suspected.
- **Dermatology Department (Building B, Floor 2, open until 16:00)** – for assessment of possible skin lesions.

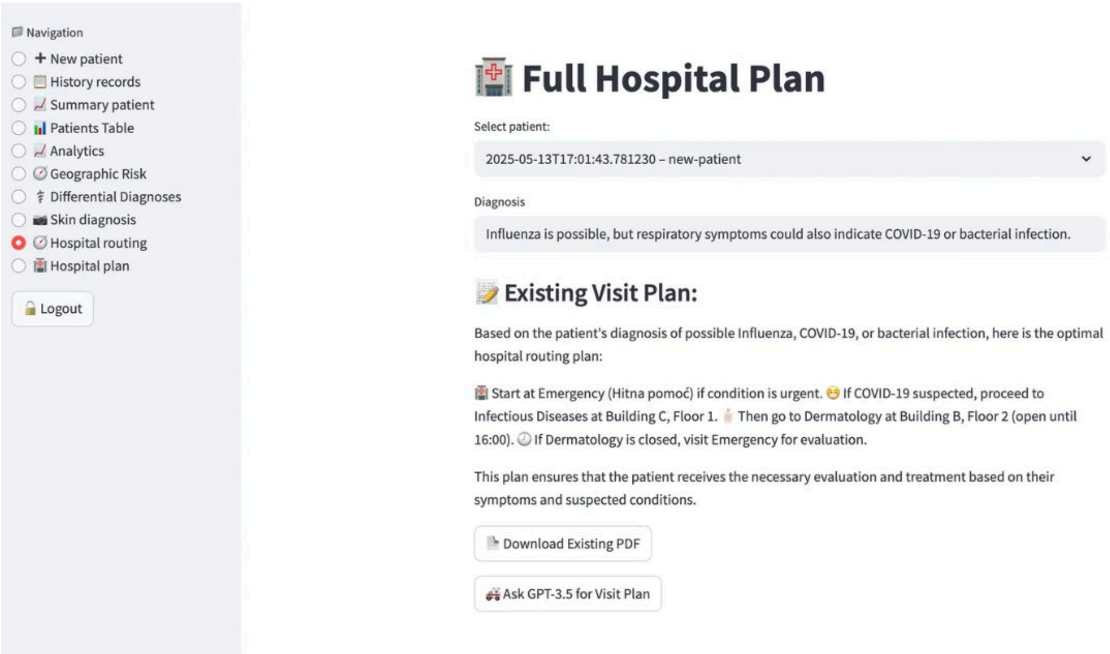

Figure 5: Automated Hospital Plan

4. Discussion

The results of MediMate AI are consistent with emerging literature on the application of generative artificial intelligence (AI) in general practice. Recent studies emphasize that AI-based assistants can reduce documentation workload, improve workflow efficiency, and support clinical reasoning in primary care (You *et al.*, 2025). Our evaluation supports these findings: the prototype reduced administrative burden through automated transcription and summarization of consultations, directly aligning with clinical reports that ambient documentation systems can mitigate GP (General practitioner) burnout. In addition to workflow automation, MediMate AI demonstrated potential in diagnostic support. Similar to earlier work showing that GPT-4 can provide differential diagnoses and triage recommendations (Gaber *et al.*, 2025), our system produced coherent diagnostic suggestions aligned with clinical guideline-based reasoning in approximately 78% of simulated cases. This performance mirrors the diagnostic accuracy observed in studies benchmarking GPT-4 in clinical reasoning tasks (Noda et al. 2025; Goh et al. 2025). Importantly, research indicates that generative AI may particularly support less experienced GPs, helping to bridge gaps in expertise (Geersing *et al.*, 2025). Nevertheless, several limitations must be acknowledged. First, our evaluation relied on simulated data rather than real patient populations, which limits generalizability. Second, while GPT-4 offers strong reasoning capabilities, prior studies have demonstrated that large language models do not always adhere to established clinical guidelines (Hager et al. 2024; Xiaomin and others 2025). This limitation was reflected in occasional inconsistencies observed in our prototype. Moreover, ethical concerns remain. (Xiong *et al.*, 2025) reported that GPT-4 may generate outputs misaligned with medical ethical principles, underscoring the need for transparency and human oversight. Additionally, scholars caution that uncritical reliance on AI may erode clinician expertise and reinforce systemic biases (Pavuluri *et al.*, 2024). Future work should address these limitations by conducting validation in real clinical environments, integrating guideline-grounded reasoning modules, and systematically analyzing outputs for bias. Further research should also explore interoperability with electronic health records, the design of explainability features to support physician trust, and patient-centered evaluations of communication modules (Kharko *et al.*, 2025). Addressing these directions will be essential to ensure safe and effective deployment of MediMate AI in general practice.

In addition to these limitations, further research should focus on bias detection, explainability, and compliance with data protection standards to ensure responsible deployment of AI in clinical settings. Transparent reasoning mechanisms and explainable outputs are essential to strengthen physician trust and allow human oversight in high-risk decisions. Adherence to GDPR and ethical frameworks for medical AI should guide all stages of data processing, model training, and deployment. Establishing structured risk management protocols—including validation pipelines, audit trails, and human-in-the-loop supervision—will be crucial to guarantee safety, accountability, and reproducibility in future versions of MediMate AI.

4.1 Future Work

Future development of MediMate AI will focus on clinical validation, EHR integration, and adding explainable reasoning modules for trustworthy decision support. Incorporating agentic AI approaches—such as planning, reflection, and ReAct-style reasoning—can improve diagnostic reliability, adaptability, and transparency (Nisa *et al.*, 2025). Future prototypes will also include federated learning for privacy-preserving updates and adaptive learning from anonymized data streams.

To ensure equitable access, MediMate AI will be developed as a low-cost, mobile, and multilingual version for rural and resource-limited settings, where AI can strengthen primary care delivery (Lamem, Sahid and Ahmed, 2025). Inspired by recent public health kiosks like HERMES, affordable, privacy-preserving AI systems could extend safe diagnostic assistance to the Global South, supporting health resilience and digital inclusion (Falahati *et al.*, 2025).

5. Conclusion

This paper introduced MediMate AI, a GPT-4–powered prototype designed to support general practitioners by reducing documentation workload and providing real-time diagnostic support. The system integrates multimodal inputs (speech, text, images) into structured clinical outputs such as consultation summaries, symptom lists, differential diagnoses, and triage recommendations. In simulated evaluations, MediMate AI achieved high transcription accuracy, produced clinically coherent summaries, and reduced documentation workload by an estimated 25–30%. Its diagnostic and triage suggestions were generally consistent with clinical guideline-based reasoning, and test users highlighted improved continuity of care through the consolidated patient dashboard. While promising, the system is intended to augment rather than replace medical expertise. Future work will focus on clinical validation, integration with electronic health record systems, and implementation of ethical safeguards. With responsible design and testing, MediMate AI demonstrates the potential of generative AI to enhance efficiency, decision support, and patient communication in primary care.

Ethics declaration: This study did not involve human participants or patient data requiring ethical approval. Prototype evaluation was conducted exclusively with synthetic patient records, simulated consultations, and publicly available datasets. Therefore, no institutional ethics clearance was required. The development and testing of MediMate AI complied with European data protection regulations (GDPR) and ethical standards for AI research. No personal or identifiable patient data were used at any stage, and all prototype evaluations were conducted exclusively on synthetic and publicly available datasets.

AI declaration: Generative AI tools (OpenAI GPT-4 and GPT-3.5) and Whisper were used in the development of the MediMate AI prototype for transcription, summarization, and diagnostic reasoning. AI was also used in the preparation of this manuscript to assist with language refinement, reference formatting, and generation of structured outputs (e.g., JSON templates). All AI-generated content was critically reviewed, edited, and verified by the authors, who take full responsibility for the final version of the paper.

References

Bongurala, A.R. *et al.* (2024) "Transforming health care with artificial intelligence: opportunities and challenges," *American Journal of Medicine*, 137(10), pp. e220–e227. Available at: https://doi.org/10.1016/j.amjmed.2024.05.004.

Fabre, B.L. *et al.* (2024) "Evaluating GPT-4 as an academic support tool for clinicians," *Education and Information Technologies*, 29, pp. 551–566. Available at: https://doi.org/10.1007/s10639-023-11910-y.

Falahati, S. *et al.* (2025) "An AI-powered Public Health Automated Kiosk System for Personalized Care: An Experimental Pilot Study." arXiv. Available at: https://doi.org/10.48550/arXiv.2504.13880.

Gaber, F. *et al.* (2025) "Evaluating large language model workflows in clinical decision support," *NPJ Digital Medicine*, 8, p. 146. Available at: https://doi.org/10.1038/s41746-025-00366-4.

Geersing, G.J. *et al.* (2025) "Generative artificial intelligence for general practice," *British Journal of General Practice*, 75(743), pp. e121–e124. Available at: https://doi.org/10.3399/bjgp25X742398.

Goh, E. *et al.* (2025) "GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial," *Nature Medicine* [Preprint].

Hager, P. *et al.* (2024) "Evaluation and mitigation of the limitations of large language models in clinical settings," *Nature Medicine*, 30(12), pp. 2411–2420. Available at: https://doi.org/10.1038/s41591-024-03222-2.

Kharko, A. *et al.* (2025) "General practitioners' opinions of generative artificial intelligence in healthcare," *Digital Health*, 11, p. 20552076251360863. Available at: https://doi.org/10.1177/20552076251360863.

Lamem, M.F.H., Sahid, M.I. and Ahmed, A. (2025) "Artificial intelligence for access to primary healthcare in rural settings," *Journal of Medicine, Surgery, and Public Health*, 5, p. 100173. Available at: https://doi.org/10.1016/j.glmedi.2024.100173.

Nisa, U. *et al.* (2025) "Agentic AI: The age of reasoning—A review," *Journal of Automation and Intelligence* [Preprint]. Available at: https://doi.org/10.1016/j.jai.2025.08.003.

Noda, R. *et al.* (2025) "GPT-4's performance in supporting physician decision-making in clinical scenarios," *Scientific Reports* [Preprint].

Pavuluri, S. *et al.* (2024) "Balancing act: the complex role of artificial intelligence in clinical practice," *Frontiers in Digital Health*, 6, p. 11344516. Available at: https://doi.org/10.3389/fdgth.2024.11344516.

Xiaomin, L. and others (2025) "Limitations of LLMs in Clinical Guideline Adherence," *Journal of Medical Systems* [Preprint].

Xiong, Y.T. *et al.* (2025) "Exploring the medical ethical limitations of GPT-4 in clinical decision-making," *Frontiers in Public Health* [Preprint].

You, J.G. *et al.* (2025) "Ambient documentation technology in clinician experience of documentation burden and burnout," *JAMA Network Open*, 8(8), p. e2528056. Available at: https://doi.org/10.1001/jamanetworkopen.2025.28056.