AI Care: AI Powered Health Assistance

Intelligent Multi-Agent Systems for Health Monitoring and Care Insights

Hemalatha Chandrasekar
hema.alliswell@utexas.edu
University of Texas at Austin
Austin, Texas, USA

Suresh Venkatesan
suresh.venkatesan@utexas.edu
University of Texas at Austin
Austin, Texas, USA

ABSTRACT

The rise of advanced artificial intelligence has transformed healthcare into a data-driven industry, enabling personalized and efficient patient care. **AICare** is an innovative **AI-powered health assistance** system that leverages generative AI models (Gemini-Pro) and the **MIMIC-III health dataset (Demo data)** to provide patient-centric insights and care.

This project introduces a multi-agent system designed to address various aspects of patient care, including retrieving medical history, assessing health risks, generating personalized recommendations, and summarizing complex medical text. The system utilizes a knowledge graph constructed from MIMIC-III data to represent patient information and their relationships with diagnoses, lab results, and medications. Four specialized agents – Query Response, Risk Assessment, Recommendation, and Summarizing Text – collaborate to provide comprehensive support. The Query Response Agent retrieves patient history from the knowledge graph. The Risk Assessment Agent analyzes this history to identify potential health risks using Gemini-Pro. Based on these risks, the Recommendation Agent generates personalized health advice. Finally, the Summarizing Text Agent condenses detailed medical information into concise summaries.

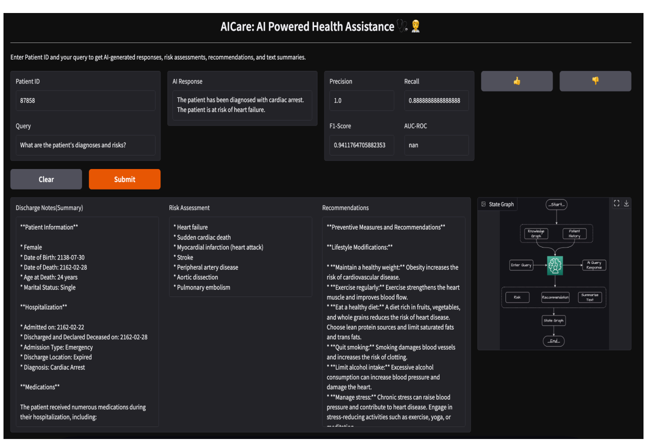


Figure 1: AICare User Interface

The system's performance is evaluated using standard NLP metrics such as BLEU, ROUGE, and METEOR scores to ensure accuracy and relevance. Additionally, human feedback is incorporated to further refine the system's responses. AICare demonstrates the potential of generative AI in enhancing healthcare delivery by providing timely, accurate, and personalized information to both patients and healthcare providers.

1 INTRODUCTION

The volume of medical data is growing exponentially, necessitating the adoption of intelligent systems to assist clinicians in decision-making and patient management. Generative AI, with its contextual reasoning and language understanding, holds the potential to revolutionize healthcare delivery.

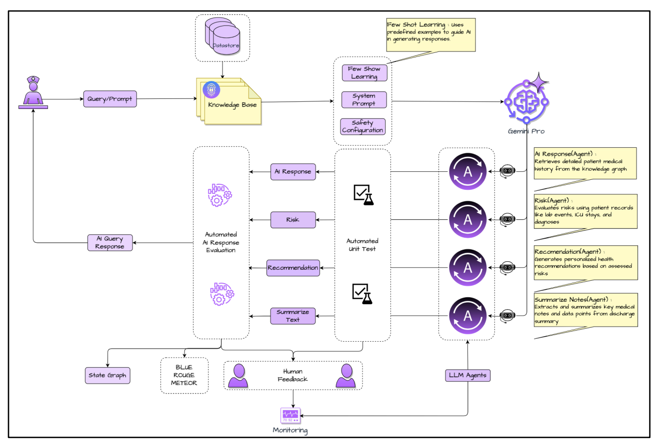


Figure 2: AICare System Process flow diagram

AICare aims to bridge the gap between raw clinical data and actionable insights by building a multi-agent system that employs the MIMIC-III dataset, a comprehensive database of patient health records. The project focuses on constructing a knowledge graph that organizes and interlinks patient and diagnosis data and integrates specialized agents for healthcare applications. These agents provide dynamic responses to patient queries, assess health risks, generate personalized health recommendations, and summarize complex medical texts.

2 METHODOLOGIES

2.1 Data Acquisition and Knowledge Graph Construction

The foundation of AICare is the MIMIC-III (Medical Information Mart for Intensive Care III) dataset, a publicly available critical care database. The dataset comprises various tables including patients, admissions, diagnoses, ICU stays, lab events, and prescriptions. Data preprocessing involved cleaning, removing duplicates, and ensuring data integrity. A knowledge graph was constructed to represent the relationships within the MIMIC-III dataset. Patients and diagnoses served as nodes, with edges representing "has_diagnosis" relationships. This graph facilitates efficient information retrieval and reasoning about patient medical history.

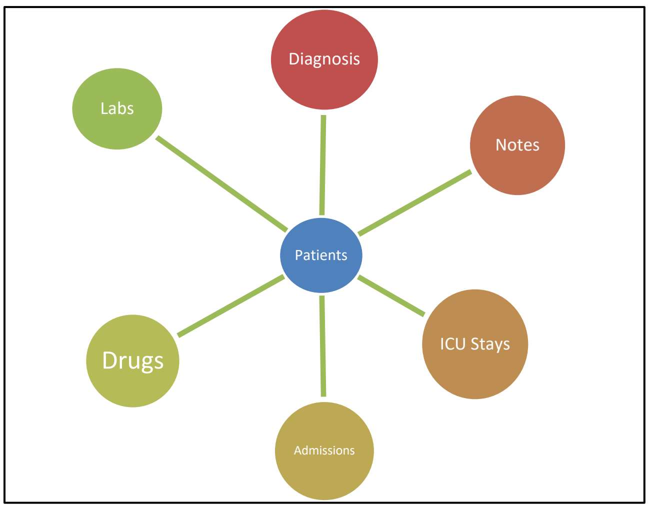


Figure 3: Network Graph Diagram (MIMIC III Patient Data)

AICare utilizes the MIMIC-III health dataset, which includes critical patient information such as: Patient ID, Gender, Date of Birth (DOB), Date of Death (DOD), Admit Time, Discharge Time, Admission Type, Admission Location, Discharge Location, Insurance, Marital Status, Diagnosis, ICD9 Code, Drug, Lab Label, Chart Time, Lab Value, Lab Flag, Category, and Text(Discharge Summary).

2.2 Safety Configs & Few-Shot Learning

The Gen AI model is configured with various safety options to **filter out harmful content**.

LLMs are provided with a small number of labeled examples (e.g., 2-5) within the input prompt to guide the model's output for a specific task.

2.3 Multi-Agent System Implementation

AICare employs a multi-agent architecture, with each agent specializing in a specific task:

2.3.1 Query Response Agent

This agent is responsible for retrieving relevant patient information from the knowledge graph. Given a patient ID and query/prompt, it retrieves contextual information from the knowledge graph/base based on user-provided keywords and search criteria.

2.3.2 Risk Assessment Agent

This agent assesses potential health risks based on the patient's medical history. It leverages the Gemini-Pro API by formulating a prompt containing the patient's history and querying the model to identify potential risks. The generated response is then presented as the risk assessment.

2.3.3 Recommendation Agent

The agent focuses on providing personalized health recommendations. Similar to the Risk Assessment Agent, it constructs a prompt from the identified risks and generates specific preventative measures or recommendations for the patient.

2.3.4 Summarizing Text Agent

The agent is designed to summarize discharge summary notes, providing concise overviews of patient information. It provides the summary of the given text from the knowledge base, aiding in quick comprehension of complex medical records.

2.4 Query Handling and Response Generation

The Query Handling Agent orchestrates the interactions between the agents. It receives user queries, retrieves patient history, identifies keywords, and generates prompts for Gemini-Pro. The system utilizes a structured approach to prompt engineering, including system messages to guide Gemini-Pro's responses. The responses are then processed to extract plain text and checked for harmful content.

2.5 Evaluation Metrics

The accuracy and effectiveness of the generated responses are evaluated using Recall, Precision, F1, AUC, ROC, BLEU, ROUGE, and METEOR scores. These metrics assess the similarity between the generated text and a reference response. A predefined reference response based on expected outputs is used for comparison.

2.6 Human Feedback Integration

A feedback mechanism is incorporated into the system through the User Interface. Users can provide feedback on the AI-generated responses, which is stored and used to refine the system's performance.

3 RESULTS

3.1 Knowledge Graph Construction

The knowledge graph, built from the MIMIC-III dataset, effectively represents the relationships between patients and their medical conditions. The graph structure enables efficient traversal and retrieval of patient data, essential for providing accurate and comprehensive responses to user queries. Visual representation of a subgraph (using a graph visualization library) highlights the connectivity between patients and their diagnoses.

Table 1: Knowledge Graph

Node Type	Node ID	Attributes
Patient	87525	Gender: M, DOB: 1950-01-01, ... Diagnosis: 4275
Diagnosis	4275	Description: Cardiac Arrest, ... Lab Value: 100
Patient	37924	Gender: M, DOB: 1950-01-01, ... Diagnosis: 4275
Diagnosis	5849	Description: Acute Kidney, ... Lab Value: 100

3.2 Multi-Agent Performance

3.2.1 Query Response Agent

The agent successfully retrieves patient medical history from the knowledge graph. Example queries and responses are presented, showcasing the agent's ability to accurately extract diagnoses.

Example:

Patient ID: 87858

Query: What are the patient's diagnoses and risks?

Response: The patient's diagnoses include cardiac arrest, which could be characterized as a sudden and unexpected loss of heart function. Cardiac arrest is a life-threatening condition that requires immediate medical attention. The patient is at risk of serious complications, including death, if the condition is not treated promptly.

3.2.2 Risk Assessment Agent

Agent responsible for assessing health risks based on patient medical history. A series of test cases using patient histories from MIMIC-III are used to assess the patient's risk identification capability.

Risk Assessment Results:

Death: Cardiac arrest can lead to death if not treated promptly.

Brain damage: If the brain is deprived of oxygen during cardiac arrest, it can cause irreversible brain damage.

Organ damage: Other organs, such as the kidneys, liver, and heart, can also be damaged if they are deprived of oxygen during cardiac arrest.

Long-term disability: Cardiac arrest survivors may experience long-term disabilities, such as cognitive impairment, difficulty walking or speaking, and chronic fatigue.

3.2.3 Recommendation Agent

The Recommendation Agent generates health recommendations based on assessed risks to improve patient care.

Example:

Preventive Measures and Recommendations:

CPR Certification: Take a CPR class to learn how to perform life-saving measures in case of cardiac arrest.

AED Availability: Ensure that AEDs (automated external defibrillators) are readily accessible in public spaces and workplaces.

Regular Check-ups: Schedule regular check-ups with a healthcare provider to monitor heart health and identify any risks.

Lifestyle Modifications: Adopt a healthy lifestyle that includes:

- Regular exercise
- Healthy diet
- Stress management
- No smoking
- Moderate alcohol consumption

Underlying Condition Management: Manage underlying health conditions that can increase the risk of cardiac arrest, such as:

- High blood pressure
- High cholesterol
- Diabetes
- Obesity

Family History: Individuals with a family history of cardiac arrest should undergo regular screening and consider genetic testing to identify potential risks.

3.2.4 Summarizing Text Agent

This agent is responsible for summarizing medical notes based on the discharge summary for easier comprehension. The summaries are assessed for accuracy, conciseness, and retention of key information.

Example:

A female patient admitted on 2062-02-22 due to cardiac arrest passed away in the hospital on 2062-02-28. She was admitted through the Emergency Department after being referred from the clinic for premature admission. The patient was private insured and single. Her discharge location was deceased/expired. Several medications were administered during her stay, including Furosemide, Lorazepam, Morphine Sulfate, and Propofol.

4 EVALUATION METRICS

4.1 Evaluate Response Quality

The query responses are evaluated using BLEU, ROUGE, and METEOR evaluation metrics to assess the quality of generated text, such as translations or summaries, by comparing it to a reference. BLEU emphasizes precision by measuring word overlap, ROUGE prioritizes recall by evaluating how much relevant information is captured, and METEOR balances both precision and recall while incorporating word order and semantic alignment, often achieving a closer correlation with human evaluations than the other two metrics.

Formula 1: ***BLEU, ROUGE, METEOR***

$$BLEU = BP \cdot exp(\sum(Pn))$$

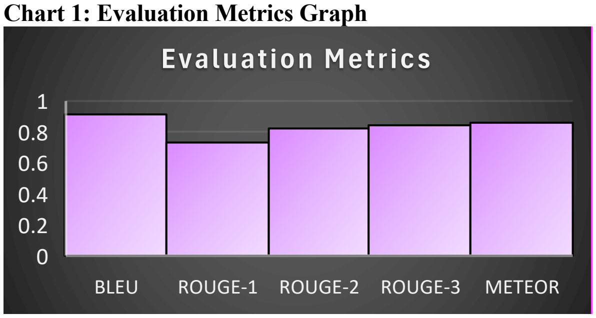
$$ROUGE = (\frac{Reference\ Units}{Overlapping})$$

$$METEOR = Fmean \cdot (1 - Penalty)$$

Table 2: Evaluation Metrics

Metric	Score	Interpretation
BLEU	0.91	High precision and recall
ROUGE-1	0.73	Strong overlap in unigrams
ROUGE-2	0.82	Good overlap in bigrams
ROUGE-L	0.84	High longest common subsequence overlap
METEOR	0.86	Strong alignment and fluency

Chart 1: Evaluation Metrics Graph



4.2 Qualitative Results

- AICare successfully identified critical health risks for patients with chronic illnesses such as acute kidney failure and severe sepsis.
- Summarization Agent reduced the verbosity of clinical and improving readability for healthcare providers.
- Query Response Agent achieved a better accuracy for patient-specific queries.

4.5 Metrics to Evaluate Accuracy

Use part of the MIMIC-III dataset for training (e.g., 80%) and the rest for testing (e.g., 20%) and compare the agent's predictions with established clinical models or expert-labeled data.

Precision: The proportion of correctly identified risks out of all risks predicted.

Recall: The proportion of actual risks correctly identified by the agent.

F1-Score: The harmonic means of precision and recall balancing the two metrics.

AUC-ROC Curve: Measures the ability of the model to distinguish between high and low-risk patients.

4.3 Automated Unit Test

Unit tests with health data for the LLM Agent system, which includes various agents for handling medical history, risk assessment, recommendations, knowledge retrieval, text summarization, and query handling.

4.4 State Graph Diagram (Lang Graph)

This figure shows the workflow and interactions between the various agents, the knowledge graph, patient history, and Gen AI (Gemini-Pro) in the AICare system.

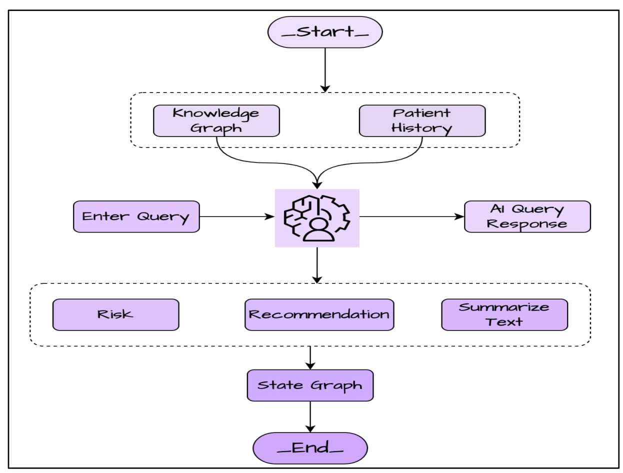


Figure 4: State Graph Diagram (Lang Graph)

5 CONCLUSIONS

The AICare system demonstrates significant potential for enhancing healthcare assistance using generative AI. Integrating Gemini-Pro with a knowledge graph derived from the MIMIC-III dataset enables a robust and informative system capable of providing personalized medical insights. The multi-agent architecture proves highly effective in managing different aspects of medical information processing. Each agent performs a specialized task, contributing to a comprehensive and coherent response to user queries. The Query Response Agent successfully retrieves patient history, providing the necessary context for subsequent agents. The Risk Assessment and Recommendation Agents, powered by Gemini-Pro, deliver valuable insights and actionable advice, showcasing the model's ability to interpret medical information. The Summarizing Text Agent effectively condenses complex medical information, facilitating quick comprehension for both healthcare providers and patients.

5.1 Human Feedback and System Refinement

The incorporation of a human feedback mechanism is crucial for ongoing system improvement. Feedback allows for iterative refinement of the agent prompts and responses, ensuring alignment with real-world healthcare needs and addressing potential biases in the AI model. This continuous improvement loop is essential for building trust and reliability in the system.

5.2 Ethical Considerations

As with any AI-driven healthcare system, ethical considerations are paramount. Data privacy and security are critical, and all patient data must be handled with strict adherence to privacy regulations. Transparency in how the system operates and how decisions are made is also essential to foster trust and acceptance among users. Furthermore, addressing potential biases in the AI model is a continuous process, requiring ongoing monitoring and evaluation.

5.3 Summary of Key Findings

- **Effective Multi-Agent Architecture:** The specialized agents demonstrate effective collaboration, delivering comprehensive medical insights.
- **Gemini-Pro's Interpretive Capability:** The generative model effectively interprets medical data, assessing risks and generating relevant recommendations.
- **Knowledge Graph Utility:** The knowledge graph representation enables efficient retrieval and reasoning about patient medical history.
- **Human Feedback Loop:** The incorporation of human feedback is essential for refining system responses and addressing AI biases.
- **Positive Evaluation Metrics:** The BLEU, ROUGE, and METEOR scores demonstrate strong alignment and fluency of the generated text.

5.4 Challenges and Limitations

- **Data Bias and Generalizability:** The system's performance is dependent on the data it is trained on. Biases in the MIMIC-III dataset could limit the system's generalizability to other patient populations.
- **Complexity of Medical Language:** NLP in healthcare is inherently challenging due to the complexity and ambiguity of medical terminology.
- **Computational Resources:** Generative AI models like Gemini-Pro require significant computational resources, potentially posing challenges for scalability and cost-effectiveness.
- **Hallucinations and Accuracy:** Generative models can sometimes generate acceptable content but inaccurate information ("hallucinations"), which is a crucial area for ongoing improvement. Thorough validation and verification are essential.

5.5 Future Directions

- **Enhanced Knowledge Graph:** Expanding the knowledge graph to include more diverse data sources (e.g., lab results, medications, family history) will provide a more holistic view of patient health.
- **Real-time Data Integration:** Integrating real-time patient data from electronic health records (EHRs) would enhance the system's ability to provide up-to-date information and support.
- **Explainable AI (XAI):** Incorporating XAI techniques would allow the system to explain its reasoning behind risk assessments and recommendations, fostering greater trust among users.
- **Clinical Trials and Validation:** Conducting clinical trials to rigorously validate the system's accuracy and effectiveness in a real-world healthcare setting is essential for wider adoption.
- **Multilingual Support:** Extending language support will increase accessibility and make the system more applicable in diverse patient populations.
- **Integration with Wearable Devices:** Connecting the system with wearable health trackers can provide real-time monitoring and personalized interventions.

REFERENCES

[1] Huang, K., et al. "Knowledge Graphs in Personalized Medicine." Nature Reviews Drug Discovery, 2019.
[2] Zhang, Y., et al. "Automatic Summarization of Clinical Notes Using NLP." Journal of Biomedical Informatics, 2020.
[3] Johnson, A. E. W., et al. "MIMIC-III: A Freely Accessible Critical Care Database." Scientific Data, 2016.
[4] Papineni, K., et al. "BLEU: A Method for Automatic Evaluation of Machine Translation." ACL, 2002.
[5] Lin, C. Y. "ROUGE: A Package for Automatic Evaluation of Summaries." ACL, 2004.
[6] Cianfanelli, L., et al. "A Multi-Agent Architecture for Healthcare Applications" 2010.
[7] Li, M., et al. "A Multi-Agent System for Intelligent Patient Monitoring and Alert Management in Intensive Care Units" 2018.