Understanding What Patients Really Need: A Multi-Intention Recognition and Planning Framework for Complex Medical Queries

Anonymous ACL submission

Abstract

LLM-based multi-agent systems have shown promise in healthcare, enhancing diagnostic accuracy and efficiency. However, most existing systems rely on simplistic and naive doctor-patient dialogues, which fail to capture the complexity of real-world clinical interactions. In practice, patients' self-descriptions are often verbose and contain hidden intents. Accurately extracting these needs and providing appropriate feedback is crucial for improving medical decision-making. To address these challenges, we propose **MIRPF**, a **M**ulti-**I**ntention **R**ecognition and **P**lanning **F**ramework designed to understand patients' complex intentions in healthcare settings. MIRPF first introduces an Intention Recognition module to extract and interpret precise medical intents from verbose queries. Next, a Dynamic Intent Orchestration Agent plans the execution sequence, taking into account the urgency and interdependencies of identified intents. Finally, based on this plan, a Multi-Agent Collaboration System, comprising intention-specific agents and a novel Chain of Thought (CoT)-based Hierarchical Progressive Decision-Making Agent, works collaboratively to complete the diagnostic process. We evaluate MIRPF on two medical dialogue benchmark datasets. The results, measured using automated metrics and expert doctor evaluations, show that MIRPF outperforms existing methods, significantly improving medical proficiency and strategic reasoning.

1 Introduction

Large Language Models (LLMs) (OpenAI, 2022; Achiam et al., 2023; OpenAI, 2024) have shown great potential in various human-machine interactions, such as negotiation (He et al., 2018) and persuasion (Wang et al., 2019). Recently, LLM-based medical assistants (Bao et al., 2023; Yang et al., 2024, 2023b; Shi, 2023; Chen et al., 2023b) have emerged as a promising solution to improve diagnostic efficiency and automate healthcare services. Previous efforts focused on integrating healthcare-specific knowledge into LLMs, achieved through strategies like building specialized knowledge databases (Li et al., 2023b) and fine-tuning models on medical data (Xiong et al., 2023). While these methods improve LLMs' understanding of medical-related questions, they often fall short in dynamic, real-world scenarios. To address this challenge in dynamic medical environments, LLM-based multi-agent systems offer significant potential by enhancing LLMs' ability to follow instructions and make decisions in realistic scenarios.

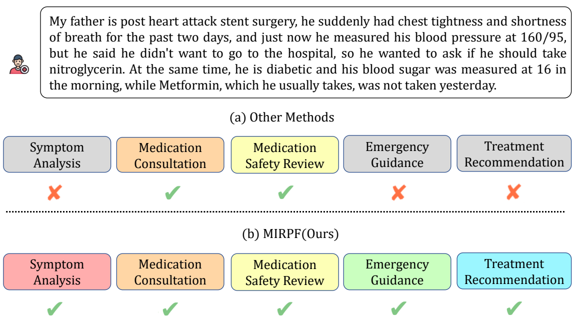


Figure 1: Comparison between other methods and MIRPF on a real-world patient case. Other methods can only identify a limited number of medical intentions from the verbose patient description. In contrast, MIRPF accurately extracts multiple intents, enabling a more comprehensive and reliable diagnostic process.

Multi-agent systems (Fan et al., 2024; Li et al., 2024), where different agents collaborate to handle diverse patient needs, can achieve more accurate medical outcomes. However, existing systems primarily focus on simple, question-answering tasks (Tang et al., 2024a), which struggle to manage the complexity of real-world medical dialogues; as shown in Figure 1. **In practice, patients' self-descriptions are often verbose and contain hidden intentions.** Accurately extracting these needs is essential to improving medical decision-making.

To address above challenges, this paper proposes MIRPF, an LLM-based multi-agent framework designed to better understand patient intentions in healthcare settings. The core idea behind MIRPF is to first extract and decompose complex medical intents through an Intention Recognition module, which enhances the model's ability to process multifaceted medical queries. Next, we introduce a Dynamic Intent Orchestration Agent (DIOA) to intelligently coordinate the execution flow among multiple agents. Finally, we build a multi-agent collaboration system that activates intent-specific agents to complete the diagnostic process and then generate a final decision via a novel Chain-of-Thoughts (CoT)-based Hierarchical Progressive Decision-Making Agent (HPDMA), which synthesizes results from various agents and ensures the coherence and accuracy of the final diagnosis. The main contributions of this work are:

- We propose MIRPF, a multi-intention recognition and planning framework designed for complex medical queries with composite healthcare intents. MIRPF offers three key modules:
  - The intention recognition module enables effective intent recognition, comprising three key components: intention extraction, context-aware decomposition, sub-query validation.
  - The dynamic intent orchestration agent selectively activates intent-specific agents, optimizing the multi-agent workflow by assigning tasks based on recognized patient intents.
  - A multi-agent system is designed, incorporating a novel CoT-based hierarchical progressive decision-making agent. This ensures that each agent's contribution is integrated in a contextualized manner, leading to a more coherent and effective decision-making process.
- Extensive experiments on two medical dialogue benchmark datasets demonstrate the effectiveness and necessity of MIRPF.
- We contribute a dataset consisting of verbose medical queries generated from real-world medical consultation records (Fan et al., 2024). This dataset provides valuable insights for the research community by offering data that more accurately reflects real-world medical conversations.

2 Related Work

LLM-driven Multi-agent Healthcare Systems LLMs like ChatGPT and GPT-4 (OpenAI, 2022; Achiam et al., 2023) excel in multidisciplinary tasks. While current LLMs (Bai, 2023; Yang et al., 2023a; AI, 2024) possess medical knowledge, they lack specialized expertise for domain-specific applications. Recent medical LLMs like HuatuoGPT, ZhongJing, PediatricsGPT, Meditron, and MedicalGPT (Zhang et al., 2023; Chen et al., 2023a; Yang et al., 2023b; Shi, 2023; Yang et al., 2024; Chen et al., 2023b; Xu, 2023; Bao et al., 2023) address this gap through various approaches such as doctor-patient conversations, expert feedback, and specialized training frameworks. While LLMs have demonstrated remarkable capabilities in medical applications, they still face challenges with hallucination and inadequate contextual comprehension in complex clinical scenarios. To address these limitations, multi-agent systems have emerged as a promising solution. Recent approaches like MDAgent and MEDAgent (Tang et al., 2024a; Kim et al., 2024) leverage dynamic multi-expert discussion mechanisms to tackle complex medical intricacies.

Multi-Intent Comprehension Despite advancements in multi-agent systems for healthcare, effectively addressing multiple intents within patient inquiries remains a significant challenge. Accurate multi-intent recognition and hierarchical medical task planning continue to encounter substantial obstacles. Traditional intent recognition systems have evolved from rule-based approaches to sophisticated neural architectures. Earlier methodologies primarily relied on sequence labeling and hierarchical classification frameworks. The emergence of LLMs has shifted the focus toward leveraging their capabilities for intent comprehension. Techniques such as prompt engineering and fine-tuning have yielded notable advancements, particularly in domain-specific intent classification. However, current LLM-based approaches, exemplified by systems like clinical agent, are constrained by predefined workflows and rigid rule sets, limiting their adaptability in dynamically managing complex multi-intent medical scenarios.

Reasoning in Medicine Traditional Chain-of-Thought (CoT) and its extensions, such as CoT-SC, Tree of Thought (ToT), Graph of Thought (GoT), and MedCoT (Wei et al., 2023; Wang et al., 2022; Yao et al., 2023; Besta et al., 2024; Liu et al., 2024) have shown promise in clinical decision-making by breaking down complex problems into intermediate, logical steps. These methods have been successfully applied in clinical error correction, diagnostic reasoning, and other medical domains. However, they often rely on static, generalized reasoning patterns that limit their adaptability in dynamic medical environments, where patient conditions and new evidence continuously evolve. In this work, we address these limitations of CoT by integrating CoT with medical knowledge graphs, enabling multi-stage, iterative reasoning that adapts to real-time changes in patient conditions and medical evidence, fostering more dynamic and context-sensitive decision-making in clinical settings.

3 Methodology

In this section, we first present an overview workflow of our proposed MIRPF. We then introduce its three main modules: (1) the Intention Recognition module, (2) the Dynamic Intent Orchestration Agent, and (3) the Multi-Agent System.

3.1 Overall Workflow of MIRPF

As shown in Fig. 2, MIRPF follows a three-stage process to achieve intent recognition and effective decision-making.

First, in the *intention recognition stage*, we propose an Intention Recognition module. This module receives a complex and verbose medical query from the patient and is capable of extracting the underlying intentions behind the detailed text.

Second, in the *planning stage*, we introduce the Dynamic Intent Orchestration Agent (DIOA). This adaptive approach dynamically arranges and executes 14 fundamental medical intents in the subsequent multi-agent system. DIOA works as an orchestrator, coordinating the execution of tasks based on recognized intentions.

Finally, in the Execution Stage, we propose a novel multi-agent collaboration system. This system dynamically engages intent-specific agents for different steps of the clinical process, with a novel CoT-based Hierarchical Progressive Decision-Making Agent (HPDMA) to execute the intentions and systematically form final decisions.

3.2 Intention Recognition Module

In this module, we leverage the LLMs to systematically decompose complex medical queries into clinically relevant sub-queries, each addressing a distinct medical intent using following components.

Intention Extraction. The Intention Extraction sub-module identifies the underlying medical intents in a complex query by analyzing its semantic structure and clinical context. Given an input query Q, the system generates a set of potential intents $I = \{i_1, i_2, \ldots, i_n\}$ using a prompt-based approach:

$$I = \text{LLM}(Q, P_{\text{intent}}),$$

where P_{intent} is a structured prompt designed to extract medical intents from the query. The output is a set of intents, each associated with a confidence score. Intents with confidence scores above a predefined threshold are retained for further processing.

Context-Aware Decomposition. Once the intents are identified, this sub-module decomposes the verbose query into sub-queries that retain the original query's clinical context while focusing on specific intents. For each intent $i_k \in I$, a sub-query q_k is generated using a context-aware prompt $P_{\text{decompose}}$:

$$q_k = \text{LLM}(Q, P_{\text{decompose}}, i_k),$$

where $P_{\text{decompose}}$ ensures that the sub-query incorporates relevant clinical details, such as patient history, symptoms, and medications.

Sub-Query Validation. To ensure the quality and relevance of the generated sub-queries, we implement a validation mechanism that evaluates the following criteria: **(a) Intent Orthogonality**: The semantic overlap between sub-queries is measured using Jaccard similarity $J(q_k, q_l)$ for all pairs (q_k, q_l). Sub-queries with $J(q_k, q_l) > 0.2$ are flagged for revision. **(b) Clinical Coverage**: The system verifies all critical clinical parameters (e.g., symptoms, vital signs, medications) mentioned in the original query are preserved in the sub-queries. **(c) Emergency Prioritization**: Sub-queries related to urgent medical conditions (e.g., chest pain, high blood pressure) are prioritized based on a predefined triage scoring system.

3.3 Dynamic Intent Orchestration Agent

We propose the Dynamic Intent Orchestration Agent (DIOA), which introduces an adaptive approach to dynamically arrange the 14 fundamental medical intents. Unlike traditional static medical workflows, our DIOA functions as a core agent that autonomously evaluates and adjusts the execution sequence based on both local and global dynamic factors. By incorporating real-time medical context and patient-specific conditions, the agent

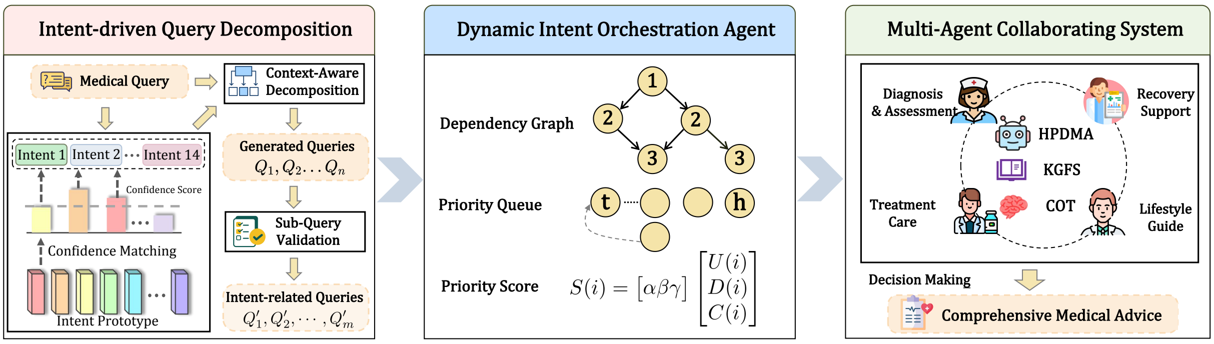


Figure 2: **Overview of the MIRPF framework.** The framework comprises three core modules: (1) Intent-driven Query Decomposition (left): The system breaks down complex queries into sub-queries by performing **Intent Recognition**, **Context-Aware Decomposition**, and **Sub-Query Validation**. (2) Dynamic Intent Orchestration Agent (center): This agent builds a **Dependency Graph** and calculates priority scores to reorder sub-query execution based on urgency, dependency, and contextual priority. (3) Multi-Agent Collaborating System (right): This module **dynamically activates** one or more of the four specialized agents—*Diagnosis & Assessment*, *Treatment Care*, *Recovery Support*, and *Lifestyle Guide*—based on recognized intents. A central **HPDMA** agent with **CoT-based Reasoning** and **KGFS** integration consolidate their insights to generate context-aware medical advice.

ensures optimal orchestration of medical intents while maintaining clinical safety and efficiency. The agent collaboratively computes a composite priority score for each intent using the following function:

$$S(i) = \alpha U(i) + \beta D(i) + \gamma C(i) \qquad (1)$$

where $U(i)$ represents medical urgency, quantified through analysis of critical keywords and vital signs. $D(i)$ represents dependency impact, measured by both incoming and outgoing relationships in the medical intent graph. $C(i)$ represents contextual priority, incorporating intent complexity and patient-specific conditions. The coefficients α, β and γ are dynamically adjusted based on the medical context, with constraints $\alpha + \beta + \gamma = 1$ and $\alpha, \beta, \gamma \geq 0$.

The DIOA employs a modified version of Kahn's topological sorting algorithm with priority-based processing. This modification ensures that medical intents are not only executed in a logically correct order but also prioritized according to clinical importance. The algorithm maintains a priority queue that performs individual reordering operations in $O(\log n)$ time, resulting in an overall complexity of $O((|V|+|E|) \log |V|)$ for the complete intent orchestration process (see Algorithm 1). The system features a dynamic dependency adjustment mechanism that transforms the intent execution graph in response to emerging medical scenarios. This is achieved through:

- Real-time monitoring of patient vital signs and clinical indicators
- Continuous evaluation of intent dependencies through a weighted DAG
- Adaptive adjustment of priority coefficients based on medical context

When new medical evidence or changes in patient condition arise, the system recalculates intent priorities using the composite scoring function in Equation 1, considering:

(a) Immediate urgency: derived from patient vital signs, clinical indicators, and medical keyword analysis with predefined weights

(b) Interdependency strength: calculated through a weighted directed acyclic graph (DAG) where edge weights w_{ij} between intents i and j are dynamically updated based on:

$$w_{ij} = \lambda R_{ij} + (1 - \lambda) H_{ij} \qquad (2)$$

where R_{ij} represents the real-time relationship strength and H_{ij} captures historical clinical significance

(c) Contextual priority: incorporating temporal medical constraints, patient-specific conditions, and intent complexity scores determined by medical domain expertise

This dynamic orchestration approach enables the system to maintain optimal intent execution sequences while adapting to changing medical scenarios, ensuring both efficiency and clinical safety.

3.4 Multi-Agent Collaborating System

In this subsection, we introduce four Intention-Specific Agents, each designed to address distinct aspects of the medical decision-making process, along with a novel CoT-based Hierarchical Progressive Decision-Making agent that integrates these contributions to enhance overall decision-making.

Intention-Specific Agents. Following the Dynamic Intent Orchestrator Agent, the corresponding Intention-Specific Agents are dynamically activated when the framework recognizes relevant and valid intents. To ensure comprehensive and diverse coverage, agent collaboration spans four key stages of healthcare: Diagnosis & Assessment, Treatment-Care, Recovery-Support, and Lifestyle-Guide.

- **Diagnosis & Assessment Agent**: This agent focuses on the initial diagnostic process, including symptom analysis, investigation of potential causes, and the interpretation of test results. It also offers second opinions to ensure accurate diagnoses. It is activated with the following intents: Department Recommendation, Symptom Analysis, Cause Investigation, Test Result Interpretation, and Second Opinion Diagnosis.

- **Treatment-Care Agent**: Responsible for providing treatment recommendations, including guidance on medication, surgery, and other care strategies. This agent ensures that treatments are both effective and safe for the patient. This agent is activated with the following intents: Treatment Recommendations, Surgery-Related Consultation, Medication Consultation, and Medication Safety Review.

- **Recovery-Support Agent**: This agent supports patients during recovery, offering rehabilitation guidance and psychological support to help patients manage both physical recovery and emotional well-being. This agent is activated with the following intents: Rehabilitation Guidance and Medical Psychological Support.

- **Lifestyle-Guide Agent**: Focused on long-term health, this agent provides advice on diet, exercise, and emergency management, promoting overall wellness and preventing further medical issues. This agent is activated with the following intents: Dietary Advice, Exercise Guidance, and Emergency Guidance.

CoT-based Hierarchical Progressive Decision-Making Agent. To incorporate advice from all four Intention-Specific Agents and integrate the information for the final decision, we propose a Hierarchical Progressive Decision-Making Agent (HPDMA) with CoT properties. This agent combines outputs from activated agents with the patient's medical history to dynamically adjust treatment protocols and formulate rehabilitation plans.

HPDMA enhances clinical decision-making by addressing the limitations of traditional CoT reasoning models. Unlike conventional CoT approaches, which often rely on generalized reasoning patterns, this agent transforms clinical decision processes into more structured, interpretable pathways. It integrates CoT with medical knowledge graphs to enable multi-stage, iterative reasoning, progressing from symptom identification to the exploration of underlying pathological mechanisms.

Specifically, we use LLM within a novel prompt engineering framework tailored to incorporate medical semantics for improved relevance and accuracy. A core element of the framework is the Context-Aware Prompt Generator, which utilizes a Knowledge-Guided Few-shot Sampling (KGFS) strategy. This strategy extracts symptom-disease-treatment triplets from established medical ontologies, such as SNOMED CT and UMLS, to generate dynamic, context-sensitive prompts. By combining these elements, the HPDMA ensures that the reasoning process is both contextually accurate and clinically valid.

4 Experiments

In this section, we first introduce the dataset used in our experiments. We then describe the experimental setup and evaluation metrics. Then, we present the results from testing our framework on public benchmarks to evaluate its overall effectiveness. Finally, we conduct multiple ablation study to validate the accuracy of intent recognition and assess the necessity of each component in the framework.

4.1 Datasets

We evaluate our framework on two medical dialogue datasets. The first is our proposed MIRPF-Datase dataset, consisting of 2,200 samples carefully extracted from real-world medical diagnosis processes. Each sample captures rich, multi-intent medical scenarios where patients express multiple medical needs within a single consultation.

The second dataset is a subset of 2,200 samples from the HuaTuo-26M dataset(Li et al., 2023a), a

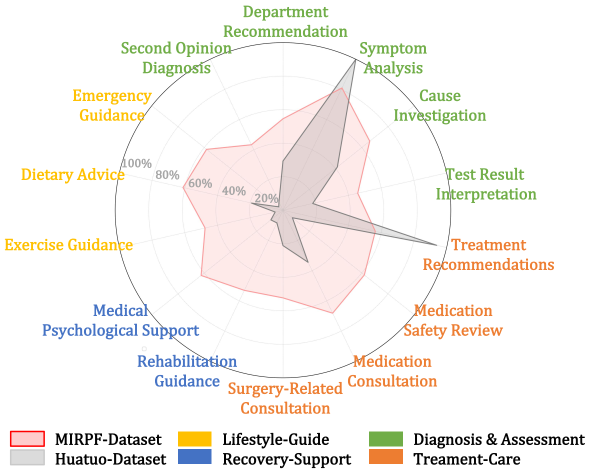


Figure 3: Radar chart comparing the distribution of medical intent categories between MIRPF-Dataset and Huatuo-Dataset. The values are normalized by dividing each data point by the maximum value across both datasets to ensure comparable scales from 0 to 1. The chart encompasses 14 different medical intent categories, with values plotted on a polar coordinate system.

large-scale Chinese medical dialogue corpus containing conversations between doctors and patients on various medical topics. We selected these samples to ensure a balanced representation of diverse medical consultation scenarios.

4.2 Experiment Setup and Evaluation Metrics

Compared Methods. We compare our method against several state-of-the-art approaches, all leveraging GPT-4 as the base model. These include both single-agent and multi-agent methods:

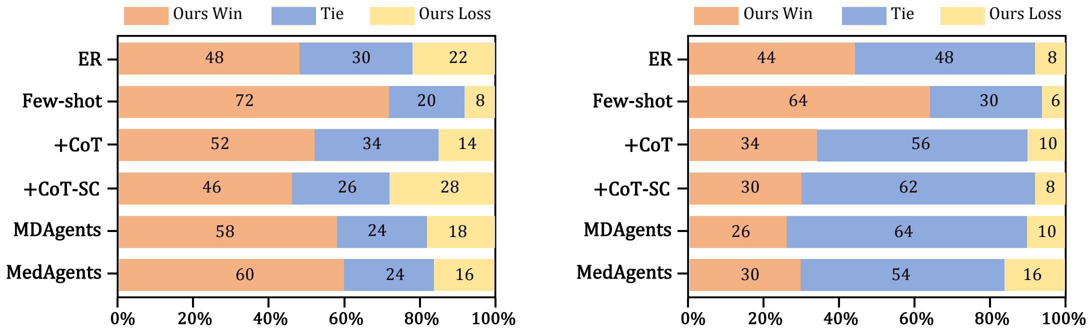
- Single-Agent Methods:
  - Zero-shot employs direct prompt-based inference without task-specific examples.
  - Few-shot improves performance by incorporating a limited set of in-context demonstrations.
  - CoT (Wei et al., 2023) extends Few-shot by integrating intermediate reasoning steps to derive the final answer.
  - CoT-SC (Wang et al., 2022) further enhances robustness by generating multiple reasoning chains and selecting the most consistent answer through majority voting.
  - Ensemble Refinement (ER) (Singhal et al., 2023) strengthens reasoning by aggregating outputs from diverse reasoning paths, ensuring more reliable and accurate results.
- Multi-Agent Methods::
  - MedAgents (Tang et al., 2024b) address medical multiple-choice questions using distinct multi-expert collaboration approaches. It utilizes 5 expert agents engaging in interactive discussions to determine answers.
  - MDAgents is then evaluated in three variants: *(a) Base (single agent)*, *(b) Collaboration (three agents with direct interaction)*, *(c) Group (three agents with structured group discussions)*. Each variant employs different collaboration mechanisms to analyze and solve medical questions.

Evaluation Metrics. We evaluate model performance through three key clinical dimensions:

- Intent Comprehension: This measures the system's ability to accurately identify not only surface-level intents but also implicit and nuanced needs, uncovering the contextual subtleties that drive clinical decision-making.
- Clinical Planning: This assesses the model's ability to strategically orchestrate and prioritize intents, creating a coherent, step-by-step reasoning pathway that aligns with medical best practices while ensuring safety and relevance.
- Response Quality: This evaluates the precision, clarity, and clinical utility of the generated answers, ensuring they are actionable and grounded in evidence-based medicine.

4.3 Experimental Results

Results on Public Benchmark. The Huatuo dataset analysis reveals distinct optimization patterns. MIRPF maintains superiority but with narrower margins (+8.4% over MDAgents-Group), indicating MIRPF's adaptive efficiency across intent complexity spectra. Notably, single-agent strategies like Few-shot prompting(+CoT-SC) achieve competitive Response Quality, suggesting chain-of-thought benefits in simpler scenarios. However, even in this lower-complexity environment, MIRPF's multi-scale intent modeling yields 11.7% improvement in Clinical Planning over vanilla multi-agent systems, validating its robust clinical decision-making framework. The results highlight our architecture's dual capability: excelling in both high-complexity edge cases and general clinical QA through hierarchical intent-resolution protocols.



(a) Verbose Medical Intent Q&A Evaluation (b) Verbose Huatuo Q&A Evaluation

Figure 4: Response comparisons of MIRPF with other baselines via doctor evaluation. In this evaluation, MDAgents employs its Group mode setting.

Table 1: This is caption

Category	Method	Verbose Huatuo Dataset			
		Intent Comprehension	Clinical Planning	Response Quality	Average
Single-Agent	Zero-shot	3.80	3.75	3.74	3.76
	Few-shot	4.19	3.64	4.03	3.95
	+CoT	4.12	3.94	4.01	4.02
	+CoT-SC	4.02	4.11	3.94	4.02
	ER	4.11	3.73	4.02	3.95
	MDAgents-Base	3.77	3.94	3.63	3.89
Multi-Agent	MedAgents	3.83	3.99	3.85	3.78
	MDAgents-Collaboration	3.88	4.02	3.81	3.89
	MDAgents-Group	4.04	4.11	4.12	4.09
	MIRPF (Ours)	**4.76**	**4.32**	**4.62**	**4.57**
Category	**Method**	**Verbose Medical-Intent Q&A Dataset**			
		Intent Comprehension	Clinical Planning	Response Quality	Average
Single-Agent	Zero-shot	4.02	3.88	3.83	3.91
	Few-shot	4.31	3.91	4.05	4.09
	+CoT	4.39	4.15	4.12	4.22
	+CoT-SC	4.37	4.32	4.10	4.26
	ER	4.26	3.91	3.98	3.99
	MDAgents-Base	4.15	4.26	3.85	4.09
Multi-Agent	MedAgents	4.19	4.31	4.03	4.18
	MDAgents-Collaboration	4.27	4.33	4.02	4.21
	MDAgents-Group	4.43	4.49	4.28	4.40
	MIRPF (Ours)	**4.89**	**4.75**	**4.66**	**4.77**

Multi-Intent Synergy in Clinical Reasoning. On our proprietary benchmark with rich multi-intent interactions, MIRPF achieves state-of-the-art performance across all metrics, demonstrating 12.1% absolute improvement over the strongest multi-agent baseline (MDAgents-Group). Single-agent approaches exhibit inherent limitations while few-shot prompting (+CoT) reaches 4.02 average score, it struggles with intent entanglement. Multi-agent variants show incremental gains through collaboration mechanisms, yet their single-model architecture limits specialized intent resolution. Our framework's dynamic intent disentanglement and parallel reasoning pathways prove critical for handling complex clinical narratives.

Expert Doctor Evaluation. We conducted a comprehensive evaluation where three medical experts assessed our approach against five baseline methods through majority voting on 50 randomly selected samples from each dataset. As shown in Figure 4, MIRPF exhibits superior performance, particularly on both medical benchmark. On our medical intent Q&A dataset, MIRPF achieves the highest win rate of 72% against Few-shot baseline with only 8% losses. The performance advantage remains strong when compared to reasoning-enhanced methods (52% wins vs CoT, 46% vs CoT-SC) and agent-based approaches (58% vs MDAgents with group-mode multi-agent collaboration, 60% vs MedAgents), with consistent tie rates around 24-34

The comparison patterns shift on Huatuo benchmark, where higher tie rates emerge across all baselines except Few-shot (64% win rate). Notably, tie rates increase to 56-62% for CoT variants and 54-64% for agent approaches, with win rates decreasing to 26-34%. This contrast between datasets (20-34% ties in MIRPF-Dataset vs 48-64% in Huatuo-dataset) demonstrates that better discriminates model capabilities through its more complex and diverse queries.

4.4 Ablation Studies

Intent Detection Analysis. The experimental results demonstrate the superior performance of our MIRPF method across all metrics in intent detection. Among baseline methods, Few-shot achieves the highest accuracy and precision among traditional approaches, while CoT-SC leads in recall, suggesting that chain-of-thought reasoning particularly helps in capturing diverse intents. How-

Method	Accuracy	Precision	Recall	F1 Score
ER	0.75	0.43	0.64	0.50
MedAgent	0.71	0.46	0.70	0.55
Zero-Shot	0.75	0.44	0.55	0.47
Few-Shot	0.79	0.61	0.72	0.64
+Cot	0.77	0.56	0.79	0.64
+Cot-SC	0.77	0.57	0.82	0.66
MDAgent	0.68	0.36	0.61	0.44
+Collaboration	0.77	0.56	0.81	0.65
+Group	0.76	0.53	0.82	0.64
MIRPF (Ours)	**0.90**	**0.82**	**0.84**	**0.82**

Table 2: Performance of Different Intent Recognition Methods. Underlined values indicate the highest values in each metric, excluding our results.

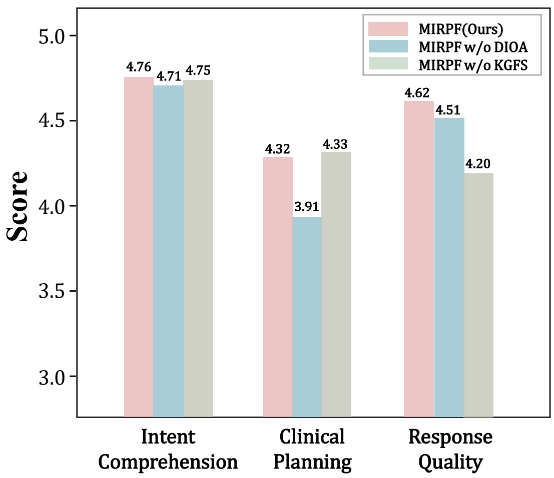


Figure 5: Ablations of DIOA and KGFS on three clinical dimensions.

ever, the generally low precision scores across baseline methods indicate a common challenge: they tend to generate redundant and incorrect intents. Multi-agent approaches like MDAgent-Collaboration show promise in improving recall through agent cooperation, but still struggle with precision. Our MIRPF method addresses these limitations effectively, achieving significant improvements across all metrics, demonstrating its capability to both reduce false positives and accurately capture true intents.

Necessity of DOIA. To evaluate the effectiveness of the Dynamic Intent Orchestration Agent (DIOA), removing DIOA leads to performance degradation across all three dimensions. Notably, Clinical Planning sees a significant drop of 10.5%, validating the importance of DIOA's dynamic priority scoring mechanism in handling complex medical scenarios. Specifically, by comprehensively considering urgency U(i), dependency impact D(i), and contextual priority C(i), DIOA better captures the intrinsic relationships between medical intents, enabling more rational planning decisions. The decreases in Intent Comprehension and Response Quality further demonstrate the necessity of dynamic intent orchestration for enhancing overall system performance.

Importance of KGFS. We further examined the role of Knowledge-Guided Few-shot Sampling (KGFS) in our Hierarchical Progressive Decision-Making Agent. Removing KGFS only marginally affects Intent Comprehension and Clinical Planning, but causes a notable 9.09% drop in Response Quality. This suggests that while the system can still detect and schedule medical intents, lacking structured medical references (e.g., symptom–disease–treatment triplets) undermines its ability to ground recommendations in precise, domain-specific reasoning, ultimately diminishing the overall quality of the generated responses.

5 Conclusion and Future Work

In this paper, we propose MIRPF, a novel multi-agent framework designed to address the critical challenge of understanding verbose medical intents in real-world healthcare scenarios. By integrating three core components—the Intention Recognition Module, Dynamic Intent Orchestration Agent (DIOA), and CoT-based Hierarchical Progressive Decision-Making Agent (HPDMA)—our framework significantly advances the capability of LLM-based systems to parse verbose patient queries, prioritize clinical actions, and synthesize evidence-based medical decisions. Our comprehensive evaluations on both proprietary datasets and established benchmarks demonstrate that MIRPF outperforms previous single-agent and multi-agent approaches in key clinical metrics, including intent comprehension, clinical planning, and response quality. By synergizing hierarchical intent understanding with adaptive multi-agent collaboration, our framework pioneers new pathways for context-aware clinical intelligence, fundamentally transforming how LLM systems process complex healthcare narratives and advance evidence-based medical reasoning.

Future Work. We plan to incorporate complementary modalities (*e.g.,* medical imaging) in the future to improve the potential of MEDAIDE in multimodal diagnostics and applications.

Limitations

Despite the successes of our framework in demonstrating promising performance in medical decision-making tasks, we recognize several limitations that open pathways for future research. First, while the current system achieves robust intent recognition for common conditions, our dataset lacks comprehensive coverage of rare diseases and specialized medications, which may limit generalizability to niche clinical scenarios. Second, the clinical validity of our CoT-based Hierarchical Progressive Decision-Making Agent (HPDMA) depends on symptom-disease-treatment triplets derived from static medical ontologies (SNOMED CT/UMLS). This foundational design choice, while ensuring structured reasoning through Knowledge-Guided Few-shot Sampling (KGFS), creates inherent constraints in adapting to 1) emerging medical discoveries not yet codified in these ontologies, and 2) region-specific clinical protocols that diverge from standardized guidelines.

Ethics Consideration

Ethical considerations are paramount in the application of medical agents to real-world clinical settings. We are fully aware of the potential impacts of our research and have taken deliberate actions to address these issues. To enhance transparency, we are committed to publicly making the drug data and medical records used in our study accessible. This will enable other researchers to validate our findings and build upon our work, fostering collaboration and advancement in this field.

We are acutely aware of the necessity for privacy and data protection. All data utilized has undergone thorough de-identification, with all sensitive information removed, and verified by a partnering medical institution. We invite doctors to perform only evaluations of model responses without involving any form of human subject research. All participants are compensated $300 for their work, which strictly adheres to the minimum hourly rate for the region in which the work is performed. For the utilization of healthcare-related data, we strictly follow the license agreements of publicly available databases. For the constructed data, we have undergone an internal ethical review by the ethics review board of our partnering medical institutions and are licensed and approved.

References

2023. Baichuan. https://github.com/baichuan-inc/Baichuan-13B.

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

Baichuan AI. 2024. Introducing baichuan4. https://platform.baichuan-ai.com/homePage.

Zhijie Bao, Wei Chen, Shengze Xiao, Kuang Ren, Jiaao Wu, Cheng Zhong, Jiajie Peng, Xuanjing Huang, and Zhongyu Wei. 2023. Disc-medllm: Bridging general large language models and real-world medical consultation. arXiv preprint arXiv:2308.14346.

Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michal Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, and Torsten Hoefler. 2024. Graph of thoughts: Solving elaborate problems with large language models. Proceedings of the AAAI Conference on Artificial Intelligence, 38(16):17682–17690.

Junying Chen, Xidong Wang, Anningzhe Gao, Feng Jiang, Shunian Chen, Hongbo Zhang, Dingjie Song, Wenya Xie, Chuyi Kong, Jianquan Li, Xiang Wan, Haizhou Li, and Benyou Wang. 2023a. Huatuogpt-ii, one-stage training for medical adaption of llms. ArXiv, abs/2311.09774.

Zeming Chen, Alejandro Hernández-Cano, Angelika Romanou, Antoine Bonnet, Kyle Matoba, Francesco Salvi, Matteo Pagliardini, Simin Fan, Andreas Köpf, Amirkeivan Mohtashami, Alexandre Sallinen, Alireza Sakhaeirad, Vinitra Swamy, Igor Krawczuk, Deniz Bayazit, Axel Marmet, Syrielle Montariol, Mary-Anne Hartley, Martin Jaggi, and Antoine Bosselut. 2023b. Meditron-70b: Scaling medical pretraining for large language models. Preprint, arXiv:2311.16079.

Zhihao Fan, Jialong Tang, Wei Chen, Siyuan Wang, Zhongyu Wei, Jun Xi, Fei Huang, and Jingren Zhou. 2024. Ai hospital: Benchmarking large language models in a multi-agent medical interaction simulator. Preprint, arXiv:2402.09742.

He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. arXiv preprint arXiv:1808.09637.

Yubin Kim, Chanwoo Park, Hyewon Jeong, Yik Siu Chan, Xuhai Xu, Daniel McDuff, Hyeonhoon Lee, Marzyeh Ghassemi, Cynthia Breazeal, and Hae Won Park. 2024. Mdagents: An adaptive collaboration of llms for medical decision-making. Preprint, arXiv:2404.15155.

Binxu Li, Tiankai Yan, Yuanting Pan, Zhe Xu, Jie Luo, Ruiyang Ji, Shilong Liu, Haoyu Dong, Zihao Lin, and Yixin Wang. 2024. Mmedagent: Learning to use medical tools with multi-modal agent. arXiv preprint arXiv:2407.02483.

Jianquan Li, Xidong Wang, Xiangbo Wu, Zhiyi Zhang, Xiaolong Xu, Jie Fu, Prayag Tiwari, Xiang Wan, and Benyou Wang. 2023a. Huatuo-26m, a large-scale chinese medical qa dataset. ArXiv, abs/2305.01526.

Yunxiang Li, Zihan Li, Kai Zhang, Ruilong Dan, Steve Jiang, and You Zhang. 2023b. Chatdoctor: A medical chat model fine-tuned on a large language model meta-ai (llama) using medical domain knowledge. Cureus, 15(6).

Jiaxiang Liu, Yuan Wang, Jiawei Du, Joey Tianyi Zhou, and Zuozhu Liu. 2024. Medcot: Medical chain of thought via hierarchical expert. Preprint, arXiv:2412.13736.

OpenAI. 2022. Introducing chatgpt. https://openai.com/blog/chatgpt.

OpenAI. 2024. Introducing o1. https://openai.com/index/introducing-openai-o1-preview/.

Liu Lin Ju Shi. 2023. Cmlm-zhongjing-2-1-8b: A state-of-the-art edge computing language model for traditional chinese medicine. https://github.com/pariskang/CMLM-ZhongJing.

Karan Singhal, Shekoofeh Azizi, Tao Tu, S. Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, Perry Payne, Martin Seneviratne, Paul Gamble, Chris Kelly, Abubakr Babiker, Nathanael Schärli, Aakanksha Chowdhery, Philip Mansfield, Dina Demner-Fushman, Blaise Agüera y Arcas, Dale Webster, Greg S. Corrado, Yossi Matias, Katherine Chou, Juraj Gottweis, Nenad Tomasev, Yun Liu, Alvin Rajkomar, Joelle Barral, Christopher Semturs, Alan Karthikesalingam, and Vivek Natarajan. 2023. Large language models encode clinical knowledge. Nature, 620(7972):172–180.

Xiangru Tang, Anni Zou, Zhuosheng Zhang, Ziming Li, Yilun Zhao, Xingyao Zhang, Arman Cohan, and Mark Gerstein. 2024a. Medagents: Large language models as collaborators for zero-shot medical reasoning. Preprint, arXiv:2311.10537.

Xiangru Tang, Anni Zou, Zhuosheng Zhang, Ziming Li, Yilun Zhao, Xingyao Zhang, Arman Cohan, and Mark Gerstein. 2024b. Medagents: Large language models as collaborators for zero-shot medical reasoning. Preprint, arXiv:2311.10537.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 5635–5649.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2022. Self-consistency improves chain of thought reasoning in language models. arXiv preprint arXiv:2203.11171.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-thought prompting elicits reasoning in large language models. Preprint, arXiv:2201.11903.

Honglin Xiong, Sheng Wang, Yitao Zhu, Zihao Zhao, Yuxiao Liu, Linlin Huang, Qian Wang, and Dinggang Shen. 2023. Doctorglm: Fine-tuning your chinese doctor is not a herculean task. arXiv preprint arXiv:2304.01097.

Ming Xu. 2023. Medicalgpt: Training medical gpt model. https://github.com/shibing624/MedicalGPT.

Aiyuan Yang, Bin Xiao, Bingning Wang, Borong Zhang, Ce Bian, Chao Yin, Chenxu Lv, Da Pan, Dian Wang, Dong Yan, et al. 2023a. Baichuan 2: Open large-scale language models. arXiv preprint arXiv:2309.10305.

Dingkang Yang, Jinjie Wei, Dongling Xiao, Shunli Wang, Tong Wu, Gang Li, Mingcheng Li, Shuaibing Wang, Jiawei Chen, Yue Jiang, et al. 2024. Pediatricsgpt: Large language models as chinese medical assistants for pediatric applications. arXiv preprint arXiv:2405.19266.

Songhua Yang, Hanjie Zhao, Senbin Zhu, Guangyu Zhou, Hongfei Xu, Yuxiang Jia, and Hongying Zan. 2023b. Zhongjing: Enhancing the chinese medical capabilities of large language model through expert feedback and real-world multi-turn dialogue. Preprint, arXiv:2308.03549.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. 2023. Tree of thoughts: Deliberate problem solving with large language models. Preprint, arXiv:2305.10601.

Hongbo Zhang, Junying Chen, Feng Jiang, Fei Yu, Zhihong Chen, Jianquan Li, Guiming Chen, Xiangbo Wu, Zhiyi Zhang, Qingying Xiao, et al. 2023. Huatuogpt, towards taming language model to be a doctor. arXiv preprint arXiv:2305.15075.

(Fan et al., 2024)

A Implementation Details of MIRPF-dataset Construction

Our Verbose Medical Intent Q&A dataset consists of data constructed from over 600 real-world medical consultation records (Fan et al., 2024). These records encompass complete medical consultation processes, including chief complaints, present illness history, past medical history, physical examination, auxiliary examination, preliminary diagnosis, diagnostic basis, differential diagnosis, treatment process, diagnostic results, and analytical summaries.

For dataset construction, we follow a systematic approach:

1. Manual Intent Annotation: We manually annotate each case with intent labels to ensure interpretability. Our annotation process focuses on identifying the rich, multi-intent nature of medical consultations where patients often express multiple medical needs within a single dialogue.

2. Representative Scenario Selection: Based on the annotated cases, we identify typical clinical scenarios that contain multiple medical intents. These scenarios serve as representative examples of complex medical consultations where patients express various medical needs simultaneously.

3. Data Enhancement and Expansion: We employ GPT-4o along with In-Context Learning (ICL) techniques to construct additional samples based on the identified intent labels. This approach allows us to maintain clinical relevance while expanding our dataset to cover a broader range of medical consultation scenarios. The constructed samples inherit the multi-intent characteristics of the original cases while introducing natural variations in expression and context. Through this process, we construct 2,200 samples, with each sample containing an average of 3.7 intents.

For comparison, we also evaluate our framework on the Verbose Huatuo Q&A dataset. Given that the intents in this dataset are primarily concentrated in symptom analysis and treatment recommendation categories and are more straightforward in nature, we utilize GPT-4 with prompt-guided intent recognition. The analysis reveals that each sample in this dataset contains an average of 2.1 intents.

B Prompt Templates

B.1 Intention Recognition

Intention Extraction Setting

```
Given the following medical query:
{{MEDICAL_QUERY}}
Given the Available INTENT_LIST:
{{BASE_INTENTS}}

Extract the underlying medical intents by following these steps:
1. Identify all potential medical intents from INTENT_LIST
2. For each identified intent:
  - Assign a confidence score (0-1)
  - Provide clinical reasoning for the score
3. Structure the output as:
  Intent: [INTENT]
  Confidence: [SCORE]
  Reasoning: [CLINICAL_RATIONALE]

Output only intents with confidence scores ≥ 0.8.
Ensure each intent is:
- Clinically relevant
- Distinct from other identified intents
- Supported by information in the query
```

Context-Aware Decomposition

```
Original Query: {{MEDICAL_QUERY}}
Target Intent: {{INTENT}} | Confidence: {{SCORE}}

Required Clinical Parameters:
[Patient] Age, Gender, History | [Symptoms] Primary, Duration, Severity, Changes |
[Context] Vitals, Medications, Allergies, Risks, Treatments |
[Intent-Specific] [DYNAMIC_PARAMETERS_BASED_ON_INTENT]

Construction Guidelines:
Maintain clinical precision | Preserve temporal relationships |
 Include relevant parameters | Use medical terminology | Specify urgency

Output Format:{ "intent": "[INTENT]",  "sub_query": { "clinical_focus": "",
"key_parameters": {}, "temporal_context": "", "urgency_level": "[1-5]"}}
```

Sub-Query Validation

```
Original Query: {{MEDICAL_QUERY]}}| Sub-queries: {{SUB_QUERY_LIST}}

Validation Criteria:
1. Content Check: Parameters [Coverage | Accuracy | Context | Urgency]
2. Similarity Analysis: Calculate Jaccard similarity between pairs (Flag if > 0.2)
3. Priority Assessment: Based on clinical urgency indicators

Emergency Scale & Indicators:
[5] Immediate: Life-threatening symptoms, Critical vitals
[4] Urgent: Requires prompt attention, Severe symptoms
[3] Semi-urgent: 24hr window, Moderate concerns
[2] Non-urgent: 72hr window, Mild symptoms
[1] Routine: General care, Follow-up

Output Format:{
"validation_results": [{"sub_query_id": "[ID]", "content": {"completeness": [SCORE],
"accuracy": [SCORE]},"similarity": {"overlap": [SCORE], "flags": []},"priority":
{"level": [1-5], "rationale": ""}}],"recommendations": {"revisions": []}}
```

B.2 Dynamic Intent Orchestration

Urgency Assessment Prompt

Role: You are a medical priority assessment system.

Task: Evaluate the urgency of the following medical intent on a scale of 0 to 1.

Context:
- Patient current vitals: {vitals_data}
- Medical intent: {intent_description}
- Clinical indicators: {clinical_data}
- Recent medical history: {history_data}

Guidelines for scoring:
- 0.0-0.2: Non-urgent, can be safely delayed
- 0.2-0.4: Low urgency, routine care
- 0.4-0.6: Moderate urgency, should be addressed soon
- 0.6-0.8: High urgency, requires prompt attention
- 0.8-1.0: Critical, immediate action required

Dependency Impact Prompt

Role: You are a medical workflow dependency analyzer.

Task: Calculate the dependency impact score of this medical intent on a scale of 0 to 1.

Context:
- Medical intent: {intent_description}
- Current medical workflow state: {workflow_state}
- Resource availability: {resource_data}
- Concurrent procedures: {concurrent_procedures}

Guidelines for scoring:
- 0.0-0.2: Almost independent, minimal dependencies
- 0.2-0.4: Low dependency, few prerequisites
- 0.4-0.6: Moderate dependencies, some coordination needed
- 0.6-0.8: High dependency, significant coordination required
- 0.8-1.0: Critical dependency, blocks multiple procedures

Contextual Priority Prompt

Role: You are a medical context priority evaluator.

Task: Assess the contextual priority of this medical intent on a scale of 0 to 1.

Context:
- Medical intent: {intent_description}
- Time of day: {current_time}
- Department status: {department_status}
- Staff availability: {staff_data}
- Hospital protocols: {protocol_data}
- Patient preferences: {patient_preferences}

Guidelines for scoring:
- 0.0-0.2: Low contextual importance
- 0.2-0.4: Basic contextual consideration
- 0.4-0.6: Moderate contextual priority
- 0.6-0.8: High contextual significance
- 0.8-1.0: Critical contextual priority

B.3 Intention-Specific Agents

Diagnosis & Assessment Agent Prompt

Role: You are a medical diagnosis and assessment specialist.

Context:
- Patient Information: {patient_info}

Reasoning Steps Required:
1. Analyze presented symptoms and clinical data
2. Compare with known medical patterns
3. Consider differential diagnoses
4. Evaluate confidence level of assessment

Treatment-Care Agent Prompt

Role: You are a medical treatment specialist.

Context:
- Patient Information: {patient_info}
- Current Medications: {medications}
- Allergies: {allergies}
- Treatment History: {treatment_history}

Reasoning Steps Required:
1. Review current diagnosis and treatment plan
2. Evaluate treatment options
3. Consider contraindications
4. Assess risk-benefit ratio

Recovery-Support Agent Prompt

Role: You are a recovery and rehabilitation specialist.

Context:
- Patient Information: {patient_info}

Reasoning Steps Required:
1. Assess recovery progress
2. Identify support needs
3. Plan rehabilitation steps
4. Consider psychological factors

Lifestyle-Guide Agent Prompt

Role: You are a lifestyle and preventive care specialist.

Context:
- Patient Information: {patient_info}

Reasoning Steps Required:
1. Evaluate current lifestyle patterns
2. Identify risk areas
3. Develop practical recommendations
4. Set achievable goals

B.4 GPT-4o Automatic evaluation

Intent Comprehension Evaluation

Role: Medical Intent Evaluation System

Input:
- Reference Intent: {ref_intent}
- Response for Evaluation {Response}

Evaluation Scale (0.0-5.0, precision: 0.1):

[4.0-5.0]
- All explicit intents captured with comprehensive responses
- Effective analysis of implicit medical needs
- Deep medical understanding with proper prioritization

[3.0-3.9]
- Most core intents identified (max one missing)
- At least one implicit intent recognized
- Adequate medical context understanding

[2.0-2.9]
- Multiple core intents missing
- Limited implicit recognition
- Surface-level medical analysis

[1.0-1.9]
- Core intents misunderstood
- Irrelevant content included
- Poor medical alignment

[0.0-0.9]
- Major intent misunderstandings
- Unrelated needs addressed
- No valid medical analysis

Clinical Planning Evaluation

Role: Medical Care Coordination Evaluator

Input:
- Identified Intents: {intent_analysis}
- Response for Evaluation: {output}

Evaluation Scale (0.0-5.0, precision: 0.1):

[4.0-5.0] Clinical Excellence
- Comprehensive medical-psycho-social integration
- Clear priority tiers (urgent/immediate/long-term)
- Cross-domain risk assessment

[3.0-3.9] Standard Clinical Care
- Basic time-phased medical planning
- Key risk identification
- Simple follow-up structure

[2.0-2.9] Basic Care
- Isolated symptom management
- Inconsistent recommendations
- Unclear implementation sequence

[1.0-1.9] Substandard Care
- Missing safety protocols
- Poor emergency prioritization
- Questionable clinical advice

[0.0-0.9] Critical Failure
- Dangerous oversights
- No coherent planning
- Potentially harmful advice

Response Quality Evaluation

Role: Medical Response Quality Evaluator

Input:
- Response for Evaluation: {output}

Evaluation Scale (0.0-5.0, precision: 0.1):

[4.0-5.0] Clinical Excellence
- Deep analysis with case-based evidence
- Detailed medication protocols with safety measures
- Actionable plans with clear monitoring metrics

[3.0-3.9] Standard Clinical Care
- Adequate analysis lacking specific cases
- Basic medication guidance with partial details
- Implementable plans needing clarification

[2.0-2.9] Basic Care
- Superficial analysis without depth
- Oversimplified medication recommendations
- Vague implementation guidance

[1.0-1.9] Substandard Care
- Off-topic or irrelevant response
- Contains medical inaccuracies
- Impractical or unsafe recommendations

[0.0-0.9] Critical Failure
- Fabricated content
- Multiple medical errors
- Potentially harmful guidance

Algorithm 1 Medical Intent Dynamic Orchestration Algorithm

```
Require:
Medical intent queries $Q = \{q_1, q_2, ..., q_n\}$
Intent types $T = \{$examination, diagnosis, treatment, inquiry$\}$
Priority function $S(v) = \alpha U(v) + \beta D(v) + \gamma C(v)$
Ensure: Sorted sequence of medical intents
function ProcessMedicalIntents($Q$)
    Initialize empty priority queue $PQ$ and result list $L$
    Initialize indegree[$v$] for all $v \in Q$
    for each query $q \in Q$ do
        Classify intent type $t \in T$ for $q$
        Calculate $U(q)$ based on medical urgency
        Calculate $C(q)$ based on intent complexity
    end for
    for each pair of queries $(q_i, q_j) \in Q$ do
        if HasMedicalDependency($q_i, q_j$) then
            Add edge $q_i \to q_j$
            indegree[$q_j$] $\leftarrow$ indegree[$q_j$] $+ 1$
        end if
    end for
    for each query $q \in Q$ do
        if indegree[$q$] $= 0$ then
            Calculate $S(q)$
            $PQ$.enqueue($q, S(q)$)
        end if
    end for
    while $PQ \neq \emptyset$ do
        $q \leftarrow PQ$.dequeue()
        $L$.append($q$)
        for each dependent query $d$ of $q$ do
            indegree[$d$] $\leftarrow$ indegree[$d$] $- 1$
            if indegree[$d$] $= 0$ then
                Calculate $S(d)$ based on current context
                $PQ$.enqueue($d, S(d)$)
            end if
        end for
        if medical context changes then
            Recalculate $S(v)$ for all $v \in PQ$
            $PQ$.reheapify()
        end if
    end while
    return $L$
end function
function HasMedicalDependency($q_i, q_j$)
    if $q_j$.type $=$ treatment $\land$ $q_i$.type $=$ diagnosis then
        return true
    end if
    if $q_j$.type $=$ diagnosis $\land$ $q_i$.type $=$ examination then
        return true
    end if
    return CheckContextDependency($q_i, q_j$)
end function
```

Algorithm 2 Hierarchical Progressive Decision-Making Agent (HPDMA)

Require:
Patient context C_p
Knowledge graph K with medical ontologies
Pre-activated intents $I = \{i_1, i_2, ..., i_n\}$ ▷ Pre-activated intent set
Intent-agent mapping $M_A : I \to A$ ▷ Mapping from intents to agents
LLM model L
Ensure: Final medical decision D_{final}
function GET ACTIVATED AGENTS(I, M_A)
  Initialize agent set $A_{active} = \{\}$
  for intent $i \in I$ **do**
    $a \leftarrow M_A(i)$ ▷ Get agent corresponding to intent
    $A_{active} \leftarrow A_{active} \cup \{a\}$
  end for
  return A_{active}
end function
function GENERATE KGFS PROMPT(C_p, K, I)
  Initialize prompt set P
  for intent $i \in I$ **do**
    $T \leftarrow$ Extract Triplets(K, C_p, i) ▷ Extract triplets based on intent
    $P \leftarrow P\cup$ Form Prompt(T, C_p, i)
  end for
  return P_{KGFS}
end function
function EXECUTE AGENTS($C_p, K, P_{KGFS}, A_{active}$)
  Initialize decisions $D = \{\}$
  for agent $a \in A_{active}$ **do**
    $d \leftarrow a(C_p, K, P_{KGFS})$ ▷ Execute agent decision
    $w \leftarrow$ Get Agent Weight(a) ▷ Get agent weight
    $D \leftarrow D \cup \{(d, w)\}$
  end for
  return D
end function
function HPDMA(C_p, K, I, M_A, L)
  $A_{active} \leftarrow$ Get Activated Agents(I, M_A)
  $P_{KGFS} \leftarrow$ Generate KGFS Prompt(C_p, K, I)
  $D \leftarrow$ Execute Agents$(C_p, K, P_{KGFS}, A_{active})$
  $D_{integrated} \leftarrow$ Integrate Decisions(D)
  $D_{final} \leftarrow L(D_{integrated}, P_{KGFS}, C_p)$ ▷ LLM final integration
  return D_{final}
end function
function INTEGRATE DECISIONS(D)
  Initialize $D_{integrated} = 0$
  for $(d, w) \in D$ **do**
    $D_{integrated} \leftarrow D_{integrated} + w \cdot d$
  end for
  return $D_{integrated}$
end function

Step 1. Intent-driven Query Decomposition

Original Medical Query

My father is heart attack stent surgery, the past two days suddenly chest tightness and shortness of breath, just measured the blood pressure 160/95, but he said he did not want to go to the hospital, just want to ask if you want to play nitroglycerin. He is diabetic and his blood sugar was measured at 16 in the morning, and the metformin he usually takes was not taken yesterday.

Medication Safety Review

Are there any risks or contraindications for my father taking nitroglycerin given his diabetic condition and his current blood sugar level of 16, especially since he missed his dose of metformin yesterday?

Symptom Analysis

What might be the cause of my father's sudden chest tightness and shortness of breath, given his recent heart attack and stent surgery?

Medication Consultation

Is it advisable for my father to take nitroglycerin given his current symptoms of chest tightness and shortness of breath, considering his medical history of heart attack and stent surgery?

Treatment Recommendations

Considering my father's recent heart attack and stent surgery, along with his symptoms of chest tightness and shortness of breath, should he use nitroglycerin at home?

Emergency Guidance

Although my father, who had a recent heart attack and stent surgery, refuses to go to the hospital, his blood pressure is at 160/95. What immediate steps should be taken if his symptoms worsen?

Step 2. Dynamic Intent Orchestration & Multi-Agent Activation

Lifestyle-Guide Agent

CRITICAL ALERT - EMERGENCY MEDICAL SITUATION
Primary Assessment:
- Blood Pressure: 160/95 - Severely elevated
- Symptoms: Chest tightness, shortness of breath
- Recent History: Heart attack, stent placement within past month
- Compounding Factors: Missed metformin, blood sugar 16

IMMEDIATE ACTIONS:
1. Call Emergency Services (911/Local Emergency Number):
   - Stay on the line with dispatcher
   - Have someone meet ambulance at entrance if possible
   - Keep doors unlocked for emergency access

2. While Waiting:
   - Sit or lie in semi-reclined position
   - Loosen any tight clothing
   - Stay still, minimize physical activity
   - Take slow, deep breaths if possible
   - Do not eat or drink anything

3. Prepare for EMT Arrival:
   - Gather current medications
   - Have medical history documents ready
   - Recent hospital discharge papers
   - List of current doctors

Diagnosis & Assessment Agent

SYMPTOM SEVERITY ANALYSIS:
Current Symptoms:
1. Chest Tightness
   - Location: [Request patient to specify]
   - Duration: [Request onset timing]
   - Character: [Request description - pressure/sharp/burning]
   - Radiation: [Check if pain spreads to arms/jaw/back]
   - Associated symptoms: Already noted shortness of breath

2. Shortness of Breath
   - Severity: [Request rating 1-10]
   - Position impact: [Check if worse lying down]
   - Activity level when occurred: [Request details]

Risk Factor Analysis:
a) Recent Cardiac History:
   - Post-stent placement (< 1 month)
   - High risk for:
     * Acute stent thrombosis
     * Early restenosis
     * Incomplete endothelialization

b) Current Vital Signs:
   - BP 160/95
     * 33% above normal systolic
     * 36% above normal diastolic
   - Indicates severe hypertensive state

c) Metabolic Status:
   - Blood glucose: 16
   - Critically low, risk for:
     * Cognitive impairment
     * Seizures
     * Loss of consciousness

Treatment-Care Agent

MEDICATION AND TREATMENT ANALYSIS:

1. Current Medication Status:
   a) Metformin:
      - Missed dose noted
      - Impact on glucose control
      - Do NOT attempt to "catch up" doses

   b) Post-Stent Medications:
      - [Request verification of]:
        * Antiplatelet therapy adherence
        * Beta-blocker status
        * ACE inhibitor/ARB compliance
        * Statin usage

2. Emergency Care Instructions:
   DO NOT:
   - Take additional medications without medical guidance
   - Attempt to self-correct blood sugar
   - Drive yourself to hospital
   - Wait to see if symptoms improve

   DO:
   - Keep nitroglycerin accessible but await EMS guidance
   - Maintain current position unless directed otherwise
   - Have someone stay with you until EMS arrives
   - Keep phone charged and nearby

3. Post-Emergency Follow-up Plan:
   - Immediate cardiology notification
   - Diabetes management review
   - Medication timing adjustment
   - Consider cardiac rehab evaluation

Step 3. Decision Making

Comprehensive Medical Advisory

Emergency Situation & Immediate Action
Your current situation constitutes a serious medical emergency requiring immediate attention. The combination of chest tightness and elevated blood pressure (160/95) following your recent heart attack and stent placement indicates a potentially critical cardiac situation. Call emergency services (911) immediately - this is not a situation that can be safely managed at home. While waiting for help, maintain a semi-reclined position and have someone stay with you.

Current Medical Risks
Your presentation combines two serious medical concerns: cardiac distress and severe metabolic imbalance. The chest tightness post-stent could indicate complications like stent thrombosis or restenosis. Your critically low blood sugar (16) and missed metformin dose creates additional risks including potential loss of consciousness or seizures. This combination of conditions significantly increases your overall risk and requires immediate professional intervention.

Interim Care Instructions
Until emergency services arrive, focus on staying still and calm. Do not take any medications, including nitroglycrine, without emergency medical guidance. Gather your current medications and recent medical documents for the emergency responders. Keep your phone charged and nearby, and ensure emergency services can access your location. Do not attempt to eat, drink, or correct your blood sugar independently.

Post-Emergency Care Needs
After immediate stabilization, you'll need comprehensive follow-up care. This event calls for an urgent review of both your cardiac and diabetes management plans. Your medication regimen will likely need adjustment to prevent future episodes, particularly focusing on the timing and coordination between your cardiac and diabetes medications. Both your cardiologist and endocrinologist should be involved in revising your care plan.

Long-term Management Strategy
Moving forward, you'll need a more integrated approach to managing your cardiac and metabolic conditions. This includes developing a more manageable medication schedule, establishing clear emergency protocols, and potentially adjusting your monitoring requirements for both blood pressure and blood sugar. Regular check-ins with your healthcare team will be crucial to prevent similar episodes in the future.

Critical Safety Note
With your medical history, any chest discomfort or breathing difficulties, especially when combined with abnormal vital signs, should be treated as an emergency. Don't wait to see if symptoms improve or try to manage them at home. Early intervention is crucial for preventing serious complications from either your cardiac or metabolic conditions.

The severity of your current symptoms, combined with your recent cardiac history and dangerous blood sugar level, makes this an absolute emergency requiring immediate professional medical care. Please ensure emergency services have been contacted if you haven't already done so.