Knowledge-Aware Co-Reasoning for Multidisciplinary Collaboration

Xurui Li[1], **Haijiao Wang**[2], **Kaisong Song**[1*], **Rui Zhu**[3*], **Haixu Tang**[4]

[1] Alibaba Group, China, [2] DiDi Global, China,

[3] School of Medicine, Yale University, USA, [4] Indiana University, Bloomington, USA

leexurui@gmail.com, wanghaijiaoleisure@gmail.com, kaisong.sks@alibaba-inc.com,

rui.zhu.rz399@yale.edu, hatang@indiana.edu

Abstract

Large language models (LLMs) have shown significant potential to improve diagnostic performance for clinical professionals. Existing multi-agent paradigms rely mainly on prompt engineering, suffering from improper agent selection and insufficient knowledge integration. In this work, we propose a novel framework **KACR** (**K**nowledge-**A**ware **C**o-**R**easoning) that integrates structured knowledge reasoning into multidisciplinary collaboration from two aspects: (1) a reinforcement learning-optimized agent that uses clinical knowledge graphs to guide dynamic discipline determination; (2) a multidisciplinary collaboration strategy that enables robust consensus through integration of domain-specific expertise and interdisciplinary persuasion mechanism. Extensive experiments conducted on both academic and real-world datasets demonstrate the effectiveness of our method.

1 Introduction

Accelerated LLM development has demonstrated remarkable potential in various clinical decision support systems (Rajashekar et al., 2024), such as automated medical documentation (Maharjan et al., 2024), intelligent diagnostic consultation (Tang et al., 2024), and AI-assisted report generation (Clusmann et al., 2023). These technological breakthroughs have enabled the proliferation of digital health platforms offering real-time consultation services, significantly enhancing treatment efficiency while substantially improving patient care experiences through optimized service delivery.

Previous single-agent paradigms face problems of knowledge limitation and model robustness. Recent multi-agent paradigms simulate clinical decision-making processes (Tang et al., 2024; Kim et al., 2024), but still face two crucial limitations in clinical applications: (1) The absence of precise patient guidance procedure forces excessive reliance on prompt engineering for discipline selection, posing substantial challenges especially in the presence of complications. (2) Inadequate interdisciplinary coordination and deficient evidence-based validation result in superficial diagnostic reasoning for clinical conclusions.

To address these challenges, a promising solution lies in the integration of external knowledge graphs (KGs) to enhance the professional competence, credibility, and interpretability of LLM agents through structured knowledge infusion (Pan et al., 2024). Although existing methods such as Think-on-Graph (ToG) (Sun et al., 2024) and Reasoning-on-Graph (RoG) (Luo et al., 2024) have demonstrated the potential of LLM-based agents for KG reasoning tasks, significant challenges persist in clinical applications. Specifically, accurately identifying relevant disciplines within clinical knowledge graphs remains particularly challenging without domain-specific training. Recent advancements exemplified by OpenAI-O1 (Zhong et al., 2024) and DeepSeek-R1 (Guo et al., 2025) demonstrate that reinforcement learning frameworks incorporating rule-based reward mechanisms can effectively optimize multi-step reasoning performance. However, current research efforts have not sufficiently explored the application of such reinforcement learning paradigms to improve LLMs' layer-by-layer reasoning capabilities on KGs.

In this paper, we present a novel framework KACR that advances clinical decision-making through structured knowledge-based reasoning. Our principal contributions are as follows:

- We propose the KACR framework to enhance multidisciplinary collaboration. This design emulates comprehensive clinical consultation through dual-phase clinical reasoning: patient guidance and multidisciplinary consultation.

* Corresponding authors.

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 13604–13620
November 4-9, 2025

- We introduce a novel reinforcement learning method for training the dynamic discipline determination module, enabling context-aware traversal by an LLM-based Actor and estimating reasoning state with a graph convolutional network-enhanced Critic.
- We propose a multidisciplinary knowledge-based collaboration through confidence-enhanced deliberations. A robust consensus can be quantified by the synergistic integration of domain-specific expertise and interdisciplinary persuasion through the context of shared reasoning graph.
- Extensive experiments conducted on eight benchmark datasets show that our KACR outperforms state-of-the-art clinical LLMs and multi-agent frameworks. [1]

2 Preliminary

The architecture of KACR is shown in Fig. 1 (a), including two modules: the reinforcement learning-optimized *discipline agent reasoning module* and the knowledge-anchored *multidisciplinary collaboration module*, which are conducted sequentially.

Step I: We start by training the *discipline agent reasoning module* using the Proximal Policy Optimization (PPO) (Schulman et al., 2017) method. This module employs an Actor-Critic architecture where the *Actor* dynamically identifies question-relevant nodes on the clinical knowledge graph, while the *Critic* evaluates the quality of node selections through value estimation. Unlike conventional approaches, our method implements LLM-powered reasoning on the clinical knowledge graph to determine relevant disciplines, concurrently generating a reasoning subgraph that contextualizes disciplinary roles through structured semantic relationships. This graph-aware selection mechanism enhances diagnostic relevance through explicit knowledge grounding.

Step II: Subsequently, the *multidisciplinary collaboration module* coordinates multiple rounds of interdisciplinary discussions among selected clinical agents. This collaborative process includes two key innovations: 1) an uncertainty quantification mechanism that dynamically weights agents' opinions based on their prediction confidence; 2) a KG-enhanced confidence reflection mechanism where agents iteratively refine their positions through evidence-based persuasion. Persuasion strength is explicitly measured using the value estimates from the *Critic* trained in Step I, ensuring that consensus-building is aligned with the underlying clinical knowledge structure. This dual mechanism effectively balances specialized expertise with collective intelligence during differential diagnosis.

Knowledge Graph. The clinical knowledge graph (CKG), denoted as $\mathcal{G}$, comprises three core components: entity set $\mathcal{V}$, structural relation set $\mathcal{E}$, and relation type set $\mathcal{R}$. The entities are classified into three distinct categories: symptom entities $\mathcal{V}_s$, disease entities $\mathcal{V}_d$, and discipline entities $\mathcal{V}_c$. Each relation is formally represented as a triplet (v_i, r, v_j), where v_i (head entity) and v_j (tail entity) are interconnected through the relation type $r \in \mathcal{R}$.

3 Methodology

Briefly, the pipeline consists of two stages: (1) Discipline reasoning training: PPO-based RL with exclusive LoRA fine-tuning on the Actor, updating RGCN/MLP in the Critic; (2) Collaborative inference: Frozen LLM for multi-agent consultations using Stage I's reasoning subgraphs, with training-free discussion via confidence-based voting.

3.1 Discipline Agent Reasoning Module

In **Step I**, the discipline reasoning module uses an *Actor* to identify relevant disciplines for a given question q through iterative layer-by-layer searches on the CKG. The Actor contains rule-based heuristic strategies and an LLM-based Actor Net. The Actor adheres to a predefined meta-path to implement iterative search: "$Question \rightarrow (Symptom \leftrightarrow Disease \leftrightarrow Discipline)_{\times N}$", where $(\cdot)_{\times N}$ indicates the number of iterations. The initial symptom entities are extracted from the question q, and a trajectory of $T = 3N$ action steps is required to collect the most relevant nodes in three types: symptoms, diseases, and disciplines. The Actor Net utilizes a frozen LLM backbone augmented with trainable Low-Rank Adaptation (LoRA) parameters (Hu et al., 2021) for text generation.

3.1.1 Actor with exploration and pruning

At the t-th step ($1 \leq t \leq T$), the Actor sequentially executes two operations: exploration and pruning. In the exploration phase, the Actor traverses along a predefined meta-path and examines first-order neighboring nodes, thereby constructing a candidate set. Subsequently, during the pruning phase,

[1] Main work done when working at Alibaba: https://anonymous.4open.science/r/KACR_RL-2B64

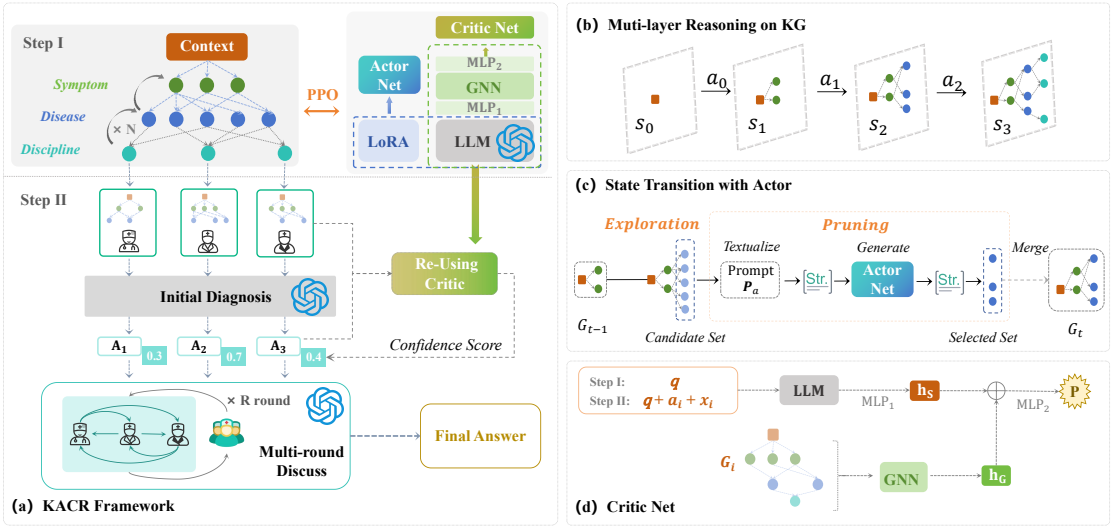

Figure 1: (a) The overall architecture of KACR. (b) Illustration of reasoning on KG layer-by-layer for symptom/disease/discipline nodes determination. (c) State transitions via Actor. (d) Details for Critic Net.

the Actor takes the action of selecting the most relevant nodes from the candidates and, together with the previously selected $\mathcal{G}_{t-1}$, forms a new subgraph $\mathcal{G}_t \subseteq \mathcal{G}$. The maximum number for disciplines is K, and that is $K^{'}$ for symptoms and diseases.

The iterative process addresses two primary challenges: 1) ambiguous patient queries that hinder effective symptom extraction; 2) inadequate symptom information for precise diagnosis. As illustrated in Fig. 2, the initial symptom extraction from the patient's statement "*I have been...flowers*" yields *Fatigue* and *Cough*. Following the forward inference path "$Symptom \rightarrow Disease \rightarrow Discipline$", *Pneumoni* emerges as a relevant diagnosis due to its established associations with these symptoms. Subsequently, this identification leads to the exploration of *Pulmonology* through discipline layer analysis. To enhance diagnostic completeness, a reverse inference path "$Discipline \rightarrow Disease \rightarrow Symptom$" is subsequently executed from the *Pulmonology* node. This bidirectional search strategy successfully identifies additional relevant nodes: disease *COVID-19* and its associated symptom *Dysgeusia*. The query-specific candidate set ultimately comprises symptom entities $\{Fatigue, Cough, Dysgeusi\}$, disease entities $\{Pneumoni, COVID\text{-}19\}$, and discipline entities $\{Pulmonology\}$. Notably, the identification of *Dysgeusia* during secondary iteration demonstrates the framework's capability to resolve ambiguous descriptions - this particular symptom aligns more accurately with the patient's metaphorical expression "*I can't even smell the fragrance of flowers*" than the initially reported symptoms.

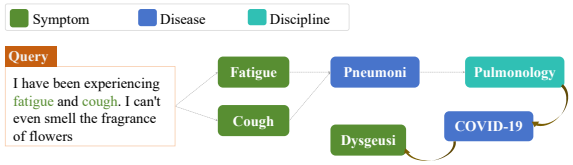

Figure 2: Illustration for iterative searching on KG.

3.1.2 Critic with text & topological info

The *Critic* estimates the state value by comprehensively analyzing the newly obtained subgraph $\mathcal{G}_t$ following the execution of the action. As shown in Figure 1 (d), the Critic's network architecture integrates both structural and semantic features through a dual-stream processing framework. The implementation involves feeding the subgraph $\mathcal{G}_t$ into a Relational Graph Convolutional Network (RGCN) (Schlichtkrull et al., 2018) layer for hierarchical feature extraction.

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}_q}\sum_{j\in\mathcal{N}_i^r}\frac{1}{|\mathcal{N}_i^r|}\mathbf{W}_r^{(l)}\mathbf{h}_j^{(l)} + \mathbf{W}_0^{(l)}\mathbf{h}_i^{(l)}\right) \quad (1)$$

where $\mathbf{h}_i^{(l)}$ denotes the node representation of v_i at the l-th layer of the RGCN, $\mathcal{N}_i^r$ is the neighbors of node v_i under relation r, $\mathbf{W}_r^{(l)}$ and $\mathbf{W}_0^{(l)}$ are weight matrices, and $\sigma(\cdot)$ is a sigmoid function. Node representations are initialized using features derived from LLM. The subgraph representation for $\mathcal{G}_t$ obtained through global average pooling of all node representations from the output layer L,

formally expressed as:

$$\mathbf{h_G} = MeanPooling\Big(\{\mathbf{h}_i^{(L)} | 1 \leq i \leq |\mathcal{V}_q|\}\Big) \quad (2)$$

In addition to topological representation, we add a multi-layer perceptron MLP_1 layer at the final transformer block of the LLM to encode the question q into a vector $\mathbf{h_S}$, which maintains identical dimensionality with the graph representation $\mathbf{h_G}$. The subsequent fusion process involves computing the element-wise average of $\mathbf{h_S}$ and $\mathbf{h_G}$, followed by a transformation through a secondary multi-layer perceptron MLP_2 to generate the final probability output $p = MLP_2(\frac{\mathbf{h_S}+\mathbf{h_G}}{2})$. Note that the Critic Network shares the identical LLM backbone with the Actor Network, ensuring efficient training and computing resource utilization.

In Step I, The Critic assesses the quality of the Actor's reasoning path to facilitate more effective exploration and pruning of candidates during training, while in Step II, it evaluates the factual consistency for the generated clinical diagnoses.

3.1.3 Actor and Critic training with PPO

The neural network architecture is meticulously designed for effectiveness and performance balance. We use LoRA to achieve optimal performance with minimal computational overhead. To mitigate potential degradation of the Critic based on LLM itself, our method combines LLM with RGCN to achieve better performance in downstream tasks.

During training, only the RGCN and MLPs for the Critic and the LoRA parameters for the Actor are updated, making training efficient. The training procedure employs PPO (Huang et al., 2024) under the assumption of an episodic setting, which is an actor-critic method that optimizes the policy based on the accumulative reward with advantage function. Let $D_k = \{(s_t, a_t, r_t, s_{t+1})\}_{t=1}^{T}$ be a trajectory that consists of a sequence of transitions. s_t is the state at step $t \in [0, T]$, formed by subgraph $\mathcal{G}_{t-1}$. The action a_t at step t represents the selection operation among the nodes of symptoms, disease, and disciplines within a dynamically evolving action space. Specifically, action includes using heuristic rules to explore candidate nodes and textualizing $\mathcal{G}_{t-1}$ to prompt the LLM to generate pruned nodes' information. By integrating $\mathcal{G}_{t-1}$ with the pruned node set parsed from the LLM output, the system progresses to the subsequent state s_{t+1} by constructing the updated subgraph $\mathcal{G}_t$. Here, r_t is the immediate reward for a_t, which returns 1 when at least one node in the selected subset matches the discipline labeled with ground-truth, and conversely assigns 0 when no alignment occurs.

Let θ denote the parameterization of the Actor network, where the importance sampling ratio is defined as the probability ratio between successive policy iterations: $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$. Concurrently, let ϕ parameterize the Critic network. Actor maximization and Critic minimization objectives in PPO can be formulated as:

$$\begin{aligned} L_{actor}^{(\theta)} &= \hat{\mathbb{E}}_t\Big[min\big(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1 \pm \epsilon)\hat{A}_t\big)\Big] \\ L_{critic}^{(\phi)} &= \hat{\mathbb{E}}_t\big(V_\phi(s_t) - \hat{R}_t\big)^2 \end{aligned} \quad (3)$$

where $\hat{\mathbb{E}}_t$ is the expectation and the clipping operator clip$(\cdot)$ regulates policy updates by constraining the probability ratio $r_t(\theta)$ within $[1-\epsilon, 1+\epsilon]$, thus preventing excessive deviation of the importance sampling ratio from 1. The cumulative future reward is $\hat{R}_t = \sum_{t'=t}^{T} \big(\gamma\big)^{t'-t} r_T$, where $\gamma \in [0, 1]$ is the reward discount factor. The advantage function $\hat{A}_t$ is computed through the generalized advantage estimation (GAE) framework (Zhao et al., 2024):

$$\hat{A}_t = \sum_{t'=t}^{T} \Big(\gamma\lambda\Big)^{t'-t} \Big(r_{t'} + \gamma V(s_{t'+1}) - V(s_{t'})\Big) \quad (4)$$

where λ is a hyper-parameter and $V(s_{t'})$ is a value function denoting the expected return at state $s_{t'}$.

3.2 Multidisciplinary Collaboration Module

In **Step II**, all participating agents employ an identical frozen LLM to facilitate multi-agent collaborative reasoning. For a user query q and its associated disciplinary nodes $\mathcal{V}_d$ identified in Step I, we instantiate $|\mathcal{V}_d|$ distinct disciplinary agents $\mathcal{A} = \{\mathcal{A}_i\}_{i=1}^{|\mathcal{V}_d|}$ to engage in a multi-round discussion. This module strengthens inter-agent collaboration through two principal mechanisms: 1) vertical domain-specific knowledge grounding to improve initial diagnostic precision, and 2) horizontal consensus formation by integrating knowledge derived from the reasoning subgraph into factual consistency check. The illustration is shown in Fig. 1 (a), following two phases: Initial diagnosis and expert discussion.

3.2.1 Initial diagnosis from separate agent.

The set of candidate diagnostic hypotheses for the question q can be represented as $\mathcal{O}_q$. For every discipline v_i node, we reverse the question-specific graph $\mathcal{G}_T$ from which v_i originates, to derive its reasoning subgraph $\mathcal{G}_T^i$. We textualize $\mathcal{G}_T^i$ into a reference description $\mathcal{R}_i$. For each domain-specific

agent $\mathcal{A}_i$ associated with discipline $v_i \in \mathcal{V}_d$, the initial diagnostic outcome can be formalized as:

$$\left(a_i^{(0)}, x_i^{(0)}, c_i^{(0)}\right) = \mathcal{A}_i\left(q, \mathcal{O}_q, \mathcal{R}_i | \mathbf{P}_g\right) \tag{5}$$

where $\mathbf{P}_g$ denotes a prompting strategy (see in Appendix), $a_i^{(0)}$ is the predicted diagnosis option, $x_i^{(0)}$ denotes the generated explanatory rationale, and $c_i^{(0)} \in [0, 1]$ quantifies the confidence level of this diagnostic decision. Prompting LLM to score its own output can cause Degeneration-of-Thought (DoT) (Liang et al., 2024) issues due to excessive self-overconfidence. To address this, we repurpose the well-trained Critic to assess the confidence score, using the discipline-specific reasoning subgraph $\mathcal{G}_T^i$ as a reference.

3.2.2 Multi-round expert discussion.

The multidisciplinary collaboration module proceeds to the R-round discussion phase. In the discussion round r, agent $\mathcal{A}_i$ uses a discussion prompt $\mathbf{P}_m$ to refine its diagnostic results. The primary discussion step for each agent $\mathcal{A}_i$ in round r can be expressed as:

$$\left(a_i^{(r)}, x_i^{(r)}, c_i^{(r)}\right) = \mathcal{A}_i\left(q, \mathcal{O}_q, \mathcal{R}_i, \mathcal{H}_i^{(r-1)} | \mathbf{P}_m\right) \tag{6}$$

Specifically, each round contains three stages:

1) Confidence assignment for all answers. For each round r, we use the Critic to estimate the confidence $c_i^{(r)}$ for each agent $\mathcal{A}_i$. This allows us to construct the grouped answer tuples $\mathcal{T}^{(r)} = \left\{(a_i^{(r)}, x_i^{(r)}, c_i^{(r)})\right\}_{i=1}^{|\mathcal{V}_d|}$ for all the agents $\mathcal{A}$.

2) Enhancing reasoning through inter-agent opinion integration. When an agent attempts to refine its cognitive process by assimilating perspectives from peer agents, we postulate that persuasive demonstrations capable of effectively shaping others' judgments can yield strategic advantages. Let $\mathcal{H}_i^{(r)}$ denote the dialogue history for agent i in round r, initialized as an empty set during the initial round. We implement a debate-oriented agent governed by $\mathbf{P}_d$ to iteratively refine the current solution $a_i^{(r)}$ into an updated version $a_i'^{(r)}$, leveraging the aggregated responses $\mathcal{T}^{(r)}$ collected among all agents. Concurrently, the agent produces a calibration rationale $x_i'^{(r)}$ that elucidates the reasoning adjustments, which is later attached to the respective dialogue history $\mathcal{H}_i^{(r)}$.

3) Consultation result aggregation. During the conclusion of each round r, the multidisciplinary collaboration module generates the consolidated response $\hat{a}^{(r)}$ through an adaptive weighting mechanism. Although conventional approaches including simple majority voting demonstrate practical viability, our empirical analysis reveals that the calibrated weight-based aggregation scheme demonstrates enhanced performance in cross-disciplinary scenarios. The formal implementation of this mechanism is formulated as below:

$$\hat{a}^{(r)} = \underset{o_j \in \mathcal{O}_q}{\arg\max} \sum_{i=1}^{|\mathcal{V}_d|} c_i^{(r)} \mathbb{I}\left(a_i'^{(r)} = o_j\right) \tag{7}$$

The indicator function $\mathbb{I}(\cdot)$ evaluates to 1 if the condition $a_i^{(r)} = o_j$ is satisfied, and 0 otherwise. The discussion continues for a maximum of R rounds or terminates prematurely when all agents achieve unanimous consensus. Details for the multidisciplinary collaboration are provided in Appendix.

4 Experiments And Results

4.1 Experimental Setting

Datasets. We evaluate our method based on various benchmark datasets. Primary experiments were conducted on the four most widely used medical datasets: **MedQA** (Jin et al., 2021), **MedMCQA** (Pal et al., 2022), **PubMedQA** (Jin et al., 2019), and **MMLU** medical topics (Hendrycks et al., 2020). These datasets include various medical questions, such as disease diagnosis, medication inquiries, and health advice. We also create a re-labeled sub-dataset for training the discipline reasoning module, using 2000 samples from the above datasets. Three experienced medical professionals annotate the appropriate disciplines for queries, with final labels determined by majority voting. The annotation shows high reliability, with a Fleiss's kappa of 0.86. The CKG is derived from the authoritative Unified Medical Language System (UMLS) (Amos et al., 2020), which is an authoritative resource published by the United States National Library of Medicine. It is continuously updated to ensure latest medical knowledge update, and we use its 2024AA version[2]. Note that our emphasis lies in enhancing multi-agent reasoning with external graph integration, not KG construction, and better KG alternatives can be involved.

To validate the generalizability of our established model in more complex clinical scenarios, we performed extended evaluations on four additional

[2] https://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html

Model	MedQA	MedMCQA	PubMedQA	MMLU							AVG
				an	ck	cm	cb	mg	pm	Avg.	
Galactica-120B	44.4	52.9	77.6	58.5	59.2	57.8	68.8	70.0	59.6	62.3	59.3
Clinical Camel-70B	53.4	47.0	74.3	62.2	69.8	67.0	79.2	69.0	71.3	69.7	61.1
PMC LLaMA-13B	56.4	56.0	77.9	61.5	63.0	52.6	59.7	70.0	64.3	61.8	63.0
Meditron-70B	60.7	65.1	80.0	62.7	72.3	62.8	82.5	77.8	77.9	72.6	69.6
Med42-70B	61.3	61.9	77.2	64.4	75.9	69.9	84.0	83.0	78.7	75.9	69.1
BiMediX-8×7B	62.8	62.7	80.2	74.1	78.9	68.2	86.1	85.0	80.5	78.8	71.1
Flan-PaLM-540B	67.6	57.6	75.2	71.9	80.4	76.3	88.9	74.0	83.5	79.1	69.6
Med-PaLM-540B	67.1	57.6	80.0	66.7	77.7	73.4	88.2	73.0	80.1	76.5	70.3
LLaMA3-8B	60.9	50.7	73.0	63.0	72.1	64.2	79.7	76.0	77.2	72.0	64.2
LLaMA3-70B	79.9	69.6	75.8	76.3	87.2	81.5	92.4	93.0	88.2	86.4	77.9
ChatGPT	64.0	68.7	73.4	64.4	78.5	84.7	76.0	82.0	74.0	76.6	70.7
GPT-4	81.4	72.4	75.2	80.0	86.0	76.9	95.1	91.0	93.0	87.0	79.0
†**KACR**$_{(LLaMA3-70B)}$	**90.4**	**81.8**	**86.5**	**89.2**	**88.9**	**87.6**	**96.1**	**91.8**	**95.8**	**91.6**	**87.6**

Table 1: Comparison with different vanilla clinical/general LLMs. † denotes models trained with PPO.

clinical datasets that reflect more intricate and real-world clinical scenarios. **DDXPlus** (Fansi Tchango et al., 2022) is a large-scale Electronic Health Record (EHR) dataset, which is used to test the performance of models in diagnosing complex and diverse medical conditions. **SymCat** (Al-Ars et al., 2023) is a synthetic dataset which includes symptom-condition samples according to ICD-10-CM. **JAMA** (Chen et al., 2025) includes real-world clinical cases collected from the JAMA Network Clinical Challenge archive. **Medbullets** (Chen et al., 2025) comprises USMLE Step 2/3 style questions collected from tweets since April 2022, which emulate common clinical scenarios. Table 9 summarizes the statistics of the datasets.

Implementation. We implement our method using PyTorch and conduct experiments on a server with 8 NVIDIA A100 80GB GPUs. We use the OpenAI API[3] to access ChatGPT and GPT-4. For all experiments, we determine hyperparameters through grid search. The Actor and Critic models are trained over maximum 50 epochs (mini-batch 8) at learning rate 5e-5, with warm-up and early-stop strategy. Training for the 70B model takes about an hour per thousand samples/epoch. The maximum input length for LLM is 8192 tokens. The hyperparameters for PPO is $\gamma = 0.9$, $\lambda = 0.95$ and $\epsilon = 0.2$. The temperature is set to 0.9 during PPO training, whereas it is set to zero during inference to ensure reproducibility. The reported results come from the following upper bound settings: discipline number $K = 5$, symptoms and diseases number $K^{'} = 10$, and the reasoning and collaboration rounds N and R are both 3. Details can be seen in computational study section. Each experiment is repeated 3 times, with a t-test result of $t \leq 0.005$.

[3] https://platform.openai.com/docs/overview

4.2 Baseline Comparisons

Comparison with different vanilla LLMs. We employ the PPO-trained †KACR$_{(LLaMA3-70B)}$ as our primary model unless otherwise specified. We first compare its performance directly with various vanilla general/clinical LLM backbones in Table 1. *Galactica* is trained on a scientific corpus (Taylor et al., 2022). *Clinical Camel* incorporates question-answering data through a dialogue-based knowledge encoding process, transforming PubMed articles and MedQA into questions and detailed answers (Toma et al., 2023). Both *PMC-LLaMA* (Wu et al., 2023) and *Meditron* (Chen et al., 2023b) perform pretraining on PubMed content and clinical texts, followed by refinements on individual multiple-choice question answer datasets. *Med42* serves as an instruction-tuned LLaMA model designed for clinical tasks (Christophe et al., 2024). *BiMediX* is a bilingual clinical mixture of experts based on Mixtral-8×7B (Pieri et al., 2024). *Flan-PaLM* is the instruction-tuned variant of PaLM, while *Med-PaLM* is its resulting model with additional prompt parameters aligned with the clinical domain (Singhal et al., 2023). *LLaMA3* is a herd of language models and we use both *Llama-3-8B-Instruct* and *Llama-3-70B-Instruct* for comparison (Dubey et al., 2024). *GPT-4* (gpt-4-turbo) is a powerful LLM accessible via API service, and we report its 5-shot performance (Achiam et al., 2023). The experimental results in Table 1, apart from our KACR algorithm, are referenced from the original reports of baseline models. It can be seen that the performance of KACR surpasses all existing state-of-the-art medical LLM backbones. Additionally, the analysis on the six subcategories of MMLU demonstrates its consistent performance across different medical scenario problems.

Method	MedQA	MedMCQA	PubMedQA	MMLU
ReConcile*	70.9	54.2	77.4	78.2
CMD*	66.0	58.0	74.0	76.4
MAD	80.8	73.2	79.6	84.2
ChatEval	81.3	76.2	82.6	86.9
Debating	82.0	72.8	80.5	85.4
DyLAN	81.9	74.0	82.8	81.4
Medagents	83.0	78.0	83.1	89.1
MDAgents	88.7	79.8	75.0	90.2
PMC-LLaMA-13B	56.4	56.0	77.9	63.0
Meditron-70B	60.7	65.1	80.0	69.6
Qwen2.5-7B	63.2	51.4	70.8	73.1
Qwen2.5-72B	78.2	75.4	78.3	84.3
†KACR$_{(PMCLLaMA)}$	65.3	64.4	81.5	72.7
†KACR$_{(Meditron)}$	74.1	69.4	83.5	78.7
†KACR$_{(Qwen2.5-7B)}$	69.1	60.4	73.5	69.7
†KACR$_{(Qwen2.5-72B)}$	87.3	80.8	83.5	89.6
LLaMA3-70B$_{(SFT)}$	80.2	71.4	76.3	78.5
KACR$_{(GPT4)}$	86.3	80.3	84.4	90.4
KACR$_{(LLaMA3-8B)}$	65.1	54.3	74.4	75.3
†KACR$_{(LLaMA3-8B)}$	70.1	62.3	78.6	77.3
KACR$_{(LLaMA3-70B)}$	84.2	78.4	80.4	87.2
†KACR$_{(LLaMA3-70B)}$	**90.4**	**81.8**	**86.5**	**91.6**

Table 2: Comparison among various multi-agent methods. The baselines in the first part use GPT4 by default. "*" indicates methods based on mixture LLMs.

Comparison with multi-agent methods. We compare our method with leading multi-agent frameworks. **ReConcile** and **CMD** facilitate collaboration among multiple LLMs (Bard, Gemini and ChatGPT) as per their original settings. (Chen et al., 2023a; Wang et al., 2024). For other methods, we use GPT-4 as LLM backbone. **MAD** promotes divergent thinking via opposing perspectives (Smit et al., 2023). **ChatEval** uses three communication strategies, favoring simultaneous-talk-with-summarizer (Chan et al., 2023). **Debating** leverages non-expert judgment of expert debates to identify correct answers (Khan et al., 2024). **DyLAN** enables agents to interact for multiple rounds in a dynamic architecture with inference-time agent selection to improve performance (Liu et al., 2023). **Medagents** leverages role-playing LLM-based agents who participate in a collaborative discussion for the clinical domain, thereby enhancing LLM proficiency and reasoning capabilities (Tang et al., 2024). **MDAgents** (Kim et al., 2024) proposes adaptive collaboration frameworks for medical decision support, yet oversimplifies discipline reasoning and inter-agent coordination through heavy reliance on prompt engineering, resulting in performance limitations. As shown in the first block of Table 2, our method outperforms all baselines on MedQA, MedMCQA, PubMedQA and MMLU. To further validate the generalizability of our established model in more complex clinical scenarios, we show extended evaluations on more complex datasets such as DDXPlus, SymCat, JAMA and Medbullets in Fig. 3. We can find that our KACR also performs the best on most datasets in these scenarios. Compared to CMD and ReConcile, our approach enhances output diversity by incorporating role-specific knowledge. Unlike general frameworks such as ChatEval, Debating, and DyLAN, our method optimizes multidisciplinary consultations by integrating professional clinical knowledge and involving clinical experts. Compared to Medagents and MDAgents that heavily relies on prompt engineering, our KACR introduces additional CKG components with PPO-optimized reasoning agents and knowledge-grounding calibration, which enhances collaboration quality.

Applying KACR on different LLM backbones. We evaluate the KACR performance by applying it on different backbones such as Qwen2.5 (Yang et al., 2024), LLaMA3 and other two clinical LLMs. Since we focus on the joint training based on reinforcement learning and knowledge graph. The closed-source models accessed via API such as GPT-4 or Claude-3.5 Sonnet can only used as baselines instead of training backbones of our KACR. Here KACR$_{(GPT4)}$ indicates we only apply the GPT-4 on the inference framework without PPO training. We also compared LLaMA3-70B$_{(SFT)}$, which is finetuned using traditional SFT method on the training set, with only a slightly improvement compared to base LLaMA3-70B while underperforming †KACR$_{(LLaMA3-70B)}$. SFT methods struggles with graph reasoning as training data lacks explicit reasoning traces, inducing pattern over-fitting. Our framework resolves this via two approaches: (1) PPO-driven integration of structured rule rewards and metapath-guided exploration; (2) Actor-Critic agents' multi-layer CKG exploration for contextualized diagnoses via progressive evidence accumulation. It can be found that KACR improves the performance for different LLM backbones. We focus on training the reasoning ability of the LLM on KG for discipline selection, without developing its generation capabilities for discussions, as this is not our primary focus. Notably, although the original performance of GPT-4 exceeds that of LLaMA3-70B, our method †KACR$_{(LLaMA3-70B)}$ outperforms KACR$_{(GPT4)}$ because of well-crafted PPO training.

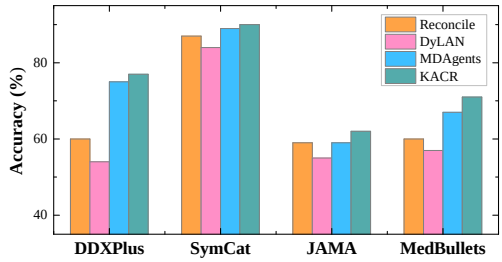

Figure 3: Comparisons on complex clinical datasets.

Method	MedQA	MedMCQA	PubMedQA
w/o CKG	83.0	78.0	83.1
w/o PPO	84.2	78.4	80.4
w/o confidence	88.2	80.2	84.3
LLM confidence	89.2	81.1	85.1
single round reasoning	88.9	80.7	84.6
single round discussion	88.7	79.6	84.8
†**KACR**$_{(LLaMA3-70B)}$	**90.4**	**81.8**	**86.5**

Table 3: Component contribution study.

4.3 Ablation Study

Component contribution study. To investigate the contributions of each component in KACR, we conduct an ablation study with several variants, as shown in Table 3. Following Medagents (Tang et al., 2024), **w/o CKG** removes the CKG-enhanced discipline reasoning module. **w/o PPO** uses the original LLM for discipline reasoning. **w/o confidence** and **LLM confidence** exclude the confidence-enhancement strategy or use the LLM for scoring instead of the Critic model. **single round reasoning** and **single round discussion** disable the iterative reasoning and discussion mechanisms, respectively. Overall, each component contributes positively to the performance.

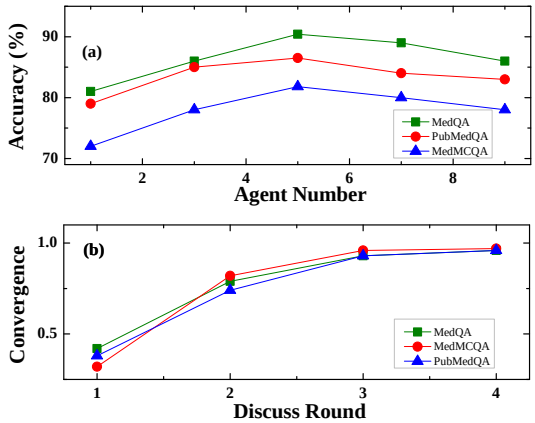

Figure 4: (a) Influence of maximum discipline agents number K on various datasets. (b) Convergence analysis for different discuss round r at 5-agent configuration.

Computational study. The impact of the number of discipline agents on various datasets is depicted in Fig. 4 (a), with all other parameters fixed. It can be observed that the optimal number of agents is 5, and this value remains consistent across different datasets. We then present the convergence patterns using the identified optimal 5-agent configuration. From Fig. 4 (b) we can see that basically more than 90% of the questions reached an agreement in the third round, and more than 95% of the questions reached an agreement in the fourth round. This finding aligns with the trends reported by other multi-agent frameworks (Kim et al., 2024).

5 Related Work

Multi-Agent Systems. Agents are a promising development in artificial intelligence, with early studies showing that assigning roles to LLMs significantly influences their output (Shanahan et al., 2023). Role playing introduces specific knowledge, making LLMs more interactive and capable of tackling complex tasks. However, LLM-based agents can introduce bias and instability, and may face DoT issues (Liang et al., 2024). Recent research on multi-agent collaboration aims to enhance LLM truthfulness by leveraging collective intelligence (Smit et al., 2023) and fostering divergent thinking for in-depth tasks. For example, Du et al. (2023) proposed a Society of Minds (SoM) where multiple agents share their answers to collaborate effectively. Chan et al. (2023) introduced ChatEval, featuring three communication strategies: one-by-one, simultaneous-talk, and simultaneous-talk-with-summarizer. However, role identity is often manually assigned or generated by LLMs, leading to errors from imprecise selections. Moreover, current multi-agent frameworks often overlook individual roles' self-evaluations. In contrast, we integrate LLMs with knowledge graphs to facilitate deeper reasoning.

LLM-KG Integrating Paradigm. LLMs often fail to capture and access factual knowledge due to their black-box characteristics. In contrast, knowledge graphs, which act as structured knowledge frameworks, can enhance LLMs by providing external knowledge and improving interpretability (Pan et al., 2024). Meanwhile, KGs struggle with complexity and updates, limiting new knowledge generation. Integrating LLMs with KGs synergizes their strengths, often by converting KG knowledge into prompts for LLMs. However, loose-

coupling restricts LLMs' role in graph reasoning. Fusion models like QA-GNN (Yasunaga et al., 2021) and GreaseLM (Zhang et al., 2022) combine LLMs and GNNs to jointly reason over text and graph knowledge. In addition, LLMs can also be treated as agents that interact with KGs to conduct reasoning. Think-on-Graph (ToG) (Sun et al., 2024) and Reasoning-on-Graph (RoG) (Luo et al., 2024) provide tight-coupling paradigms where KGs and LLMs work in tandem, complementing each other's capabilities in each step of graph reasoning. However, due to the complexity of graph patterns and target tasks, LLMs often struggle with accurate inference without task-specific optimization. Furthermore, studies rarely effectively integrate KGs into different clinical phases, including domestic knowledge provision, confidence scoring, and multi-agent collaboration.

Generalizability to Similar Scenarios. Our KACR framework can be easily generalized to various scenarios, such as law and finance. On the one hand, many domains have high-quality knowledge graphs, and users can freely define meta-paths by analogy with medical scenarios. On the other hand, each domain has different agent roles, who reach a consensus conclusion through discussion. Finally, knowledge graphs are not always necessary, and our framework will degenerate into a multi-agent collaborative framework without external KG. For example, a financial investment in the "new energy vehicle industry chain" can benefit from joint discussions by different expert agents (e.g., "automotive" and "battery" experts). A meta-path like "customer needs $\rightarrow$ related products $\rightarrow$ core companies $\rightarrow$ niche domains" can guide the selection of domain-specific expert agents.

6 Conclusion

Our work introduces KACR, a knowledge-aware framework that bridges structured clinical knowledge with multi-agent collaboration to advance diagnostic reasoning. By integrating PPO-optimized discipline reasoning and knowledge-anchored discussion via confidence enhancement, KACR alleviates critical limitations in multi-agent collaboration for complex clinical decision-making. Extensive experiments conducted on eight clinical benchmarks demonstrate that our method achieves the best performance.

7 Limitations

This study focuses on enhancing the reasoning ability of LLM through knowledge graph integration. Additional training could improve performance. Although hand-crafted prompts for knowledge node selection, clinical diagnosis, and multidisciplinary consultation are iteratively refined through multi-phase experimentation, their current implementations do not represent theoretically optimal configurations. Future investigations will prioritize the development of a systematic framework integrating knowledge graph embeddings and neural architecture search to automate prompt optimization, thus establishing a robust paradigm for dynamic prompt engineering in medical decision-support systems. In this paper, we adopt UMLS for its authoritative and semi-annual updates, ensuring sustained knowledge currency. Note that our implementation focuses on using KGs as contextual references for LLM agents rather than graph-construction itself, and better KG alternatives could further improve the performance.

8 Ethical Considerations

This study presents a decision-support framework conceptualized as a supplementary tool for healthcare professionals and end-users, providing data-driven insights to improve clinical decision-making processes. It is imperative to emphasize that diagnostic determinations and therapeutic interventions must remain in the purview of licensed medical practitioners. The model outputs should undergo rigorous clinical validation and be interpreted within comprehensive diagnostic contexts, as uncritical reliance on algorithmic recommendations without adequate human oversight may introduce clinical risks. The proposed system is conceived as complementary enhancements to, rather than substitutes for, professional medical judgment and domain expertise. All experimental resources (i.e., datasets, KGs and LLMs) utilize exclusively publicly accessible datasets derived from established medical repositories and research resources that have undergone extensive validation within the scientific community.

References

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*.

Zaid Al-Ars, Obinna Agba, Zhuoran Guo, Christiaan Boerkamp, Ziyaad Jaber, and Tareq Jaber. 2023. Nlice: Synthetic medical record generation for effective primary healthcare differential diagnosis. In *2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 397–402. IEEE.

Liz Amos, David Anderson, Stacy Brody, Anna Ripple, and Betsy L Humphreys. 2020. Umls users and uses: a current overview. *Journal of the American Medical Informatics Association*, 27(10):1606–1611.

Chi-Min Chan, Weize Chen, Yusheng Su, Jianxuan Yu, Wei Xue, Shanghang Zhang, Jie Fu, and Zhiyuan Liu. 2023. Chateval: Towards better llm-based evaluators through multi-agent debate. In *Proceedings of the ICLR*.

Hanjie Chen, Zhouxiang Fang, Yash Singla, and Mark Dredze. 2025. Benchmarking large language models on answering and explaining challenging medical questions. In *Proceedings of the NAACL: Human Language Technologies (Volume 1: Long Papers)*, pages 3563–3599.

Justin Chih-Yao Chen, Swarnadeep Saha, and Mohit Bansal. 2023a. Reconcile: Round-table conference improves reasoning via consensus among diverse llms. *arXiv preprint arXiv:2309.13007*.

Zeming Chen, Alejandro Hernández Cano, Angelika Romanou, Antoine Bonnet, Kyle Matoba, Francesco Salvi, Matteo Pagliardini, Simin Fan, Andreas Köpf, Amirkeivan Mohtashami, et al. 2023b. Meditron-70b: Scaling medical pretraining for large language models. *arXiv preprint arXiv:2311.16079*.

Clément Christophe, Praveen K Kanithi, Prateek Munjal, Tathagata Raha, Nasir Hayat, Ronnie Rajan, Ahmed Al-Mahrooqi, Avani Gupta, Muhammad Umar Salman, Gurpreet Gosal, et al. 2024. Med42–evaluating fine-tuning strategies for medical llms: Full-parameter vs. parameter-efficient approaches. *arXiv preprint arXiv:2404.14779*.

Jan Clusmann, Fiona R Kolbinger, Hannah Sophie Muti, Zunamys I Carrero, Jan-Niklas Eckardt, Narmin Ghaffari Laleh, Chiara Maria Lavinia Löffler, Sophie-Caroline Schwarzkopf, Michaela Unger, Gregory P Veldhuizen, et al. 2023. The future landscape of large language models in medicine. *Communications medicine*, 3(1):141.

Yilun Du, Shuang Li, Antonio Torralba, Joshua B Tenenbaum, and Igor Mordatch. 2023. Improving factuality and reasoning in language models through multiagent debate. In *Proceedings of the ICML*.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Arsene Fansi Tchango, Rishab Goel, Zhi Wen, Julien Martel, and Joumana Ghosn. 2022. Ddxplus: A new dataset for automatic medical diagnosis. *Advances in Neural Information Processing Systems*, 35:31306–31318.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2020. Measuring massive multitask language understanding. *arXiv preprint arXiv:2009.03300*.

Edward J Hu, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2021. Lora: Low-rank adaptation of large language models. In *Proceedings of the ICLR*.

Nai-Chieh Huang, Ping-Chun Hsieh, Kuo-Hao Ho, and I-Chen Wu. 2024. Ppo-clip attains global optimality: Towards deeper understandings of clipping. In *Proceedings of the AAAI*, volume 38, pages 12600–12607.

Di Jin, Eileen Pan, Nassim Oufattole, Wei-Hung Weng, Hanyi Fang, and Peter Szolovits. 2021. What disease does this patient have? a large-scale open domain question answering dataset from medical exams. *Applied Sciences*, 11(14):6421.

Qiao Jin, Bhuwan Dhingra, Zhengping Liu, William Cohen, and Xinghua Lu. 2019. Pubmedqa: A dataset for biomedical research question answering. In *Proceedings of the EMNLP-IJCNLP*, pages 2567–2577.

Akbir Khan, John Hughes, Dan Valentine, Laura Ruis, Kshitij Sachan, Ansh Radhakrishnan, Edward Grefenstette, Samuel R Bowman, Tim Rocktäschel, and Ethan Perez. 2024. Debating with more persuasive llms leads to more truthful answers. In *Proceedings of the ICML*.

Yubin Kim, Chanwoo Park, Hyewon Jeong, Yik S Chan, Xuhai Xu, Daniel McDuff, Hyeonhoon Lee, Marzyeh Ghassemi, Cynthia Breazeal, and Hae W Park. 2024. Mdagents: An adaptive collaboration of llms for medical decision-making. *Advances in Neural Information Processing Systems*, 37:79410–79452.

Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. 2024. Encouraging divergent thinking in large language models through multi-agent debate. In *Proceedings of the EMNLP*, pages 17889–17904.

Zijun Liu, Yanzhe Zhang, Peng Li, Yang Liu, and Diyi Yang. 2023. Dynamic llm-agent network: An llm-agent collaboration framework with agent team optimization. *arXiv preprint arXiv:2310.02170*.

Linhao Luo, Yuan-Fang Li, Reza Haf, and Shirui Pan. 2024. Reasoning on graphs: Faithful and interpretable large language model reasoning. In *Proceedings of the ICLR*.

Jenish Maharjan, Anurag Garikipati, Navan Preet Singh, Leo Cyrus, Mayank Sharma, Madalina Ciobanu, Gina Barnes, Rahul Thapa, Qingqing Mao, and Ritankar Das. 2024. Openmedlm: prompt engineering can out-perform fine-tuning in medical question-answering with open-source large language models. *Scientific Reports*, 14(1):14156.

Ankit Pal, Logesh Kumar Umapathi, and Malaikannan Sankarasubbu. 2022. Medmcqa: A large-scale multi-subject multi-choice dataset for medical domain question answering. In *Conference on health, inference, and learning*, pages 248–260. PMLR.

Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. 2024. Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*.

Sara Pieri, Sahal Shaji Mullappilly, Fahad Shahbaz Khan, Rao Muhammad Anwer, Salman Khan, Timothy Baldwin, and Hisham Cholakkal. 2024. Bimedix: Bilingual medical mixture of experts llm. *arXiv preprint arXiv:2402.13253*.

Niroop Channa Rajashekar, Yeo Eun Shin, Yuan Pu, Sunny Chung, Kisung You, Mauro Giuffre, Colleen E Chan, Theo Saarinen, Allen Hsiao, Jasjeet Sekhon, et al. 2024. Human-algorithmic interaction using a large language model-augmented artificial intelligence clinical decision support system. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–20.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*, pages 593–607.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *CoRR*, abs/1707.06347.

Murray Shanahan, Kyle McDonell, and Laria Reynolds. 2023. Role play with large language models. *Nature*, 623(7987):493–498.

Xiaoming Shi, Zeming Liu, Li Du, Yuxuan Wang, Hongru Wang, Yuhang Guo, Tong Ruan, Jie Xu, Xiaofan Zhang, and Shaoting Zhang. 2024. Medical dialogue system: A survey of categories, methods, evaluation and challenges. *Findings of the Association for Computational Linguistics ACL 2024*, pages 2840–2861.

Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, et al. 2023. Large language models encode clinical knowledge. *Nature*, 620(7972):172–180.

Andries Petrus Smit, Nathan Grinsztajn, Paul Duckworth, Thomas D Barrett, and Arnu Pretorius. 2023. Should we be going mad? a look at multi-agent debate strategies for llms. In *Proceedings of the ICML*.

Jiashuo Sun, Chengjin Xu, Lumingyuan Tang, Saizhuo Wang, Chen Lin, Yeyun Gong, Lionel Ni, Heung-Yeung Shum, and Jian Guo. 2024. Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph. In *Proceedings of the ICLR*.

Xiangru Tang, Anni Zou, Zhuosheng Zhang, Ziming Li, Yilun Zhao, Xingyao Zhang, Arman Cohan, and Mark Gerstein. 2024. Medagents: Large language models as collaborators for zero-shot medical reasoning. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 599–621.

Ross Taylor, Marcin Kardas, Guillem Cucurull, Thomas Scialom, Anthony Hartshorn, Elvis Saravia, Andrew Poulton, Viktor Kerkez, and Robert Stojnic. 2022. Galactica: A large language model for science. *arXiv preprint arXiv:2211.09085*.

Augustin Toma, Patrick R Lawler, Jimmy Ba, Rahul G Krishnan, Barry B Rubin, and Bo Wang. 2023. Clinical camel: An open expert-level medical language model with dialogue-based knowledge encoding. *arXiv preprint arXiv:2305.12031*.

Ferhat Tuncel, Basri Mumcu, and Senem Tanberk. 2021. A chatbot for preliminary patient guidance system. In *2021 29th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4. IEEE.

Qineng Wang, Zihao Wang, Ying Su, Hanghang Tong, and Yangqiu Song. 2024. Rethinking the bounds of llm reasoning: Are multi-agent discussions the key? *arXiv preprint arXiv:2402.18272*.

Chaoyi Wu, Xiaoman Zhang, Ya Zhang, Yanfeng Wang, and Weidi Xie. 2023. Pmc-llama: Further finetuning llama on medical papers. *arXiv preprint arXiv:2304.14454*, 2(5):6.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, et al. 2024. Qwen2. 5 technical report. *arXiv preprint arXiv:2412.15115*.

Michihiro Yasunaga, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. 2021. Qa-gnn: Reasoning with language models and knowledge graphs for question answering. In *Proceedings of the NAACL*, pages 535–546.

Xikun Zhang, Antoine Bosselut, Michihiro Yasunaga, Hongyu Ren, Percy Liang, Christopher D Manning, and Jure Leskovec. 2022. Greaselm: Graph reasoning enhanced language models. In *Proceedings of the ICLR*.

Hanyang Zhao, Wenpin Tang, and David Yao. 2024. Policy optimization for continuous reinforcement learning. *Advances in Neural Information Processing Systems*, 36.

Tianyang Zhong, Zhengliang Liu, Yi Pan, Yutong Zhang, Yifan Zhou, Shizhe Liang, Zihao Wu, Yanjun Lyu, Peng Shu, Xiaowei Yu, et al. 2024. Evaluation of openai o1: Opportunities and challenges of agi. *arXiv preprint arXiv:2409.18486*.

A Graph Description Template Exemplar

Graph Structure Description: In this heterogeneous graph, Symptoms nodes are connected to Diseases nodes through "Related to" edges, indicating the associations between symptoms and diseases. Diseases nodes are connected to Disciplines nodes through "Belongs to Discipline" edges, indicating the medical disciplines to which the diseases belong. This structure clearly demonstrates the complex relationships among symptoms, diseases, and disciplines.

Node Types:

- **Symptoms**: Represent various symptoms exhibited by patients, such as headache, fever, etc.
- **Diseases**: Represent specific types of diseases, such as influenza, pneumonia, etc.
- **Disciplines**: Represent medical disciplines, such as internal medicine, surgery, etc.

Edge Types:

- **Related to**: Connects Symptoms and Diseases, indicating the association between symptoms and diseases.
- **Belongs to Discipline**: Connects Diseases and Disciplines, indicating the medical discipline to which a disease belongs.

Node Sets:

- **Symptoms Node Set**: $S = \{\text{Symptom}_1, \text{Symptom}_2, \ldots, \text{Symptom}_L\}$
- **Diseases Node Set**: $D = \{\text{Disease}_1, \text{Disease}_2, \ldots, \text{Disease}_M\}$
- **Disciplines Node Set**: $C = \{\text{Discipline}_1, \text{Discipline}_2, \ldots, \text{Discipline}_K\}$

Edge Sets:

- **Related to Edge Set**: $E_{\text{related}} = \{(\text{Symptom}_i, \text{Disease}_j) \mid i \in [1, L], j \in [1, M]\}$
- **Belongs to Discipline Edge Set**: $E_{\text{belongs}} = \{(\text{Disease}_j, \text{Discipline}_k) \mid j \in [1, M], k \in [1, K]\}$

Table 4: Graph Description Template Exemplar.

#Background: You are a clinical assistant, and a user consults you with a clinical question: *{{question q}}*. From the question, we have initially extract potential symptoms *{{initial symptoms}}*

#Current reasoning subgraph: *{{$\mathcal{G}_t$ Description}}*

#Candidate node set for current step: From the current reasoning graph $\mathcal{G}_t$, we have explored from a clinical knowledge graph that the question is related to potential symptoms (or diseases / disciplines): *{{symptoms $\mathcal{V}_s$}}* $\big($or *{{diseases $\mathcal{V}_d$}}* / *{{disciplines $\mathcal{V}_c$}}*$\big)$.

#Instruction: Select *{{top-K}}* symptoms (or diseases / disciplines) from the candidate set $\mathcal{N}_t$ that are relevant to the question.

Table 5: $\mathbf{P}_a$: Exemplar prompt template for Actor Net to prune nodes from candidates.

#Background: You are a clinical assistant in *{discipline v_i}*, a user consults you with a clinical question: *{{question q}}*.

#References: There is a reference information which contains all the reasoning paths from question q to discipline v_i. The reference description is $\mathcal{R}_i$.

#Options: *{{options $\mathcal{O}_q$}}*

#Instruction: Given the references, you should conduct step-by-step reasoning and select the best answer from the options.

Table 6: $\mathbf{P}_g$: Exemplar prompt template for initial diagnose generation.

B Prompt Template Exemplar

#Background You are a clinical assistant in *{discipline v_i}*, a user consults you with a clinical question: *{{question q}}*.

#References: There is a reference information which contains all the reasoning paths from question q to discipline v_i. The reference description is $\mathcal{R}_i$. In addition, from last round of discussion, there is a *{{temporary consensus result $\hat{a}^{(r)}$}}*. You have also considered the diagnosis results provided by other experts in various disciplines, and put the calibrating explanation from the debater agent into your *{{chat history $\mathcal{H}_i^{(r)}$}}*.

#Options: *{{options $\mathcal{O}_q$}}*

#Instruction: Considering the references, please think step-by step, and select the best answer, with detailed explanation.

Table 7: $\mathbf{P}_m$: Exemplar prompt template for multi-round expert discussion.

#Background: You are a clinical debating assistant, there is a clinical question: *{{question q}}. Another clinical assistant in {discipline v_i} has initially give the {{temporary answer $a_i^{(r)}$}}, please help check the answer and calibrate it if needed.*

#References: There is a reference information which contains the temporary answers, explanations and confidence scores from all agents: *{{grouped information $\mathcal{T}^{(r)}$}}.*

#Options: *{{options $\mathcal{O}_q$}}*

#Instruction: Given the references, please think step-by-step, and select the most proper answer from the options, with your calibrating explanation .

Table 8: $\mathbf{P}_d$: Exemplar prompt template for the debater DEB.

.

C KG-enhanced Multidisciplinary Collaboration

The pseudo-code for multidisciplinary collaboration is shown in Algorithm 1.

Algorithm 1 KG-enhanced Multidisciplinary Collaboration

Require: Agents number $|\mathcal{V}_d|$, discuss turn R, a group of agents $\mathcal{A} = \{\mathcal{A}_i\}_{i=1}^{|\mathcal{V}_d|}$, chat history of each agent at each round r $\mathcal{H}^{(r)} = \{\mathcal{H}_i^{(r)}\}_{i=1}^{|\mathcal{V}_d|}$, debater DEB;
Ensure: Final answer ANS.

for $r \leftarrow 1, R$ **do**
  for $i \leftarrow 1, |\mathcal{V}_d|$ **do**
    $(a_i^{(r)}, x_i^{(r)}) \leftarrow \mathcal{A}_i(q, \mathcal{O}_q, \mathcal{R}_i, \hat{a}^{(r-1)}, \mathcal{H}_i^{(r-1)})$;
    $c_i^{(r)} \leftarrow Critic(q, a_i^{(r)}, x_i^{(r)}, \mathcal{G}_T^i)$;
    $\mathcal{T}^{(r)} \leftarrow \mathcal{T}^{(r-1)} + [(a_i^{(r)}, x_i^{(r)}, c_i^{(r)})]$;
  end for
  for $i \leftarrow 1, |\mathcal{V}_d|$ **do**
    $(a_i'^{(r)}, x_i'^{(r)}) \leftarrow DEB(a_i^{(r)}, \mathcal{T}^{(r)})$;
    $\mathcal{H}_i^{(r)} \leftarrow \mathcal{H}_i^{(r-1)} + [x_i'^{(r)}]$
  end for
  $\hat{a}^{(r)} = \arg\max_{o_j \in \mathcal{O}_q} \sum_{i=1}^{|\mathcal{V}_d|} c_i^{(r)} \mathbb{I}\left(a_i'^{(r)} = o_j\right)$
end for
$ANS \leftarrow \hat{a}^{(R)}$
return ANS

D PPO Algorithm

Our PPO optimization with clipped surrogate objective follows OpenAI's instruction [4], the pseudo-code is shown in Algorithm 2.

[4] https://spinningup.openai.com/en/latest/algorithms/ppo.html

Algorithm 2 PPO

Require: Initial policy parameters θ_0 and value function parameters and ϕ_0;
for $k = 0, 1, 2, ...$ **do**
  Collect set of trajectories $\mathcal{D}_k = \{\tau_i\}$ by running policy $\pi_k = \pi(\theta_k)$ in the environment.
  Compute rewards-to-go $\hat{R}_t$.
  Compute advantage estimates, $\hat{A}_t$ based on the current value function $V_{\phi k}$.
  Update policy by maximizing PPO-Clip objective:

$$\theta_{k+1} = \arg\max_{\theta} \frac{1}{\|\mathcal{D}_k\|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T} \Big[min\big(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1 \pm \epsilon)\hat{A}_t\big)\Big]$$

  , typically via stochastic gradient with Adam.
  Fit value function by regression on means-squared error:
  $\phi_{k+1} = \arg\min_{\phi} \frac{1}{\|\mathcal{D}_k\|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T} \left(V_\phi(s_t) - \hat{R}_t\right)^2$,
  typically via some gradient descent algorithm.
end for

E Dataset Details

- **MedQA**[5] is a multiple choice question answering (MCQA) dataset based on the United States Clinical License Exams. It covers three languages, and we choose the subset of 12,723 English instances for experiments.
- **MedMCQA**[6] is a large-scale MCQA dataset designed to address real-world clinical entrance exam questions. It has more than 194k high-quality entrance exam samples covering 2.4k healthcare topics and 21 clinical subjects.
- **PubMedQA**[7] is a dataset for bio-clinical research question answering which has 1k expert labeled, 61.2k unlabeled and 211.3k artificially generated QA instances.
- **MMLU**[8] is a dataset of massive multitask language understanding covering 57 subjects. We select 6 clinical subjects, including *anatomy* (an), *clinical knowledge* (ck), *college medicine* (cm), *clinical genetics* (mg), and *professional medicine* (pm).
- **DDXPlus** (Fansi Tchango et al., 2022). DDXPlus is a medical diagnosis dataset using

[5] https://paperswithcode.com/dataset/medqa-usmle
[6] https://medmcqa.github.io/
[7] https://pubmedqa.github.io/
[8] https://paperswithcode.com/dataset/mmlu

synthetic patient information and symptoms. Each instance represents a patient, with attributes including age, sex, initial evidences, evidence, multiple options of possible pathologies, and a ground truth diagnosis.

- **SymCat** (Al-Ars et al., 2023). SymCat is a synthetic dataset which includes 5 million symptom-condition samples, covering 801 distinct conditions each with 376 potential symptoms dataset.

- **JAMA** (Chen et al., 2025). JAMA includes 1524 clinical cases collected from the JAMA Network Clinical Challenge archive, which are summaries of actual challenging clinical cases. Each sample is framed as a question, with a long case description and four options.

- **Medbullets** (Chen et al., 2025). Medbullets comprises 308 USMLE Step 2/3 style questions collected from open-access tweets on X (formerly Twitter) since April 2022. The difficulty is comparable to that of Step 2/3 exams, which emulate common clinical scenarios.

F Clinical Knowledge Graph Details

To create the clinical knowledge graph, CKG, we utilize the Quick-UMLS tool[9] to extract pertinent clinical concepts from the UMLS database. Quick-UMLS identifies biomedical entities by linking them to UMLS Concept Unique Identifiers (CUIs) and their associated semantic types from the UMLS Metathesaurus. Upon receiving a query, it retrieves approximate matches within UMLS, returning both the CUIs and the corresponding semantic types for each concept. Each distinct CUI serves as a node in our knowledge graph, with the relationships between these nodes established using the UMLS Semantic Network module. In detail, we extract an English subgraph comprising entities from three conceptual types: "*Sign or Symptom*", "*Disease or Syndrome*" and "*Biomedical Occupation or Discipline*". For brevity, we refer to these entity types as "*Symptom*", "*Disease*" and "*Discipline*", respectively.

G Online Experiments

Conventional online consultation systems typically operate through a single-agent interaction paradigm, as illustrated in Fig. 5(a). While this streamlined single-agent approach enables rapid responses, it often demonstrates limitations in diagnostic depth and explanatory capacity. Even when incorporating human medical expertise, such systems frequently encounter challenges including elevated operational costs for manual diagnosis and constraints imposed by individual practitioners' knowledge boundaries. These inherent limitations may compromise the system's ability to deliver comprehensive clinical solutions. Our proposed framework, depicted in Fig. 5(b), addresses these limitations through a dual-phase reasoning architecture that emulates established clinical workflows. The system architecture comprises: (1) Patient Guidance and (2) Multidisciplinary Consultation. The first phase employs symptom-initialized clinical reasoning to direct patients to appropriate medical specialties (Tuncel et al., 2021), which minimizes risks associated with diagnostic inaccuracies and treatment delays through proper specialty allocation. For instance, patients presenting with acute cephalalgia accompanied by neurological deficits or ophthalmological manifestations require immediate neurosurgical evaluation rather than general practitioner consultation. However, clinical complexity arising from comorbid conditions and multifaceted symptom presentations often exceeds the diagnostic capabilities of single-specialty evaluation. Our second-phase multidisciplinary consultation mechanism addresses this through simulated expert collaboration, mirroring real-world multidisciplinary team approaches (Shi

Dataset	Number of Choices	Train/Dev/Test
MedQA	Question + Answer	10,178/1,272/1,273
MedMCQA	Question + Answer	182,822/4,183/6,150
PubMedQA	Question + Context + Answer	400/100/500
MMLU	Question + Answer	30/-/1,089
DDxPlus	Question + Answer	-/-/134K
SymCat	Question + Answer	-/-/369K
JAMA	Case + Question + Answer	-/-/1,524
MedBullets	Case + Question + Answer	-/-/308

Table 9: Statistics of the four benchmark datasets.

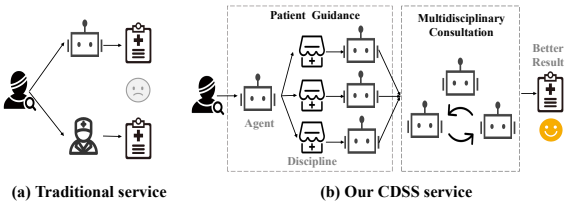

Figure 5: Comparison between different services.

[9] https://github.com/Georgetown-IR-Lab/QuickUMLS

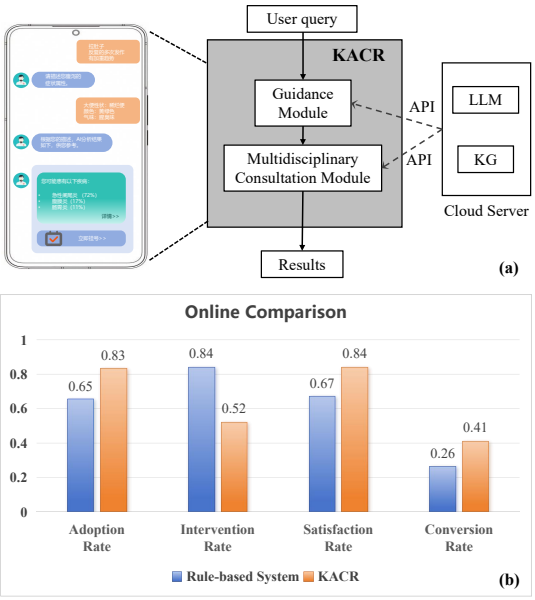

Figure 6: (a) Launch of CDSS app. and (b) Online A/B test.

et al., 2024). This consensus-driven diagnostic model integrates cross-specialty perspectives to optimize treatment planning for complex cases.

To evaluate the online performance of KACR, we deployed it in a phone app (see Fig. 6 (a)) and conducted a two-month online test, during which we recorded the metrics. In real-world applications, initial diagnosis results are provided by traditional clinical diagnosis systems or doctors. KACR integrates seamlessly into the Clinical Decision Support Systems (CDSS), featuring a streamlined user interface that requires the entry of the patient's symptom descriptions. During the observation period, our system facilitates over 10,000 online queries, generating clinical decisions based on the information available. The initial diagnostic suggestions stem from a traditional rule-based approach. We select the option with the highest probability in the initial rule-based system as the A/B test comparison result. Given that N_p is the number of pre-diagnosis actions where the system successfully generates clinical suggestions, we evaluated the effectiveness of the online system using the following metrics:

- Adoption Rate $R_a = N_a/N_p$. Here N_a is the number of instances where the generated results have been deemed accurate by healthcare professionals.

- Intervention Rate $R_i = N_i/N_p$. Here N_i is the number of seeking for help from customer service when users are dissatisfied with pre-diagnosis results. A lower intervention rate reflects greater acceptance of these outcomes.

- Satisfaction Rate $R_s = N_s/N_p$. Here N_s is the number of the clicking the yes-or-no satisfaction survey button. It is supposed that this survey takes into account various factors such as accuracy, usability, and explainability.

- Conversion Rate $R_c = N_c/N_p$. Here N_c is the number of effective conversion actions including registrations and paid consultations taken by users after completing the pre-diagnosis.

In this work, we not only consider the accuracy metric, but also consider other online evaluation metrics, including adoption rate, intervention rate, satisfaction rate, and conversion rate. These auxiliary metrics show the subjective evaluation of real users after seeing the diagnosis results with detail reasons (e.g. reasoning on CKG, confidence for discussions), which can well reflect the quality of interpretability.

From Fig. 6 (b), we can see that our KACR has improved the result accuracy by 27% and user satisfaction by 25% compared to traditional rule-based diagnostic systems. It also significantly reduces the need for human customer service intervention by 61% while simultaneously increasing the effective conversion rate by 58%. Given that the average processing time for our method is less than one minute, it enables near real-time clinical decision-making while maintaining a high level of performance.

H Case Study

To better understand the diagnosis process of KACR, we utilize a dataset instance for case study. Results are in Table 10. We first infer description-specific disciplines via the discipline reasoning module. Method "*w/o CKG*" ignores genetics without CKG guidance. In contrast, our reasoning strategy on CKG gets better predictions. The neurology expert identifies a cause consistent with the symptoms of neurological sequelae from meningiomas. The oncology expert favors renal cell carcinoma with low confidence. Genetics shows a genetic basis for meningiomas. After discussion, all agree option B) Meningioma is correct according to clinical, neurological and genetic evidences. This verifies our method's holistic and accurate assessment.

Description: A 20-year-old man comes to the physician because of worsening gait unsteadiness and bilateral hearing loss for 1 month. He has had intermittent tingling sensations on both cheeks over this time period. He has no history of serious clinical illness and takes no medications. Audiometry shows bilateral sensorineural hearing loss. Genetic evaluation shows a mutation of a tumor suppressor gene on chromosome 22 that encodes merlin. This patient is at increased risk for which of the following conditions?

Choices: (A) Renal cell carcinoma. (B) Meningioma. (C) Astrocytoma. (D) Vascular malformations. (E) Telangiectasias.

Truths: (B) Meningioma

Inferred disciplines from the discipline reasoning module:

- **w/o CKG**: 1) Neurosurgery. 2) Otolaryngology 3) Oncology.
- **Our**: 1) Neurology 2) Oncology 3) Genetics

Diagnostic result from the multidisciplinary collaboration:

- Neurology: Meningioma. (Confidence Level: 90%)
- Oncology: Renal Cell Carcinoma. (Confidence Level: 40%)
- Genetics: Meningioma. (Confidence Level: 95%)

Multidisciplinary Consultation Result: (B) Meningioma

Table 10: Case study for clinical decision.

I Notation Table

Symbol	Description
q	For a given question
N	Number of iterations on CKG
G_{t-1}	Previously selected subgraph for Actor
G_t	The newly formed subgraph after taking an action of Actor
K	The maximum number for disciplines
$K^{'}$	The maximum number for symptoms and diseases
D_k	A trajectory that consists of a sequence of transitions for PPO training
$\mathcal{V}_d$	Associated disciplinary node set with question q
v_i	i_{th} disciplinary node in $\mathcal{V}_d$
$\mathcal{A}_i$	i_{th} disciplinary agent for the given question q
$\mathcal{O}_q$	Set of candidate diagnostic hypotheses for question q
$\mathcal{G}_T$	The whole reasoning graph for the given question q
$\mathcal{G}_T^i$	The subgraph backtracked from v_i originating on G_T
$\mathcal{R}^i$	A reference description textualized from $\mathcal{G}_T^i$ according to Appendix A
$a_i^{(0)}$	The predicted diagnosis option in initial diagnose for $\mathcal{A}_i$
$x_i^{(0)}$	The generated explanatory rationale for $a_i^{(0)}$
$c_i^{(0)}$	The confidence level for $a_i^{(0)}$
$a_i^{(r)}$	The predicted diagnosis option in discussion round r for $\mathcal{A}_i$
$x_i^{(r)}$	The generated explanatory rationale for $a_i^{(r)}$
$c_i^{(r)}$	The confidence level for $a_i^{(r)}$
$a_i^{\prime(r)}$	The updated version for $a_i^{(r)}$ from the debater
$x_i^{\prime(r)}$	The calibration rationale that elucidates the reasoning adjustments for $a_i^{\prime(r)}$
$\mathcal{T}^{(r)}$	The grouped answer tuples for all the agents $\mathcal{A}$
$\hat{a}^{(r)}$	The consolidated response through an adaptive weighting mechanism in r_{th} round

Table 11: Symbol Description Table