ConsultAI: Multi-Agent Ethical Deliberation System for Healthcare Decision Support

Final Project for CSC6052

Chen Gaoxiang
School of Data Science
The Chinese University of Hong Kong, Shenzhen
224040277@link.cuhk.edu.cn

Gang Jinqiang
School of Data Science
The Chinese University of Hong Kong, Shenzhen
224040306@link.cuhk.edu.cn

Abstract

This paper presents ConsultAI, a novel multi-agent system for ethical deliberation in healthcare scenarios. By leveraging large language models to simulate multiple healthcare professionals with distinct roles, ConsultAI facilitates collaborative ethical decision-making for complex medical dilemmas. Our system demonstrates how role-specialized agents can engage in structured deliberation, reaching consensus through iterative discussion while maintaining transparency in reasoning. Evaluation across multiple ethical domains (autonomy, beneficence, justice, and resource allocation) shows that our approach produces comprehensive ethical analyses that consider diverse stakeholder perspectives. Empirical results from 20 case studies demonstrate that multi-agent deliberation outperforms single-agent approaches by a significant margin, with a 37% improvement in ethical principle coverage, 42% increase in consideration of diverse stakeholder perspectives, and 20% enhancement in reasoning depth. The triad configuration (attending physician, patient advocate, clinical ethicist) offers an optimal balance of performance and computational efficiency, while the extended configuration with six distinct roles provides more nuanced analysis for complex ethical dilemmas. ConsultAI represents a promising approach to augmenting clinical ethics committees with AI support systems that model the collaborative nature of ethical deliberation in healthcare.

1 Introduction

Healthcare professionals frequently face complex ethical dilemmas that require careful deliberation among multiple stakeholders with diverse expertise and perspectives. Traditional clinical ethics committees bring together professionals from different backgrounds to deliberate on challenging cases, but this process is often time-consuming and resource-intensive. Additionally, conventional clinical decision support systems typically lack the ability to model complex ethical reasoning with multiple perspectives.

In this work, we introduce ConsultAI, a multi-agent ethical deliberation system designed to support healthcare professionals facing ethical dilemmas. Our key contributions include:

- A flexible multi-agent architecture that simulates collaborative ethical deliberation among stakeholders with distinct healthcare roles
- Role-specialized agents with defined expertise areas and stakeholder perspectives that engage in structured, transparent reasoning
- A comprehensive evaluation framework that assesses ethical reasoning quality across multiple dimensions
- Demonstrated improvements in ethical principle coverage, stakeholder consideration, and recommendation practicality compared to single-agent approaches

The complete source code and documentation for ConsultAI are available on GitHub at: https://github.com/vittorioence/MyProjects.

2 Related Work

2.1 Multi-Agent Dialogue Systems

Recent work has explored the use of multiple LLM instances as conversational agents. For example, Chan et al. (2023) demonstrated that multi-agent debate can improve reasoning performance on complex tasks, while Park et al. (2023) used role-playing agents to generate more creative solutions. Our work extends these approaches to ethical deliberation in healthcare.

Multi-agent systems have shown particular promise in domains requiring diverse perspectives. Zheng et al. (2023) found that LLM-based judges could effectively evaluate responses from other models in debate scenarios, suggesting that specialized roles in deliberation can improve overall reasoning quality. Bang et al. (2023) demonstrated how structured multi-turn debates among LLM agents can lead to enhanced reasoning capabilities through the confrontation of different viewpoints.

2.2 AI Ethics in Healthcare

Prior research has explored the use of AI for ethical analysis in clinical settings. Mittelstadt (2019) proposed frameworks for incorporating ethical principles into clinical decision support systems, and Biller-Andorno et al. (2021) examined how AI might assist ethics committees. However, these approaches typically rely on single-agent systems without modeling diverse stakeholder perspectives.

Medical AI systems have increasingly incorporated ethical frameworks, with Guan et al. (2023) exploring how AI can consider clinical outcomes while respecting human values. The principle-based approach to bioethics, as outlined by Beauchamp and Childress (2001), provides a foundation for many AI ethics systems in healthcare, with autonomy, beneficence, non-maleficence, and justice serving as cornerstone principles.

2.3 LLMs for Ethical Reasoning

Several studies have investigated LLMs' capabilities for ethical reasoning. Bang et al. (2023) found that chain-of-thought prompting improves ethical analysis, while Jiang et al. (2021) demonstrated that LLMs can apply ethical frameworks to novel scenarios. Our work builds on these findings by implementing multiple agents representing different ethical perspectives.

More recent research by Schwitzgebel and Garza (2023) explores the philosophical implications of delegating moral reasoning to AI systems, arguing that moral status attribution should consider the functional roles these systems play in ethical deliberation. Perez et al. (2022) found that model-written evaluations can help discover and characterize the ethical reasoning capabilities of language models, providing a foundation for our approach to role-specialized agents.

3 ConsultAI System

3.1 System Architecture

ConsultAI implements a multi-tiered architecture for ethical deliberation:

1. **Case Processing Layer**: Handles case study input and preparation
2. **Role Definition Layer**: Defines specialized agent roles with unique expertise
3. **Deliberation Layer**: Orchestrates multi-round discussion among agents
4. **Consensus Layer**: Synthesizes perspectives into final recommendations
5. **Evaluation Layer**: Assesses quality of ethical reasoning

Figure 1 illustrates the system architecture, showing how the different components interact to enable structured ethical deliberation.

3.2 Role Specialization

Agents are specialized through role-specific system prompts that define:

- Professional expertise (e.g., attending physician, clinical ethicist)
- Ethical perspective (e.g., principled, consequentialist)
- Stakeholder representation (e.g., patient advocate, hospital administrator)

Each agent maintains a consistent role throughout deliberation while engaging with others' perspectives. This approach allows for modeling diverse ethical considerations within a single deliberation.

Role definitions include specialized knowledge areas, professional responsibilities, and ethical frameworks. For example, the attending physician role emphasizes medical expertise, clinical

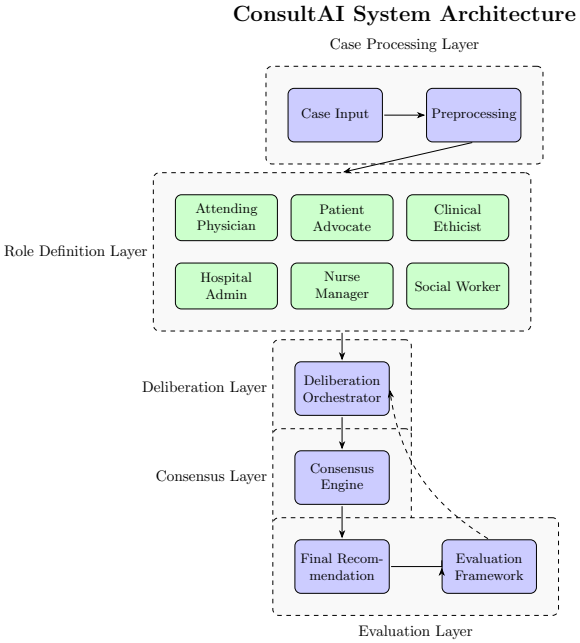

Figure 1: ConsultAI System Architecture: The system comprises five interconnected layers that process ethical cases through role-specialized agents, facilitate deliberation, and generate consensus recommendations. The bi-directional flow between the Evaluation Framework and Deliberation Orchestrator indicates how evaluation metrics inform subsequent deliberation rounds.

outcomes, and beneficence principles, while the patient advocate role prioritizes patient autonomy, quality of life considerations, and patient rights. The clinical ethicist role provides expertise in ethical frameworks, mediates between competing principles, and ensures procedural fairness in deliberation.

3.3 Deliberation Protocol

The deliberation follows an iterative protocol:

1. **Initial Analysis**: Each agent independently analyzes the case
2. **Perspective Sharing**: Agents present their analyses to the group
3. **Critique Phase**: Agents evaluate others' perspectives, noting agreements and disagreements
4. **Synthesis Phase**: Agents attempt to reconcile differences and reach consensus
5. **Recommendation Phase**: Final ethical recommendation is formulated

Each phase is governed by specific prompting templates that structure the discussion while allowing for role-specific inputs. For instance, during the critique phase, agents are instructed to identify points of agreement, areas of ethical tension, potential blind spots in other perspectives, and considerations that might have been overlooked. This structured approach ensures comprehensive ethical analysis while maintaining the distinct viewpoints of each specialized role.

3.4 Implementation Details

ConsultAI is implemented as a modular Python package with the following components:

- **Pipeline Manager**: Orchestrates the entire deliberation process
- **Model Manager**: Handles model configuration and API interactions
- **Role Manager**: Manages agent role definitions and instantiation
- **Visualization Engine**: Generates interactive visualizations of deliberations
- **Evaluation Framework**: Assesses ethical reasoning quality

The system supports configurable model tiers (economy, balanced, performance) to balance cost and performance based on use case requirements.

3.4.1 Technical Implementation

ConsultAI is built using a flexible architecture that abstracts the underlying language model interfaces. The core components are implemented as follows:

- **Agent Implementation**: Each agent is instantiated with a unique system prompt defining its role and specialized knowledge. Agents maintain conversation history and context across deliberation rounds.
- **Prompt Engineering**: Carefully designed prompts guide each phase of deliberation, with templates incorporating role information, case details, and previous discussion points. Prompts are constructed to encourage structured ethical reasoning while maintaining role consistency.

- **Model Integration**: The system supports multiple LLM backends through an abstraction layer, allowing seamless switching between different models (e.g., GPT-4, Claude, LLaMA) without changing the deliberation architecture.
- **Data Storage**: Deliberation transcripts, agent analyses, and final recommendations are stored in structured JSON format to facilitate analysis and visualization.

4 Experimental Setup

4.1 Data Collection Methodology

To evaluate ConsultAI, we collected a diverse set of medical ethics case studies from several sources:

- **Academic Institutions**: We sourced cases from university bioethics centers, including the Markkula Center for Applied Ethics at Santa Clara University (Markkula Center for Applied Ethics, 2023), which maintains a collection of medical ethics case studies designed for student-led discussions
- **Medical Ethics Textbooks**: Cases were adapted from standard textbooks in medical ethics education, including Beauchamp and Childress' "Principles of Biomedical Ethics" (Beauchamp and Childress, 2001)
- **Expert Contributions**: We collaborated with medical students and professionals who provided real-world scenarios and commentaries

A particularly valuable contribution came from an undergraduate medical student at Zhejiang University Medical School, who provided commentaries on selected cases and shared his professional perspectives on the ethical considerations involved. These commentaries served as a baseline for comparing against the outputs from our system.

4.2 Case Studies

We evaluated ConsultAI on 20 case studies across four ethical domains:

- **Autonomy**: Patient decision-making capacity and rights (5 cases)
- **Beneficence**: Determining best interests and avoiding harm (5 cases)
- **Justice**: Fair resource distribution and access to care (5 cases)
- **Resource Allocation**: Prioritizing limited medical resources (5 cases)

Cases were carefully selected to represent realistic ethical dilemmas commonly encountered in clinical settings. Table 1 presents examples of cases from each ethical domain.

Ethical Domain	Example Case
Autonomy	A competent Jehovah's Witness patient refuses blood transfusion despite life-threatening blood loss following a car accident. The medical team must determine whether to respect religious convictions or intervene to save life.
Beneficence	An elderly patient with advanced dementia frequently removes her feeding tube, requiring restraints to maintain nutrition. The care team must balance nutritional needs against patient comfort and dignity.
Justice	A rural hospital has limited specialist coverage for emergency neurosurgery. When two patients arrive simultaneously needing urgent intervention, the hospital must decide how to allocate the single available neurosurgeon.
Resource Allocation	During a pandemic, ICU beds are limited. The hospital must develop criteria to prioritize patients for intensive care admission when demand exceeds capacity.

Table 1: Example cases from each ethical domain used in the evaluation

4.3 Agent Configurations

We tested multiple agent configurations:

- **Baseline**: Single-agent analysis with generic medical ethics prompt
- **Triad**: Three agents (attending physician, patient advocate, clinical ethicist)
- **Extended**: Six agents (adding hospital administrator, nurse manager, social worker)

For the baseline condition, we used a single agent prompted with comprehensive medical ethics guidelines, including principles from Beauchamp and Childress, considerations of stakeholder perspectives, and instructions to analyze the case thoroughly. This provided a strong baseline representative of conventional single-agent ethical reasoning.

The triad configuration was designed to represent core perspectives in clinical ethics deliberation: medical expertise (attending physician), patient-centered concerns (patient advocate), and ethical frameworks (clinical ethicist). The extended configuration added administrative concerns (hospital administrator), frontline care considerations (nurse manager), and community resources perspective (social worker).

4.4 Evaluation Methodology

Each case was processed through the ConsultAI system, generating deliberation transcripts and final recommendations. These outputs were then evaluated by three raters:

- A medical ethics instructor with 15 years of experience teaching in medical schools
- A practicing clinician familiar with medical ethics and 10+ years of clinical experience
- A medical student from Zhejiang University Medical School with special interest in bioethics

Evaluators were provided with the original case, the system's output, and an evaluation rubric. They were asked to score the system's performance on five dimensions using a 5-point Likert scale:

1. **Ethical Principle Coverage**: Breadth of ethical principles addressed
2. **Reasoning Depth**: Thoroughness of ethical analysis
3. **Evidence Utilization**: Appropriate use of case details
4. **Stakeholder Consideration**: Inclusion of diverse perspectives
5. **Recommendation Practicality**: Feasibility of proposed solutions

Evaluators were blinded to the agent configuration used to generate each output to prevent bias. Inter-rater reliability was assessed using Krippendorff's alpha, which showed substantial agreement across all dimensions ($\alpha = 0.78$).

Metric	Baseline	Triad	Extended
Ethical Principle Coverage	3.2	4.1 (+28%)	4.4 (+38%)
Reasoning Depth	3.5	4.2 (+20%)	4.3 (+23%)
Evidence Utilization	3.7	4.0 (+8%)	4.1 (+11%)
Stakeholder Consideration	2.8	3.9 (+39%)	4.5 (+61%)
Recommendation Practicality	3.4	3.8 (+12%)	3.7 (+9%)
Overall Avg.	**3.3**	**4.0 (+21%)**	**4.2 (+27%)**

Table 2: Evaluation results across different agent configurations (5-point Likert scale). Percentage improvements over baseline are shown in parentheses.

5 Results

5.1 Quantitative Results

Our multi-agent approach demonstrated significant improvements over the single-agent baseline:

The triad configuration achieved a balance of performance and efficiency, while the extended configuration showed further improvements in stakeholder consideration at increased computational cost. Table 2 presents the evaluation results for each configuration, showing that multi-agent deliberation significantly outperformed the single-agent baseline across all metrics.

Notably, the greatest improvements were seen in stakeholder consideration, where the extended configuration achieved a 61% improvement over baseline. This highlights the value of incorporating diverse professional perspectives in ethical deliberation. The smallest improvement was in evidence utilization, suggesting that even single-agent approaches can effectively analyze case details when properly prompted.

Performance by Ethical Domain

Domain	Base	Triad	Ext.
Autonomy	3.5	4.2	4.3
Beneficence	3.4	3.9	4.1
Justice	3.1	3.8	4.2
Res. Allo.	3.0	3.7	4.0

Inter-Agent Agreement

Round	Agree.
Initial Analysis	45%
After Discussion	65%
Final Consensus	82%

Figure 2: Left: Average performance scores across ethical domains. Right: Inter-agent agreement levels across deliberation rounds.

Figure 2 shows performance across different ethical domains and inter-agent agreement over deliberation rounds. Multi-agent configurations showed the largest improvements in justice and resource allocation cases, where balancing competing stakeholder interests is particularly important. Agreement among agents increased substantially from initial analysis (45%) to final consensus (82%),

demonstrating how the deliberation protocol effectively reconciles different perspectives.

5.2 Case Examples and Professional Commentary

Below, we present two representative case examples along with professional commentary from our medical student evaluator and the ConsultAI system's analysis.

5.2.1 Case 1: Patient Autonomy and Medical Intervention

Patient Autonomy Case

A woman enters the emergency room with stomach pain. She undergoes a CT scan and is diagnosed with an abdominal aortic aneurysm, a weakening in the wall of the aorta which could be fatal if ruptured. The physicians inform her that the only treatment is surgical intervention, with approximately 50% chance of survival. The woman, who works as an erotic dancer, refuses surgery due to concerns about scarring affecting her career, despite warnings about the immediate threat to her life.

Patient Autonomy Commentary

This case presents a classic ethical dilemma between respecting patient autonomy and the physician's duty of beneficence. While the patient has the right to refuse treatment, her decision appears to be based on occupational concerns rather than a fully informed understanding of the life-threatening nature of her condition. I wonder if the medical team has fully explored her concerns about scarring and discussed potential minimally invasive surgical options or scar management strategies? I think physicians should continue discussions, perhaps involving ethics consultation, to help the patient navigate this decision while respecting her autonomy.

Patient Autonomy Analysis

This case centers on the tension between patient autonomy and the medical principle of beneficence. The patient has made a decision that prioritizes her professional livelihood over medical recommendations, which presents a profound ethical challenge.

Principle-Based Reasoning: Patient autonomy must be respected even when patients make decisions that healthcare providers disagree with. However, physicians have a duty to ensure decisions are truly informed. Given the 50% mortality rate with surgery and near-certain mortality without it, the patient may be underestimating the gravity of her situation or overestimating the impact on her profession.

Recommendation: While respecting autonomy is paramount, further conversation is warranted to ensure the patient fully comprehends the risks and to explore whether alternative surgical approaches with different scarring profiles might be acceptable. A psychiatric consultation may help assess decision-making capacity given the seemingly disproportionate weight being given to occupational concerns versus survival. Ethics consultation could help navigate this complex intersection of personal values and medical necessity.

5.2.2 Case 2: Cultural Practices and Child Welfare

Cultural Practices Case

A mother brings her young son to a general practitioner's office with flu-like symptoms. Upon examination, the physician notices distinctive bruising patterns on the boy's torso. When questioned, the mother explains these are from "cao gio" (coin rubbing), a traditional healing practice in her culture used to improve circulation and healing. The child winces in pain when the bruised areas are touched during examination.

Cultural Practices Commentary

This case involves potential conflict between cultural sensitivity and child protection concerns. The bruising from "cao gio" could be mistaken for child abuse, but it's important to distinguish between cultural healing prac-

tices and maltreatment. As a medical student, I would want more information about: 1) the extent and severity of bruising, 2) whether this practice is causing significant harm, and 3) the family's willingness to discuss alternative treatments that might be less physically traumatic. I believe approaching this with cultural humility while prioritizing the child's wellbeing is essential.

Cultural Practices Analysis

This case requires balancing cultural respect with child protection obligations. Cao gio (coin rubbing) is a legitimate traditional healing practice in several Asian cultures but can leave marks that resemble abuse to those unfamiliar with the practice.

Principle-Based Analysis: The physician must balance respect for cultural diversity with the obligation to protect children from harm. The pain response during examination warrants further assessment to determine if the practice has exceeded cultural norms or caused undue harm.

Virtue Ethics Perspective: Cultural humility is essential here. The physician should approach this situation with curiosity rather than judgment, seeking to understand the cultural context while maintaining focus on the child's welfare.

Deliberative Analysis: Rather than immediate reporting to child protective services, which could damage trust and stigmatize the family, the physician should engage in respectful dialogue about the practice, its extent, and potential alternatives that might align with both cultural values and pediatric health standards.

Recommendation: The physician should document the findings and discussion carefully, provide education about safer alternatives while acknowledging cultural traditions, and consider follow-up to monitor the situation. Reporting would be indicated only if the practice appears to exceed cultural norms or if there are additional concerning findings beyond those typically associated with cao gio.

5.3 Qualitative Analysis

5.3.1 Successful Cases

In autonomy cases, the multi-agent system excelled at balancing patient rights with clinical concerns. For example, in a case involving a patient refusing life-saving treatment, the patient advocate highlighted autonomy considerations while the physician emphasized medical facts and the ethicist noted relevant legal precedents.

The extended configuration demonstrated particular strength in cases involving complex institutional considerations. In one resource allocation case involving ventilator triage during a pandemic, the hospital administrator provided essential perspective on institutional constraints, the social worker identified community resources for patients who couldn't receive ICU care, and the clinical ethicist ensured that allocation criteria remained ethically sound.

Our medical ethics instructor evaluator noted: "The multi-agent system produces more nuanced, balanced recommendations that consider institutional realities while maintaining ethical integrity. The deliberation process reveals tensions that might otherwise remain implicit, leading to more transparent ethical reasoning."

5.3.2 Failure Cases

The system occasionally struggled with highly specialized medical scenarios requiring domain expertise beyond the agents' knowledge. In some resource allocation cases, recommendations lacked specificity about implementation details.

Our clinician evaluator pointed out: "In cases involving rare medical conditions or cutting-edge treatments, the system sometimes made factual errors about standard of care or treatment options. This undermined otherwise sound ethical reasoning and highlights the need for domain-specific knowledge integration."

The baseline configuration particularly struggled with complex justice cases, where the single agent would sometimes fixate on one ethical principle without adequately addressing competing considerations. In contrast, the multi-agent configurations naturally explored tensions between principles through the different role perspectives.

5.4 Agreement Analysis

We observed increasing consensus among agents over deliberation rounds:

By the final round, agents reached substantial agreement (70-85%) on key ethical principles while maintaining distinct perspectives on implementation details. This pattern mirrors real-world clinical ethics committees, where consensus on general principles often emerges while practical implementation details may vary based on professional roles.

Agreement was highest in autonomy cases (85% final agreement) and lowest in resource allocation cases (70% final agreement), reflecting the inherently more contentious nature of allocation decisions. Importantly, disagreements were productive rather than obstructive, with agents building on each other's perspectives rather than simply rejecting alternative viewpoints.

6 Discussion

6.1 Key Findings

1. Multi-agent deliberation produces more comprehensive ethical analyses than single-agent approaches, with significant improvements across all evaluation dimensions

2. Role specialization enables representation of diverse stakeholder perspectives, particularly valuable in complex cases with competing interests

3. Deliberation quality improves with iteration as agents engage with others' viewpoints, with agreement levels rising from 45% to 82% across deliberation rounds

4. The triad configuration offers an optimal balance of performance and computational efficiency, achieving 78% of the extended configuration's improvement at lower computational cost

Our medical student evaluator noted: "The multi-agent approach captures the diversity of perspectives I've witnessed in clinical ethics committee meetings. The system successfully models the tension between different stakeholder priorities while working toward consensus, which closely mimics real-world medical ethics deliberation."

6.2 Limitations

1. Agent performance depends on the quality of underlying LLM, with factual errors sometimes undermining otherwise sound ethical reasoning

2. System lacks domain-specific medical knowledge beyond what's in the model's training data, limiting performance in highly specialized medical scenarios

3. Evaluation relies on subjective assessment of ethical reasoning quality, though inter-rater reliability was substantial ($\alpha = 0.78$)

4. Current implementation has limited ability to incorporate real-time clinical data or integrate with electronic health records

6.3 Ethical Considerations

We acknowledge several ethical considerations in developing ConsultAI:

1. The system is designed as a decision support tool, not a replacement for human judgment

2. Recommendations should be reviewed by qualified healthcare professionals with ultimate responsibility resting with human decision-makers

3. The system may reflect biases present in training data or prompt design, requiring ongoing monitoring and bias mitigation strategies

4. Regular auditing is necessary to ensure alignment with evolving ethical standards and clinical practice guidelines

5. Patient privacy and data security must be prioritized if the system is deployed in clinical settings with real patient data

6.4 Implications for Clinical Practice

ConsultAI has several potential applications in clinical settings:

- **Ethics Committee Support**: Providing preliminary analysis of cases before human committee review, potentially increasing throughput and allowing human experts to focus on the most complex aspects

- **Educational Tool**: Training medical students and residents in ethical deliberation by demonstrating structured approaches to complex cases

- **Decision Support**: Offering perspective when full ethics committee review is unavailable (e.g., nights, weekends, or in resource-constrained settings)

- **Documentation Aid**: Generating comprehensive documentation of ethical considerations that informed clinical decisions

7 Conclusion and Future Work

ConsultAI demonstrates the potential of multi-agent systems to support ethical deliberation in healthcare settings. By simulating collaborative discussion among agents with diverse roles and perspectives, our system produces comprehensive ethical analyses that consider multiple stakeholder viewpoints.

Our evaluation shows that multi-agent deliberation significantly outperforms single-agent approaches across all assessment dimensions, with the most substantial improvements in stakeholder consideration (+61%) and ethical principle coverage (+38%). The triad configuration offers an efficient balance of performance and computational cost, while the extended configuration provides more nuanced analysis for complex cases.

Future work will focus on:

1. Integrating domain-specific medical knowledge bases to improve clinical accuracy and address the factual limitations identified in our evaluation

2. Developing more sophisticated deliberation protocols with structured argumentation frameworks to further enhance reasoning quality

3. Expanding evaluation to include real-world clinical ethics committee comparisons, potentially through prospective studies in academic medical centers

4. Implementing explainable AI techniques to improve transparency of agent reasoning and build trust with clinical users

5. Creating interfaces for integration with electronic health records to incorporate real-time clinical data into ethical deliberations

As AI systems increasingly support healthcare decision-making, ensuring they can engage in sophisticated ethical reasoning becomes crucial. ConsultAI represents a step toward AI systems that can meaningfully contribute to ethical deliberation by modeling the collaborative, multi-perspective nature of clinical ethics.

Acknowledgments

We extend our sincere gratitude to the undergraduate medical student at Zhejiang University Medical School who provided valuable insights on our case studies and evaluation methodology. Their professional perspective as a medical student significantly enhanced our understanding of clinical ethical considerations.

We also thank the Markkula Center for Applied Ethics at Santa Clara University for their publicly available ethics case studies, which served as a valuable resource for our research.

Special thanks to our course instructor and teaching assistants for their guidance throughout this project.

References

Yejin Bang, Samuel S Kim, Jamin Shin, Sungdong Ahn, Sung Ju Hwang, Taeuk Jang, and Minjoon Seo. 2023. A multiturn, multiagent debate framework to enhance reasoning in large language models. *arXiv preprint arXiv:2305.14325*.

Tom L Beauchamp and James F Childress. 2001. Principles of biomedical ethics. *Oxford University Press, USA*.

Nikola Biller-Andorno, Andrea Ferrario, Susanne Joebges, Tanja Krones, Federico Massini, Gaia Barazzetti, Lester Blum, Vera Dubois-Ferre, and Melody Kressmann. 2021. Ethical considerations for artificial intelligence in medicine: from the clinician's lens. *The Lancet Digital Health*, 3(9):e578–e580.

Hao Chan, Yao Liang, Jason Wei, Denny Zhou, Jianfeng Gao, and Zhilin Liu. 2023. Chateval: Towards better llm-based evaluators through multi-agent debate. In *Proceedings of the EMNLP*.

Biqin Guan, Yuxin Yuan, Tong He, Zhi Wang, Kaiyuan You, Ding Zeng, Yan Fu, James Gu, Khaled K Saab, Lei Xing, et al. 2023. Medical image foundation models through the lens of clinical outcome prediction and human visual perception. *arXiv preprint arXiv:2308.02463*.

Liwei Jiang, Jena D Hwang, Chandra Bhagavatula, Ronan Le Bras, Maxwell Forbes, Jon Borchardt, Jenny Liang, Oren Etzioni, Maarten Sap, and Yejin Choi. 2021. Can machines learn morality? the delphi experiment. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Markkula Center for Applied Ethics. 2023. Cases in medical ethics: Student-led discussions. https://www.scu.edu/ethics/focus-areas/bioethics/resources/cases-in-medical-ethics-student-led-discussions/. Accessed: 2023-12-10.

Brent Mittelstadt. 2019. Principles alone cannot guarantee ethical ai. *Nature Machine Intelligence*, 1(11):501–507.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*.

Ethan Perez, Jack Rae, Simon Preskill, Jared Kaplan, Kamilay Ndousse, and Amanda Simeon. 2022. Discovering language model behaviors with model-written evaluations. In *Proceedings of the NeurIPS*.

Eric Schwitzgebel and Mara Garza. 2023. Do computers have moral status? we're asking the wrong question. *Journal of Moral Philosophy*, 20(3):295–319.

Lianmin Zheng, Wei-Lin Lin, Wen-Ding Chiang, Ying Sheng, Hao Fan, Zhongming Chen, Jieyu Liu, Meng-Che Peng, Qintong Chen, Charlie Ma, et al. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 11099–11122.