Conversational Agent Debating Capability

Mohamad Faraj Makkawi
IMT Atlantique
Rennes, France
mohamad.makkawi@imt-atlantique.net

Hassan Khan
IMT Atlantique
Rennes, France
hassan.khan@imt-atlantique.net

Abstract—This article discusses the evolving capabilities of conversational agents to participate in sophisticated debates in terms of structured argumentation and symbolic reasoning strategies. A thorough review on salient computational approaches such as Argument Graphs, the ASPIC methodology, and Multi-Attribute Argumentation Frameworks, and evaluation of how they enhance logical consistency, comprehensibility, and rhetorical power in AI debates is done. By comparing different debating architectures such as rule-based, retrieval-based, hybrid, hierarchical, and explainable agent, the paper highlights the benefits and trade-offs in different designs. Symbolic approaches emphasize user trust and verifiability of the system. Despite significant progress, problems such as uncertainty management, knowledge acquisition bottlenecks, and real-time adaptability remain challenging. This work highlights the need to incorporate formal strictness with adaptive methods in expoiting the optimal potential of AI debaters in advanced argumentative debate.

Index Terms—Debate Systems, Argumentation, Conversational Agents, Argumentation Service Platform with Integrated Components (ASPIC), Retrieval-Augmented Generation (RAG), Large Language Model (LLM).

I. Introduction

Dialogue systems capable of debating at a level of sophistication are a growing area of research in Artificial Intelligence. Not only do these systems need to understand and generate arguments, but they must stick to principles of logic and persuasion, the same as advanced human interactions. Creating arguing agents that can present justified positions, decide controversies, and rebuttal counterarguments in a logical manner is a significant technical challenge with tremendous potential for application ranging from educational software to sophisticated decision support systems. To meet these challenges, some approaches have been formulated having their merits and disadvantages.

Among the several approaches, rule based agents are one of the earliest and straightforward method. These systems function on the basis of pre-established templates and logical rules picked from expert knowledge. Although they provide good predictability, their use of predefined templates restrict their domain and may lead to irrelevant responses [1].

Alternatively, retrieval-based agents, sometimes combined with techniques like Retrieval-Augmented Generation (RAG), directly retrieve arguments from massive databases or text corpora [2]. Using vast amounts of existing knowledge, this allows for fast generation of contextually diverse responses. However, ensuring the factual and logical consistency of the retrieved arguments remains a significant challenge for such systems.

Hybrid argumentation agents seek to combine clear logical rules with adaptable responses learned from data. By typically integrating retrieval mechanisms and explicit reasoning mechanisms (e.g. rule-based logic or knowledge graphs), these agents aim to produce more flexible, relevant, and logically coherent arguments than purely rule-based or retrieval-based systems alone can [2].

Further complexity is added by hierarchical persuasion agents. These agents model the interaction of argumentation itself, resulting in step-by-step construction of arguments based on dialogue context or feedback from the user. This maximizes the overall quality and persuasiveness of the debate [3].

Finally, explainable debate agents prioritize transparency in their argumentative process [4]. By making their claims and reasoning steps explicitly justified, these agents provide the insights to logical validity of the arguments, which also enhances the user trust.

Although many methodologies are facilitating AI debate capabilities, this research investigates the contributions of structured argumentation theory-based and symbolic reasoning-based methods to make the agent able to conduct a good debate. Three argumentation techniques are discussed in the scope of this paper to make the conversational agents able to debate, starting with argument graphs which provide intuitive visualization of argument relationships. Second is ASPIC+, offering formal structure for logical argument and third is Multi-Attribute Frameworks, enabling multidimensional argument comparison. Together, these improve argumentation robustness and comprehensibility. In detail, analysis is conducted on how these techniques facilitate structured reasoning and knowledge representation, highlighting, in particular, their advantages of explainability and logical consistency [5]. Meanwhile, we acknowledge and resolve inherent challenges for such systems, such as managing uncertainty and dynamically revising reasoning protocols in real-time debate contexts [5].

Although Large Language Models (LLMs) being an important part of the research in conversational agents due to their fluency and generative AI, this paper directs its research on symbolic reasoning approaches. We acknowledge the very great capabilities of LLMs, but the focus of this research is more on formal argumentation structures like ASPIC+, Argument Graphs, and corresponding symbolic methods. These techniques guarantee logical consistency, valid arguments,

and immanent explainability. These are the qualities that are critical in designing debating agents that can deploy the meaningful and logically consistent arguments that lead to good argumentation.

This paper will initially develop a brief definition of debate and effective argumentation criteria. Section III explores leading computational approaches (Argument Graphs, ASPIC, and Multi-Attribute frameworks) before comparatively evaluating a number of agent architectures for debate. The approach further demonstrates that symbolic reasoning is critical for achieving explainability and maintaining logical soundness. Section IV, outlines the broad restrictions and difficulties in creating skilled AI debaters, laying the groundwork for our final findings.

II. DEBATE DEFINITION AND WHAT MAKES A GOOD DEBATE

Debate is a process in which group of debaters take turns presenting systematic arguments and counterarguments on some prescribed subject, under established rules and time constraints, with the objective of advancing, defending, and challenging positions for judgment by an adjudicator or audience members [6],[7]. It can be seen as a scenario where argument exchanges are governed by pre-established rules, such as number of debaters, strict alternation, timing and sequence of the speeches, and the requirement to rebut opposite arguments in a pre-established format [8]. What exactly constitutes a "good" debate remains an open question, as criteria can vary depending on context, format, and evaluative priorities. For example, if the debate is academic, logical consistency and evidence are the important factors. But if the debate is political, clarity of delivery and public persuasion are the most important factors. A good debate consists of empirical evidence-based arguments that are well supported, clear explanations that spell out reasoning using simple language, and an awareness of counterarguments to back persuasive attempts. Moreover, logical coherence between premises and the categorization of argument quality, that is, how well an argument is supported, clear, relevant, and logically sound, are elements that make responses effective as a whole [8].

In order to evaluate debate performance well, judges look for ordered argumentation with an organized structure, logical flow, and successful refutations. The quality of evidence is extremely pertinent and should be based on actual facts, credible sources, and firm resolutions. Teams must prioritize argumentation and provide insights into the "point of clash". Judges are also responsible for recognizing fabricated evidence, ad hominem attacks, or rule violations. Qualitative evaluation focuses on the strength of the argument.

III. CONVERSATIONAL AGENT DEBATING TECHNIQUES AND APPLICATIONS

Debating Agents use argumentation techniques to engage themselves into meaningful and persuasive debate to go through complex debating topics. Under this section, key argumentation techniques are examined.

- **Argument graphs:** A visual representations of interrelations between arguments, which can be formally represented by a graph, in which there are vertices that represent arguments and the edges represents support and attack relations. The visualization through such graphical representations can make complicated interrelations between arguments easy to understand, making them understandable by audience members in a more readable way. Though argument graphs efficiently promote understanding and involvement, their persuasiveness relies on the argument quality. Therefore, they get a positive assessment for their ability to provide clarity in persuasive conversations [9],[8].

- **ASPIC framework:** The ASPIC+ framework provides a formal and structured format for argument presentation and evaluation, particularly valuable in contexts like debate where logical rigor is essential. At its core, ASPIC+ specifies how arguments can be made systematically by using rules to derive claims from facts. Strict rules and defeasible rules are the two rules which are used in ASPIC+. To start with strict rules which represent deductive, logically necessary inferences (e.g., "if X is a square, then it is also a rectangle"), this ensures that the conclusion is correct if the premises is true. Defeasible rules, which use common-sense, or presumptive reasoning where the conclusion typically holds but it still has a probability of getting defeated by an attacking argument (e.g., "birds typically fly"). Arguments are so chains that link premises to a claim via sequences of these strict and defeasible rules. ASPIC+ then explicitly defines various ways arguments can conflict through different types of attack relations, such as challenging a premise used in an argument or rebutting a conclusion derived via a defeasible rule. Crucially, preferences between arguments are often used to determine which attacks are successful, leading to a formal assessment of which arguments are ultimately justified. With this formal setup, revealing the step-by-step application of specific rules, there comes the potential for unambiguous logical demonstration. While the full network of interacting arguments and attacks typically forms a complex argument graph, the structured nature allows specific lines of reasoning supporting a claim to be clearly traced. This explicit structure, revealing the logical connections and dependencies, helps fortify an argument for persuasion by making its reasoning transparent to an audience for evaluation. Evaluative perspective shows that ASPIC+'s strong logical structure is well-suited to enable sound arguments, thereby aiding in the promotion of reasoned persuasion [8].

- **Multi-Attribute Argumentation Framework:** This technique ranks arguments on multiple criteria such as (logical soundness, relevance, clarity), and each criterion is attached with weights and score is evaluated for the argument. The multi-aspect analysis process allows for sensitive rankings so that arguments correctly align with audience's values to increase the persuasiveness of

this technique. This approach is widely valued for its capability to accommodate heterogeneous preferences, enhancing the persuasiveness of arguments [8].

While Argument Graphs Framework is proficient at visual clarity for complex debating structure and Multi-Attribute Argumentation Framework excel at ranking arguments based on criteria to improve the persuasiveness. However the ASPIC+ framework stands out for providing conversational agent with core debating capability. Its primary advantage lies in its formal, rule-based approach, which ensures the construction of logical arguments and provides a mechanism for evaluating their justification through well-defined attack relations and preferences.

IV. COMPARATIVE ANALYSIS AND EVALUATION OF DEBATE AGENTS

To facilitate an easy understanding of the landscape of debate agents, this paper has two significant tables. Table I has a simplified overview of different types of conversational agents utilized in debate use cases, including their approaches, strengths, and weaknesses. Table II takes it further by listing a comparison of evaluation metrics for these agents in tabular form. These tables are designed to be an easy reader reference, enabling us to make useful comparisons and contrasts between different agent architectures and assessment criteria in the field.

TABLE I
OVERVIEW OF DIFFERENT CONVERSATIONAL AGENT TYPES FOR DEBATE APPLICATIONS

Agent Type	Approach	Use Cases	Strengths	Weaknesses
Rule-Based	Predefined logical templates	Compliance systems [1]	Predictable outcomes	Rigid structure
Retrieval-Based	DB queries + RAG	Customer support chatbots [6]	Dynamic data handling	Retrieval dependency
Hybrid (Rule-Retrieval Based)	Rule-retrieval integration	Legal argumentation [2]	Context-aware reasoning	Complex setup
Hierarchical	Feedback-driven adaptation [3]	Policy negotiation	Personalized arguments	Feedback latency
Explainable	Justification chains [4]	Diagnostics, Legal systems	Transparent explanations	Depth limitations

[1] Bronsdon G. *Evaluating AI Agent Performance: Benchmarks for Real-World Tasks*. Accessed: Apr. 3, 2025. URL: https://www.galileo.ai/blog/evaluating-ai-agent-performance-benchmarks-real-world-tasks.

[2] IBM. *AI Agent Evaluation*. Accessed: Apr. 3, 2025. URL: https://www.ibm.com/think/topics/ai-agent-evaluation.

[3] SuperAnnotate. *AI Agent Evaluation*. Accessed: Apr. 3, 2025. URL: https://www.superannotate.com/blog/ai-agent-evaluation.

[4] Smythos. *AI Agent Performance Measurement*. Accessed: Apr. 3, 2025. URL: https://smythos.com/ai-agents/agent-architectures/ai-agent-performance-measurement/.

TABLE II
EVALUATION METRICS FOR CONVERSATIONAL AGENTS IN DEBATE APPLICATIONS

Evaluation criteria	Rule-Based	Retrieval	Hybrid	Hierarchical [1]	Explainable [2,3]
Accuracy	High in narrow domains [2]	Moderate (depends on data quality) [4]	Balanced across tasks	Varies by hierarchy level	Moderate (prioritizes transparency over optimization)
Response Time [1]	Fast (deterministic)	Moderate (retrieval latency)	Slow (multi-layer processing)	Slowest (sequential decision-making)	Moderate (additional explainability overhead)
Cost Efficiency	Low (static rules)	Moderate (retrieval infrastructure)	High (compute-intensive)	High (maintenance complexity)	Moderate (explanation generation costs)
Scalability	Poor (rule complexity)	High (modular data updates)	Moderate (resource-intensive)	Best (layered abstraction)	Limited (Model Size Complexity)
Stability	High (deterministic)	Moderate (data drift sensitivity)	High (adaptive tuning)	High (isolated failures)	Moderate (explanation consistency)
User Satisfaction	Low (inflexible)	Moderate (context-aware)	High (balanced UX)	Moderate (learning curve)	Highest (trust through clarity)

V. Symbolic Methods: Explainability and Validity in Logical Reasoning

Symbolic techniques represent knowledge as ordered, legible symbols derived from sound logical reasoning to argue about a problem. They use formal logic, rule systems, and knowledge graphs. Through explainability, these techniques defend the argument that has been selected, and this is what makes them stand out [5].

In symbolic methods, every step follows logically from earlier premises without contradiction, usually by formal entailment tests or propositional logic truth tables. A formal test entailment examines if a conclusion can be derived from a set of premises based on formal laws of logic, like truth table, to guarantee conclusion is only made when the set of premises utilized are valid. To achieve soundness in debating systems, streams of reasoning are analyzed to check if they support each other to reach a conclusion. It is defined as a sequence of results, each supported by antecedents and founded on accepted rules, forming an intelligible and traceable chain from premises to conclusion. comparing them with symbolic ground truths (for example, mathematical theorems or logical derivations), and cross-verifying against recognized rules [5]. This explainability enhances user trust in the system and its usability [5].

VI. Limitations of Conversational Debating Agents

While debating agents have been promising in a variety of applications, there exist several challenges that restrict them to engage in actual debates efficiently. The most important of these challenges are:

- **Uncertainty and Big Data Management:** AI Debate Systems might work with large data from multiple sources, that may contain noisy or inconsistent information [5].
- **Knowledge Acquisition Bottleneck:** Symbolic AI programs typically require hand-coded rules and knowledge, leading to high human effort and cost in translating real-world problems into system inputs [5].
- **Real-time Dynamic Assessments:** AI debate systems face challenges in reacting dynamically to emerging arguments or rapidly changing contexts in real-time scenarios [5].
- **Logical Inconsistencies:** Internal logical inconsistency is one of the major concerns in designing secure debate systems. The problem highlights the need to incorporate constrained logical verification methods, that prove AI reasoning against predetermined logical constraints or rules, or consistency inference tools as part of the AI system architecture [10].

VII. Discussion

The development of conversational agents capable of conducting sophisticated debate is one of the major areas of AI research, involving not only language understanding but also logic consistency and persuasion [6]. This paper explored a number of approaches, highlighting the relative advantages and disadvantages of rule-based, retrieval-based, hybrid, hierarchical, and explainable agents (Table I). Hybrid and retrieval-based systems, although offering flexibility and access to large data sources [2], it is challenging to ensure logical correctness and factual accuracy. Rule-based systems are predictable and ensure logical consistency but only for a restrictive domain [1]. Our work placed particular emphasis on the function of formal argumentation structures and symbolic reasoning. Techniques like Argument Graphs offer good visualization alternatives, enhancing users' comprehensibility of complex argument relations [9],[8]. ASPIC enables the formal definition of logically valid arguments necessary to build convincing and verifiable chains of arguments [8]. Furthermore, Multi-Attribute Argumentation Frameworks facilitate advanced assessment of arguments against multiple attributes and can augment persuasive power by appealing to values in the audience [6]. Such systematic approaches address important ingredients of a "good debate", such as organization, logical consistency, and adequately supported reasoning [6]. The comparison analysis (Table II) also brings out the inherent trade-offs. For instance, formal structure-based methods (e.g., rule-based or augmented hybrid/explainable systems) can be very precise in usage and very explainable, and thus achieve user trust, but very poor in response time, scalability, or adapt-ability compared to pure retrieval-based systems [4]. Symbolic methods, in particular, are the foundation of explainability and logical soundness necessary for user trust and system verification. The ability to outline the steps of reasoning, as represented by symbolic methods and explicit chains of justification, is to be compared with the typically black-box character of data models alone. However, there remain important weaknesses in AI debate agents. Dealing with noisy data is a major issue. Knowledge acquisition bottleneck, especially for symbol-based systems and rule-based programs hand-coded, remains costly and time-consuming. Furthermore, in order to make agents dynamic and respond in real time, this is inappropriate for many modern architectures [5]. Internal logical consistency is required for system integrity and to avoid integrating logically inconsistent argumentation with heterogeneous information or reasoning protocols [10]. Eliminating these weaknesses is essential to future research on strong and effective AI debaters.

---

[5] Lee, L. and Hockenmaier, J. *Evaluating Step-by-step Reasoning Traces: A Survey*. Accessed: Apr. 3, 2025. URL: https://arxiv.org/html/2502.12289v1.

[6] Debating for Everyone. *How to Judge a Debate*. Accessed: Apr. 3, 2025. URL: https://www.debatingforeveryone.com/resources/how-debating-works/how-to-judge-a-debate.

VIII. Conclusion

This paper examines the skills required for conversational agents to engage in a good debate, pointing to the potential of structured theories of argumentation and symbolic reasoning. This paper examines the foundational concepts defining a good debate [6],[7] and outlined the key computational approaches, including Argument Graphs [9], the ASPIC framework, and Multi-Attribute Argumentation Frameworks [8]. Techniques that are discussed in this paper are important for explainability and logical consistency. They also provide a proper structure for the arguments that are used in the debates. In this paper, we have examined the limitations of the existing debating agents and how symbolic techniques can help overcome these techniques by using logical validity. However, there are still some challenges, particularly in data management, knowledge acquisition, real-time responsiveness, and maintenance of logical consistency [10],[5]. Lastly, the integration of the strengths of symbolic argumentation and structured reasoning can potentially produce AI agents that can not only participate in debates but do so in a manner that is logical, coherent, and consistent. Research in the future must tackle the limitations recognized, perhaps in new hybrid architectures compromising formal rigour with data-driven plasticity, in order to be in a position to maximize the strengths of AI for high-level argumentative discourse.

References

[1] H. H. L. C. Monte-Alto, A. T. Possebom, M. M. M. Morveli-Espinoza, and C. A. Tacla. "A rule-based argumentation framework for distributed contextual reasoning in dynamic environments". *DYNA* 88.217 (2021), pp. 120–130. DOI: 10.15446/dyna.v88n217.90858.

[2] I. Kulatska. "ArgueBot: Enabling debates through a hybrid retrieval-generation-based chatbot". MA thesis. University of Twente, 2019.

[3] K. Sakai, R. Higashinaka, Y. Yoshikawa, H. Ishiguro, and J. Tomita. "Hierarchical argumentation structure for persuasive argumentative dialogue generation". *IEICE Transactions on Information and Systems* 103.2 (2020), pp. 424–434.

[4] B. Ali, S. Pawar, G. K. Palshikar, and R. Singh. "Constructing a dataset of support and attack relations in legal arguments in court judgements using linguistic rules". *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*. Retrieved from https://www.elra.info. European Language Resources Association (ELRA). 2022, pp. 491–500.

[5] S. Kasif. "A Trilogy of AI Safety Frameworks: Paths from Facts and Knowledge Gaps to Reliable Predictions and New Knowledge". *arXiv preprint arXiv:2410.06946* (2024). Department of Biomedical Engineering, Program in Bioinformatics, Department of Computer Science, Boston University. URL: https://arxiv.org/abs/2410.06946.

[6] G. Rakshit, K. K. Bowden, L. Reed, A. Misra, and M. Walker. "Debbie, the debate bot of the future". *Advanced Social Interaction with Agents: 8th International Workshop on Spoken Dialog Systems*. Springer, 2019, pp. 45–52.

[7] C. Tan, V. Niculae, C. Danescu-Niculescu-Mizil, and L. Lee. "Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions". *Proceedings of the 25th International Conference on World Wide Web*. 2016, pp. 613–624.

[8] D. Engelmann, J. Damasio, A. R. Panisson, V. Mascardi, and R. H. Bordini. "Argumentation as a method for explainable AI: A systematic literature review". *17th IEEE Iberian Conference on Information Systems and Technologies (CISTI)*. IEEE, 2022, pp. 1–6.

[9] L. A. Chalaguine and A. Hunter. "A persuasive chatbot using a crowdsourced argument graph and concerns". *Computational Models of Argument*. IOS Press, 2020, pp. 9–20.

[10] E. Ilkou and M. Koutraki. "Symbolic vs Sub-symbolic AI Methods: Friends or Enemies?" *Proceedings of the CIKM 2020 Workshops*. Accessed: Mar. 23, 2025. CEUR Workshop Proceedings. 2020. URL: http://ceur-ws.org.