Towards Neurosymbolic Argumentative Agents

Alison R. Panisson[1,2], Guilherme Trajano[1]

[1]Department of Computing – Federal University of Santa Catarina (UFSC), Brazil
[2]Graduate Program in Computer Science (PPGCC) – Federal University of Santa Catarina (UFSC), Brazil

Abstract

This paper presents our research on Neurosymbolic Argumentative Agents, which bridge a symbolic argumentation framework, grounded in argumentation schemes, with the capabilities of Large Language Models (LLMs). LLMs serve as essential interfaces for understanding and generating natural language argumentation, enabling key tasks such as translating arguments into computational representations, generating natural language arguments, guided argument mining from unstructured text, and, critically, reconstructing enthymemes by inferring missing components using argumentation scheme structures. This neurosymbolic integration leverages the linguistic fluency of LLMs to enhance the formal rigor of symbolic reasoning, advancing Human-AI Hybrid Intelligence toward richer, argumentation-based interactions.

Keywords

Neurosymbolic Agents, Large Language Models, Argumentation, Hybrid Intelligence

1. Introduction

The field of computational models of arguments has experienced a significant evolution over the past decades. Early research focused primarily on abstract argumentation formalisms, such as that proposed by Dung [e.g., 1], which provided crucial insights into argument acceptability but lacked the internal structure necessary for detailed representation. This foundational work naturally progressed toward structured argumentation models, which explicitly define the internal composition of arguments and the nature of attack relations [e.g., 2, 3, 4, 5].

We argue that the conceptualization of Argumentation Schemes (AS) is vital for modeling human-like presumptive reasoning within these structured frameworks [6, 7]. AS represent recurring patterns of presumptive arguments employed in both everyday discourse and specialized contexts, such as legal and scientific reasoning. Their central role in modeling cognitive capabilities has led the Artificial Intelligence (AI) community to regard AS as a key component in the implementation of intelligent-agent technologies.

Our research builds directly on this line of development, particularly through the formulation and implementation of a computational model of argumentation schemes for Multi-Agent Systems (MAS) [8]. This framework, which forms the symbolic core of our agents, was developed during Panisson's postgraduate research (Master's and Ph.D. theses), along with many collaborators, between 2014 and 2019, and formally defined to capture the specific structure of AS within practical MAS frameworks. It preserves the essence of Walton's methodology by enabling the system to manage implicit information, especially that elicited by Critical Questions (CQs), without requiring these elements to be explicitly represented as premises or complex undercutting arguments. This symbolic foundation allows agents, implemented in Agent-Oriented Programming Languages (AOPLs) such as Jason [9], to construct, communicate, and reason with structured arguments.

However, the traditional symbolic paradigm faces considerable challenges when interfacing with human users, who communicate through the fluidity and ambiguity of Natural Language (NL). Hu-

CMNA'25: The 25th International Workshop on Computational Models of Natural Argument, Dec 12, 2025, Online
*Corresponding author.
†These authors contributed equally.
alison.panisson@ufsc.br (A. R. Panisson); huilhermetrajano@gmail.com (G. Trajano)
0000-0002-9438-5508 (A. R. Panisson)

man arguments are often enthymemes, arguments with omitted premises or conclusions, requiring sophisticated inferential capabilities from the listener.

In our previous work, we have pursued to bridge the gap between symbolic argumentative agents and natural language argumentation by incorporating technologies such as chatbot platforms for argument understanding [10, 11, 12, 13], as well as exploring distinct strategies for natural language argument generation [14, 15, 16].

The emergence of LLMs has, however, opened new horizons for Computational Models of Natural Arguments (CMNA). LLMs, built on architectures such as the Transformer and comprising billions of parameters [17, 18], offer unparalleled capabilities in Natural Language Understanding (NLU) and generation. This positions LLMs as a promising bridge between the formal rigor of symbolic reasoning and the complexity of human communication. The integration of these two paradigms defines Neurosymbolic Argumentative Agents, a key step toward realizing Hybrid Intelligence (HI) [19], grounded in the argumentative capabilities of both AI agents and humans.

This paper presents our progress from symbolic argumentative agents [8] to the proposed neurosymbolic architecture by synthesizing the core advancements across four interconnected areas of research:

1. **Symbolic Argumentative Agents (reflection):** We describe the computational model of AS for MAS, developed by our research group from 2014 to the present [8]. This model successfully operationalizes AS, including the inherent implicitness of critical questions, within agent-oriented programming frameworks.
2. **Interface between Natural Language and Symbolic Representation (reflection and horizon):** We outline our early approaches to creating an interface between natural language and the computational representation of arguments [16, 15, 10, 11]. We then present our recent work, in which LLMs are employed to implement this interface by translating natural language arguments into computational, symbolic representations defined by AS [20, 21] and vice versa. We argue that this capability is crucial for integrating human dialogue into agent reasoning systems.
3. **Neurosymbolic Argument Mining (horizon):** We leverage the NLU capabilities of LLMs, guided by the structural patterns of AS, to perform argument mining from unstructured texts (e.g., blog posts or documents) [21]. This approach supports the identification and extraction of argumentative components embedded within extended, non-argumentative discourse, thereby integrating sub-symbolic processes (extraction) with the agent's symbolic reasoning engine.
4. **Neurosymbolic Enthymeme Reconstruction (reflection and horizon):** We outline our initial approaches to encoding and decoding enthymemes within a symbolic representation [22, 23, 24, 25]. We then highlight the advanced functionality in which LLMs, guided by AS structures, infer and reconstruct missing premises or conclusions in enthymemes expressed in natural language [26]. This process produces complete, formally represented arguments (e.g., inferring missing contextual premises such as knows(doctor, health)), thereby making implicit human reasoning explicit for the AI agent's reasoning mechanisms.

Collectively, these contributions outline the approach required for what we envision as the next generation of neurosymbolic argumentative agents, systems capable of fluently processing natural language while preserving the logical rigor and structure necessary for sophisticated autonomous reasoning.

2. Symbolic Argumentative Agents

The proposed symbolic argumentative agents approach operates based on a foundational computational model of AS specifically modeled for MAS [8]. This framework constitutes the symbolic core of the agents and is formally defined and implemented within Jason AOPL and platform [9]. This symbolic foundation enables agents to directly construct, communicate, and reason with structured arguments [8].

The framework centers on representing the general structure of AS [8]:

- **Argumentation Scheme Definition:** An AS is represented as a tuple $\langle sn, C, P, CQ\rangle$, where sn is the scheme identifier, C is the conclusion, P is the set of premises, and CQ is the set of associated Critical Questions.
- **Computational Encoding:** The structure of an AS $\langle sn, C, P, CQ\rangle$ is computationally encoded using a non-ground defeasible inference rule of the form $(p_i, \ldots, p_j \Rightarrow p_k)[sn]$, where the set of premises $P = \{p_i, \ldots, p_j\}$ and the conclusion $C = p_k$. The rule is annotated[1] with the scheme name $[sn]$, which serves as the explicit reference to the associated Critical Questions $\{p_0[sn], \ldots, p_n[sn]\}$.
- **Argument Instantiation:** An argument is an instance generated from an AS, represented as a tuple $\langle S, c\rangle_{\theta_{sn}}$. Here, θ is a most-general unifier, S is the support (containing all instantiated premises $P\theta$ and the inference rule), and c is the instantiated conclusion $C\theta$.

A key contribution of this symbolic model is its mechanism for handling the implicit information inherent to CQs. The approach preserves the essence of Walton's methodology by keeping the CQs implicit at the implementation level, thus avoiding the necessity of representing them explicitly as additional premises or complex undercutting arguments.

The revelation of relevant implicit information is achieved through the matching between the constructed argument instance and its underlying reasoning pattern $[sn]$. An argument $\langle S, c\rangle_{\theta_{sn}}$ is an acceptable instance of its AS to an agent ag (with knowledge Δ_{ag}) only if: (i) all premises $p \in S$ are supported or inferred from Δ_{ag} (the agent ag's knowledge); and (ii) all associated Critical Questions $Cq_i \in CQ$ are positively answered by ag ($\forall Cq_i \in CQ, \Delta_{ag} \models Cq_i\theta$). This structure allows CQs to address crucial factors like criticizing premises, pointing out exceptional situations (e.g., source reliability), and representing contextual conditions for the scheme's use.

This symbolic framework supports agent capabilities in both reasoning and communication:

- **Argumentation-Based Reasoning:** Agents construct and individually evaluate argument acceptability, and subsequently define the collectively acceptable arguments based on argumentation semantics, considering conflicts (attacks) among different arguments. The approach allows for AS of different levels of specificity, including chained argumentation schemes.
- **Argumentation-Based Dialogues:** The structure facilitates communication by explicitly addressing the problem of scheme awareness. Dialogue protocols define specific performatives, such as `question_scheme` and `inform_scheme`, to allow agents to retrieve the specific reasoning pattern used by an opponent when necessary.

Example 1 (Argumentation Scheme Role to Know) *The argumentation scheme Role to Know, denoted as* `role_to_know`*, is computationally represented as follows:*

```
( role(Agent,Role), role_to_know(Role,Domain), asserts(Agent,Conclusion),
  about(Conclusion,Domain) ⇒ Conclusion )[as(role_to_know)]
```

with the argumentation scheme name sn = `role_to_know`*, the conclusion* C = `Conclusion`*, and premises* P = { `role(Agent,Role)`, `role_to_know(Role,Domain)`, `asserts(Agent,Conclusion)`, `about(Conclusion,Domain)`}.

The associated critical questions CQ *are as follows:*

- `role_to_know(Role,Conclusion)`*[as(role_to_know)].*
- `reliable(Agent)`*[as(role_to_know)].*
- `asserts(Agent,Conclusion)`*[as(role_to_know)]*
- `role(Agent,Role)`*[as(role_to_know)].*

[1] Following the formal representation presented in [27].

Example 2 (Argument Instantiation) *As an example of instantiating the Role to Know scheme, consider a scenario in which an agent knows that* mary *plays the role of a doctor –* role(mary, doctor). *Furthermore, the agent is aware that doctors are knowledgeable about cancer –* role_to_know(doctor, cancer). *Suppose* mary *asserts that "smoking causes cancer" –* asserts(mary, causes(smoking, cancer)) *– and the agent recognizes that the causes of cancer pertain to the domain of cancer –* about(causes(smoking, cancer), cancer). *Based on this information, the agent can instantiate the argumentation scheme as follows:*

```
( role(mary,doctor), role_to_know(doctor,cancer),
  asserts(mary,causes(smoking,cancer)), about(causes(smoking,cancer),cancer)
  ⇒ causes(smoking,cancer) )[as(role_to_know)]
```

The agent can then automatically associate the corresponding critical questions. For instance, it may consider whether mary *is a reliable doctor, expressed as* reliable(mary)[as(role_to_know)], *according to the unification function* $\theta =$ Role $\mapsto$ doctor, Agent $\mapsto$ mary, Domain $\mapsto$ cancer, Conclusion $\mapsto$ causes(smoking, cancer).

3. Towards a Bidirectional Interface between Natural Language and Symbolic Representation

The necessity to integrate human dialogue into agent reasoning systems is crucial for the development of HI [19, 20, 21]. Our initial work and subsequent research focused on creating an effective interface between natural language used by humans and the computational (symbolic) representation of arguments employed by MAS [8]. Early approaches focused on two main directions: translating symbolic representation to NL (the output side) for explainability, and using chatbot technologies to classify and understand NL arguments (the input side).

We developed a method to translate arguments from a computational representation to natural language using natural language templates associated with AS [15]. The main objective was to enable agents to explain their reasoning and decision-making to human users by translating computational arguments into human-readable NL arguments [15, 8]. For instance, the NL template for an AS uses variables (e.g., <Agent>, <Role>) that are populated during the unification process, transforming instantiated computational arguments into comprehensible NL forms [15]. This method focused on the agent's output side, supporting Explainable AI (XAI) [8, 15].

Example 3 (Natural Language Template) *An example of a natural language template for the argumentation scheme Role to Know is presented below:*

⟨*"<Agent> is a <Role>, and <Role>s know about <Domain>. <Agent> asserts that <Conc>; therefore, we should believe that <Conc>."*⟩[as(role_to_know)]

Following the Example 2, this template is instantiated as follows:

⟨*"Mary is a doctor, and doctors know about cancer. Mary asserts that smoking causes cancer; therefore, we should believe that smoking causes cancer."*⟩[as(role_to_know)]

On the input side, we investigated the the use of chatbot technologies (specifically the Rasa framework) and its NLU module to classify arguments in NL according to the AS used to instantiate them [11, 10]. This approach aimed to allow agents to understand arguments uttered by humans, including incomplete arguments known as enthymemes [11]. Experiments showed that the NLU model achieved good accuracy in classifying both full arguments and enthymemes based on their underlying AS structure, although challenges remained in accurately extracting specific numbered premises from enthymemes [11]. This approaches were the first steps towards addressing the translation challenge: converting complex, unstructured human arguments (in NL) into the formal symbolic representations

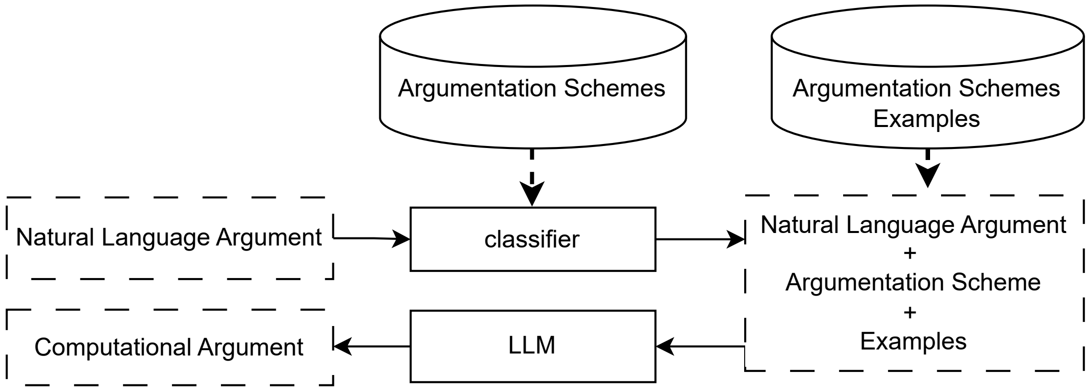

Figure 1: Overview for Translating Natural Language into Computational Arguments [20].

necessary for agent reasoning (symbolic) [20, 21]. This capability is crucial for enabling agents to reason over human input.

Later, we proposed the use LLMs to perform this translation from NL arguments into computational, symbolic arguments [20]. The approach is grounded in the use of AS to classify the input argument, providing crucial context and structure to the LLM for the symbolic translation task [20]. By leveraging a Retrieval Augmented Generation (RAG) methodology, the LLM is guided to instantiate variables and predicates accurately into the required computational representation [20]. Figure 1 ilustrates an overview of the proposed approach. The evaluation demonstrated that LLMs efficiently translate simple argument structures and improve performance on complex arguments when provided with increased context (more examples) [20].

This neurosymbolic approach also address the reconstruction of enthymemes, which involves combining the generative capabilities of LLMs with the precise, structured guidance of AS [21]. This is essential because human communication frequently relies on implicit information and enthymemes [21]. In this framework, the LLM is guided by the AS structure to perform inference, reconstructing the missing components (premises or conclusions) and subsequently generating a complete computational representation of the intended argument [21]. This capability, enabled by the LLMs' semantic understanding and commonsense knowledge, successfully bridges the gap between structured symbolic frameworks and informal natural language expressions, laying the groundwork for sophisticated human-agent interaction within HI systems [21]. The ultimate goal is to enable distributed AI systems to interact meaningfully with human users through shared argumentative structures [21].

4. Neurosymbolic Argument Mining

Another direction of our research is to leverage the NLU capabilities of LLMs, guided by the structural patterns of AS, to perform argument mining from unstructured texts (e.g., blog posts or documents) [21]. This approach supports the identification and extraction of argumentative components embedded within extended, non-argumentative discourse, thereby integrating sub-symbolic processes (extraction and translation via LLMs) with the agent's symbolic reasoning engine [8].

The approach provides two main novel contributions: (i) demonstrating how argument mining is performed using AS structures to guide LLMs in argument extraction, and (ii) showing how the extracted arguments are translated from natural language into formal symbolic representations. The method operates as an LLM-based argument extraction pipeline that processes input text, searches for argumentative sentences, performs AS classification, and translates the findings into a computational form.

1. **Text Pre-processing:** The input document (e.g., everyday discourse or formal documents) is pre-processed by an initial LLM instance.

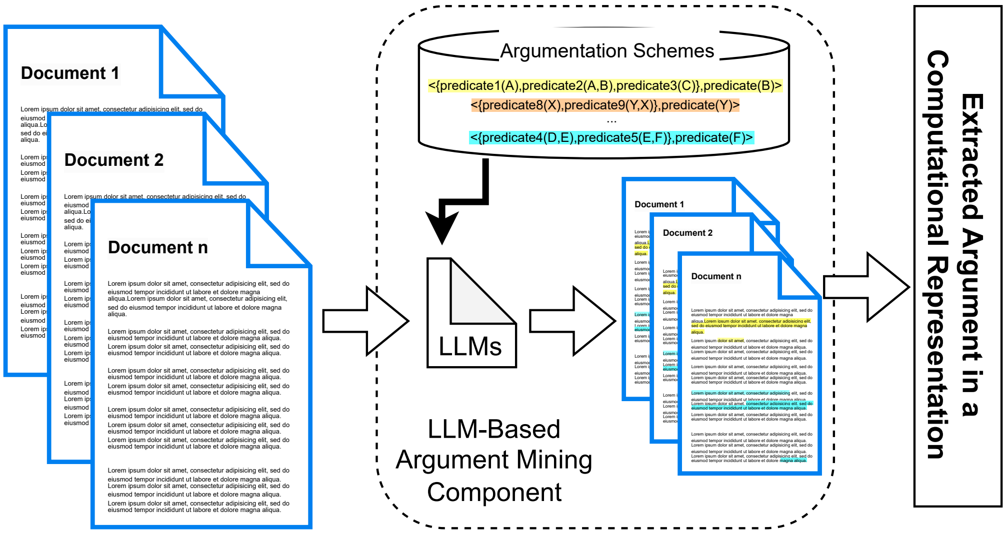

Figure 2: Overview of the Argument Mining Approach [21].

2. **AS-Guided Extraction:** The pre-processed text, along with a predefined set of AS and a prompt, is provided to a second LLM instance. The prompt, defining the task, provides AS examples (including natural language description and computational representation), and instructs the model to match text excerpts with the appropriate scheme.
3. **Enthymeme Handling (Inference):** The LLM is specifically tasked with finding (or inferring) missing premises or conclusions, i.e., enthymemes, based on the contextual information and the structure defined by the AS. This capability allows the construction of complete argument structures from naturally occurring, often incomplete, discourse.
4. **Symbolic Translation and Instantiation:** The LLM performs instantiation by selecting the appropriate scheme (e.g., Role to Know) and mapping variables (e.g., `Agent`, `Role`, `Domain`, `Conclusion`) to specific terms and predicates found or inferred from the text.

The interface translates the mined arguments into a formal computational representation compatible with existing computational argumentation frameworks used by agents, such as our approach [8]. This representation follows a structural approach expressed as a first-order formalism discussed in Section 2. This symbolic output is then processed by neurosymbolic argumentative agents to perform tasks such as evaluating argument acceptability by checking associated CQs or determining relationships (support, attack) between arguments in a multi-agent system. This capability enables AI agents to directly interpret, reason about, and communicate arguments extracted from diverse sources, leveraging them for advanced reasoning and interaction. This strategy lays the foundation for distributed AI systems capable of engaging meaningfully with human users, both synchronously and asynchronously, within HI and shared (real or virtual) environments.

5. Neurosymbolic Enthymeme Reconstruction

Firstly, we have proposed approaches to encoding and decoding enthymemes within a symbolic representation [22, 23, 24, 25], enabling more efficient argumentation-based dialogue among computational agents. We then moved to the advanced functionality in which LLMs, guided by AS structures, infer and reconstruct missing premises or conclusions in enthymemes expressed in natural language [26]. This process produces complete, formally represented arguments (e.g., inferring missing contextual premises such as role_to_know(doctor, cancer)), thereby making implicit human reasoning explicit for the AI agent's reasoning mechanisms.

Example 4 (Natural Language Enthymeme) *An example of natural language enthymeme, instantiated of the argumentation scheme Role to Know, is as follows:*

"Mary asserts that smoking causes cancer; therefore, we should believe that smoking causes cancer"

In contrast to the natural language argument generated in Example 3, the following argument omits the premises "Mary is a doctor" and "doctors know about cancer." However, the proposed approach is capable of efficiently reconstructing the complete argument by using the AS as a guide. It does so by aligning variables across different predicates and leveraging the general contextual knowledge encoded within the LLM. For this particular example, the system generates an output equivalent to the computational argument presented at the end of Example 1.

Enthymemes, defined formally as an incomplete argument $\langle S', c'\rangle_{\theta_{sn}}$ where components (Δ) are omitted from the full argument $\langle S, c\rangle_{\theta_{sn}}$ such that $(S' \cup c') = ((S \cup c) \setminus \Delta)$, represent a complex yet natural form of human communication. The challenge for intelligent agents in hybrid systems is the decoding process, reconstructing the missing content based on implicit shared knowledge or context.

Our methodology proposes a neurosymbolic interface that leverages the LLMs' advanced capabilities in NLU to handle ambiguity and context-dependency inherent in enthymematic discourse, combining them with the structured reasoning provided by AS. AS serve as abstract, structuring rules, acting as templates to guide the reconstruction process and enable agents to access the implicit knowledge embedded within reasoning patterns.

The core of the approach is an LLM-based reconstruction component that processes a natural language enthymeme $\langle S', c'\rangle$ and outputs a complete computational argument $\langle S, c\rangle_{\theta_{sn}}$.

1. **AS Classification:** The LLM first classifies the input enthymeme based on a predefined set of AS (Δ_{AS}) to identify the underlying reasoning pattern (e.g., Argument from Role to Know). This is guided by prompt engineering that includes constructive examples of AS in both natural language and computational form.
2. **Inference and Reconstruction:** Guided by the classified AS structure, the LLM infers the missing components (Δ), premises or conclusions, that were omitted in the original enthymeme, making the implicit human reasoning explicit. The LLM utilizes its extensive linguistic and world knowledge to hypothesize these missing components, effectively modeling what is "taken for granted" in the communication context.
3. **Symbolic Instantiation:** The LLM performs instantiation by mapping the variables in the selected scheme to specific terms and predicates found in, or inferred from, the text. For instance, in Example 4, the LLM infers the missing premises `role_to_know(doctor, cancer)` and `about(causes(smoking, cancer), cancer)` to complete the structure of the Role to Know scheme.
4. **Computational Output Generation:** The final output is the complete argument represented in a computational form, ensuring compatibility with computational argumentation frameworks used, for example [8].

This reconstruction process enhances agent capabilities, allowing them to comprehend human communication more fully and make informed decisions based on the enriched, structured input. Our evaluation, in [26], demonstrates that LLMs achieve good accuracy in classifying enthymemes and effectively reconstruct arguments, even with significant implicitness (e.g., 50% of missing premises/conclusion), provided they are anchored to these structured AS guides.

6. Are Neurosymbolic Approaches the Future for Argumentative Agents?

The future of argumentative agents remains uncertain. While LLMs evolve and expand their applicability across diverse domains, they continue to transform our understanding of what machines can achieve in communication, reasoning, and collaboration. LLMs have already demonstrated extraordinary capabilities in comprehending and generating natural language, supporting an ever-growing range of cognitive and decision-making tasks that were traditionally handled by symbolic AI [28]. Their flexibility, contextual awareness, and apparent ability to reason across vast semantic spaces position them as powerful tools for implementing the next generation of argumentation-based systems.

However, while these sub-symbolic systems offer unprecedented performance in language processing and knowledge synthesis, their inherent opacity and lack of formal guarantees challenge their suitability for domains where reasoning must be transparent, accountable, and logically sound [29]. Argumentation, as a field rooted in logical structure and epistemic justification, demands such rigor. This tension reveals a key insight: the future of argumentative agents might not be defined by the dominance of either symbolic or sub-symbolic paradigms alone, but by their synthesis.

Hybrid, neurosymbolic approaches emerge as a compelling middle ground, combining the interpretability, verifiability, and formal soundness of symbolic reasoning with the expressiveness, adaptability, and contextual fluency of neural models. In this hybrid vision, LLMs act not as autonomous reasoners but as perceptual and linguistic interfaces that ground and enrich the agent's symbolic reasoning processes [29]. AS provide the structural basis through which LLMs can anchor their generative potential to logical form, ensuring that reconstructed, translated, or mined arguments remain consistent with the agent's epistemic framework.

Beyond this technical synthesis, hybrid architectures open new horizons for integrating human-inspired reasoning mechanisms, such as Theory of Mind (ToM), the capacity to model and anticipate the beliefs, intentions, and emotions of others. Embedding ToM-like capabilities within argumentative agents allows them to interpret communicative intent, adapt argumentative strategies, and engage in more context-sensitive, cooperative, and persuasive dialogues [25, 22]. By reasoning about the mental states of their interlocutors, these agents can better align their arguments with human expectations, thus advancing both communicative effectiveness and epistemic alignment.

The development of such architectures points toward a new class of Neurosymbolic Argumentative Agents, capable of combining diverse forms of reasoning, deductive, abductive, and presumptive, with mechanisms for perspective-taking and communicative adaptation. These agents could operate in dynamic, open environments, where they interpret, evaluate, and generate arguments in ways that align both with human communication norms and with formal reasoning requirements.

However, achieving this vision will require overcoming significant challenges. Future research must address the alignment of LLM outputs with formal semantics, the incorporation of grounding and truth-maintenance mechanisms, and the establishment of robust evaluation methodologies that assess both argumentative quality and epistemic reliability. Moreover, as neurosymbolic systems become more deeply embedded in human decision-making processes, ethical and epistemological questions regarding responsibility, bias, and interpretability will become central to their design.

In this context, we envision a trajectory where neurosymbolic argumentative agents become foundational components of HI systems, ones that do not merely imitate human reasoning but extend it through structured, transparent collaboration between human and artificial agents. It seems to be a promising direction for developing argumentative agents that are not only powerful but also trustworthy, explainable, and capable of reflective, socially aware reasoning.

7. Conclusion

In this paper, we provided an overview of our research on argumentative agents, reflecting on the conceptual and technical evolution of our work from symbolic to neurosymbolic paradigms. Our reflection began with the foundational computational model of AS for MASs [8], which established the symbolic core of our agents. This model operationalizes Walton's theory [6, 7] of presumptive reasoning within computational frameworks, enabling agents to construct, reason with, and communicate structured arguments while maintaining logical rigor and transparency.

Building upon this foundation, we extended our research toward new horizons that explore the intersection between natural language and symbolic reasoning. We incorporated Natural Language Processing (NLP) methods and, more recently, LLMs, to bridge the gap between formal argument representation and human communication. These efforts culminated in neurosymbolic approaches that employ LLMs for argument translation [20], argument mining [21], and enthymeme reconstruction [26], tasks that combine the contextual and linguistic capabilities of neural models with the epistemic precision of symbolic reasoning.

Through these developments, we outlined a trajectory toward Neurosymbolic Argumentative Agents, which we see as a promising horizon for the next generation of research on computational models of natural argument. In such systems, LLMs provide adaptive understanding and generative capacity, while symbolic structures guarantee consistency, interpretability, and accountability. Together, they embody a step toward HI, where human and artificial reasoning processes are integrated through shared argumentative structures.

Our reflection over the past years and our horizon for the future converge on a central insight: progress in computational argumentation depends not only on increasing linguistic or computational capacity but also on preserving the principles of transparency, justification, and reasoning integrity that define argumentation itself. As we look toward the coming decades, we envision hybrid, neurosymbolic argumentative agents that advance these principles, agents capable of reasoning, explaining, and engaging in dialogue with humans in ways that are both intellectually rigorous and deeply aligned with the goals of collaborative, trustworthy AI.

Declaration on Generative AI

During the preparation of this work, the author(s) used Generative AI in order to grammar, formal tone, and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artificial Intelligence 77 (1995) 321–357.

[2] S. Modgil, H. Prakken, The aspic+ framework for structured argumentation: a tutorial, Argument & Computation 5 (2014) 31–62.

[3] P. Besnard, A. Hunter, Constructing argument graphs with deductive arguments: a tutorial, Argument & Computation 5 (2014) 5–30.

[4] A. J. García, G. R. Simari, Defeasible logic programming: Delp-servers, contextual queries, and explanations for answers, Argument & Computation 5 (2014) 63–88.

[5] F. Toni, A tutorial on assumption-based argumentation, Argument & Computation 5 (2014) 89–117.

[6] D. Walton, Argumentation schemes for presumptive reasoning, Routledge, 1996.

[7] D. Walton, C. Reed, F. Macagno, Argumentation Schemes, Cambridge University Press, 2008.

[8] A. R. Panisson, P. McBurney, R. H. Bordini, A computational model of argumentation schemes for multi-agent systems, Argument & Computation 12 (2021) 357–395.

[9] R. H. Bordini, J. F. Hübner, M. Wooldridge, Programming Multi-Agent Systems in AgentSpeak using Jason (Wiley Series in Agent Technology), John Wiley & Sons, 2007.

[10] L. H. H. de Sousa, G. Trajano, A. S. Morales, S. Sarkadi, A. R. Panisson, Using chatbot technologies to understand natural language arguments, in: International Conference on Agents and Artificial Intelligence, Springer, 2024, pp. 320–341.

[11] L. H. H. de Sousa, G. Trajano, A. S. Morales, S. Sarkadi, A. R. Panisson, Using chatbot technologies to support argumentation, in: 16th International Conference on Agents and Artificial Intelligence (ICAART 2024), SciTePress, 2024.

[12] M. d. S. Custódio, et al., Rasa4jaca: Uma interface entre sistemas multiagentes e tecnologias chatbots open source (2022).
[13] D. Engelmann, J. Damasio, T. Krausburg, O. Borges, M. Colissi, A. R. Panisson, R. H. Bordini, Dial4jaca–a communication interface between multi-agent systems and chatbots, in: International conference on practical applications of agents and multi-agent systems, Springer, 2021, pp. 77–88.
[14] A. R. Panisson, D. C. Engelmann, R. H. Bordini, Engineering explainable agents: an argumentation-based approach, in: International Workshop on Engineering Multi-Agent Systems, Springer, 2021, pp. 273–291.
[15] C. E. A. Ferreira, D. C. Engelmann, R. H. Bordini, J. L. Carbonera, A. R. Panisson, A knowledge base of argumentation schemes for multi-agent systems, in: International Conference on Enterprise Information Systems (ICEIS), 2024, pp. 1–15.
[16] C. E. A. Ferreira, A. R. Panisson, D. C. Engelmann, R. Vieira, V. Mascardi, R. H. Bordini, Explaining semantic reasoning using argumentation, in: International Conference on Practical Applications of Agents and Multi-Agent Systems, Springer, 2022, pp. 153–165.
[17] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. e. a. Dong, A survey of large language models, arXiv preprint arXiv:230318223 (2023).
[18] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, et al., Gpt-4 technical report, arXiv preprint arXiv:230308774 (2023).
[19] Z. Akata, D. Balliet, M. de Rijke, F. Dignum, V. Dignum, G. Eiben, A. Fokkens, D. Grossi, K. Hindriks, H. e. a. Hoos, A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence, Computer 53 (2020) 18–28.
[20] G. Trajano, D. C. Engelmann, R. H. Bordini, S. Sarkadi, J. Mumford, A. R. Panisson, Translating natural language arguments to computational arguments using llms, in: Computational Models of Argument (COMMA), IOS Press, 2024, pp. 289–300.
[21] G. Trajano, S. Sarkadi, A. L. Gonçalves, A. R. Panisson, Towards argument mining agents: A neuro-symbolic approach based on argumentation schemes and llms, in: IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2025.
[22] A. R. Panisson, P. McBurney, R. H. Bordini, Towards an enthymeme-based communication framework in multi-agent systems, in: Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning, KR 2022, Haifa, Israel. July 31-August 5, 2022, 2022, Israel., 2022.
[23] A. R. Panisson, P. McBurney, R. H. Bordini, Towards an enthymeme-based communication framework, in: 21st International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2022, Auckland, New Zealand, May 9-13, 2022, 2022, Nova Zelândia., 2022.
[24] A. R. Panisson, Argumentation schemes and enthymemes in multi-agent systems, in: Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, 2017, pp. 1849–1850.
[25] A. R. Panisson, R. H. Bordini, Uttering only what is needed: Enthymemes in multi-agent systems, in: Proceedings of the 16th International Conference on Autonomous Agents & Multiagent Systems (AAMAS-2017), 2017, Brasil., 2017.
[26] G. Trajano, S. Hoffmann, S. Sarkadi, A. R. Panisson, Toward enthymeme reconstruction: A neuro-symbolic interface, in: IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2025.
[27] A. R. Panisson, M-arguments, in: 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), IEEE, 2020, pp. 161–168.
[28] S. J. Russell, P. Norvig, Artificial intelligence: A modern approach, global edition 4e (2021).
[29] U. Nawaz, M. Anees-ur Rahaman, Z. Saeed, A review of neuro-symbolic ai integrating reasoning and learning for advanced cognitive systems, Intelligent Systems with Applications (2025) 200541.