The A-BDI Metamodel for Human-Level AI:

Argumentation as Balancing, Dialogue and Inference

Liuwen Yu[1 [0000−0002−7200−6001]] and Leendert van der Torre[1,2 [0000−0003−4330−3717]]

[1] University of Luxembourg, Luxembourg
[2] Zhejiang University, China

Abstract. In this paper, we introduce A-BDI, the first metamodel for formal and computational argumentation. It contains three models, conceptualizing argumentation as balancing, argumentation as dialogue, and argumentation as inference respectively. Each model looks at argumentation from a different perspective, addressing its own concerns and using its own formal and computational methods. Whereas balancing is inspired by the scale metaphor and uses quantitative techniques typically found in theories in economics and neural computing, dialogue is developed in multiagent communication and interaction and uses chatbot and Large Language Models (LLMs) technology, and inference is derived from theoretical investigations in knowledge representation and reasoning and uses techniques from symbolic reasoning. By bringing together new and traditional Artificial Intelligence (AI) approaches, the A-BDI metamodel provides a formal and computational framework for human-level, neuro-symbolic, and hybrid AI.

Keywords: Metamodel for Argumentation· Methodology of Argumentation · Knowledge Representation and Reasoning · Hybrid Artificial Intelligence

1 Introduction

The research area of argumentation in Artificial Intelligence (AI) is highly diverse: there is natural, formal, and computational argumentation; there are various models of argumentation, and there is a diversity in the formal methods used [6, 16, 54]. To deal with this diversity, we are developing our A-BDI metamodel for formal and computational argumentation, which can be used in knowledge engineering of human-level AI systems. In this conference paper, we do not have the space to discuss the technical aspects of our A-BDI metamodel, but instead, our overall research question is: What are the motivation, methodology, and application of A-BDI metamodel? This overall question breaks down into the following questions:

1. What are the models and conceptualizations of argumentation used in traditional and new-generation AI?

2. What are the relations, similarities, and distinctions between the models?
3. How to use the A-BDI metamodel?
4. What are the motivations for a metamodel?

In our approach, the three models are BDI: balancing, dialogue, and inference. A metamodel provides a higher-level abstraction of how these models interrelate and can be integrated. Concretely, the metamodel we present uses extended Dung's abstract argumentation graphs to relate the models. These extended argumentation frameworks can be flattened into basic ones, demonstrating the universality of attack [54].

To answer the above four questions, we adopt the following approach, which also structures our paper.

1. We use the literature study underlying our chapter in the handbook to roughly delineate the three models of argumentation used in AI, and we illustrate all three approaches on a single example, highlighting their individual specifics.
2. We perform an analysis of the results obtained from the literature study.
3. As a first step towards a user guide for formal and computational argumentation in AI, we discuss how to choose a model for an application.
4. We use our work on the project Logics for New Generation Artificial Intelligence (LNGAI) [3] and workshop Causality, Agents and Large Models (CALM) [4] to provide motivations from a human-level, neurosymbolic hybrid AI.

Our aim in writing this paper and presenting it at CLAR is to receive feedback from the community on our proposal. In this sense, our paper is written as a discussion paper, with several questions and open issues identified. This paper builds on our work as editor of the handbook of formal argumentation, coPI of the logic for new generation AI projects in China, and creators of a new workshop series on causality and Argumentation using LLMs. Our vision for the shift in our work from traditional to current AI and from traditional argumentation methods to new ones resulted in thirteen challenges for formal and computational argumentation, as introduced and discussed in our chapter in the third volume of the handbook.

2 The Balancing, Dialogue, and Inference (BDI) models

This section presents three models of argumentation. Each of them embodies a unique perspective on the construction and purpose of argumentation, a set of formal methods, and application across different disciplines and contexts. Table 1 provides an overview of the three models. We use a legal child custody case [53] to illustrate the three models.

[3] https://www.zlaire.net/lngai/
[4] https://www.ciad-lab.fr/prima-causal-ai-workshop/

Models	Process	Theories and Formal Approaches	Application
Argumentation as balancing	Balancing pros and cons to reach a justified decision	Multi-criteria decision theory, case-based reasoning	Deliberative decision-making in law, ethics, and economics
Argumentation as dialogue	Dynamic verbal interaction between stakeholders to exchange information or resolve conflicts of opinion	Speech act theory, game theory, axiomatic semantics, operational semantics	Debating technologies, chatbots, recommender systems
Argumentation as inference	Logical structure and reasoning to derive conclusions from incomplete and inconsistent premises	Graph theory, nonmonotonic logic, computational logic, causal reasoning, Bayesian reasoning	Automated reasoning systems, knowledge representation, expert systems

Table 1: Three models of A-BDI

2.1 Argumentation as Balancing

Argumentation as balancing is identified by Gordon [19]. It involves weighing the pros and cons of an issue in order to reach a balanced decision or judgment. It is applicable not only when resolving conflicts of opinion in persuasion dialogues but also, e.g., when deciding courses of action in deliberation dialogues [19]. In such a system, pro and con arguments for alternative resolutions of the issues (options or positions) are put forward, evaluated, resolved, and balanced. The formal methods used are multi-criteria decision theory and case-based reasoning, and they are applied in the realms of law, ethics, and economics.

One model of argumentation as balancing is the Carneades Argumentation System [20]. The conception of argument graphs in Carneades is similar to Pollock's conception of an inference graph. There are nodes in the graph representing statements (propositions) and links that indicate inference relations between statements. In particular, the system distinguishes between pro and con arguments. Semantically, con arguments are instances of presumptive inference rules for negating the conclusion. If the premises of a con argument hold, this justifies rejecting the conclusion or, equivalently, accepting its logical complement. With pro and con arguments, some statements need to be ordered or otherwise aggregated to resolve the conflict. Then there are several proof standards used to balance the pros and cons. Here are three examples:

SE (Scintilla of Evidence): A statement meets this standard iff it is supported by at least one defensible pro argument.

BA (Best Argument): A statement meets this standard iff it is supported by some defensible pro argument with priority over all defensible con arguments.

DV (Dialectical Validity): A statement meets this standard iff it is supported by at least one defensible pro argument and none of its con arguments are defensible.

We use a child custody example to illustrate how Carneades works, as visualized in Figure 1.

Example 1 (Child custody in Carneades). Statements are depicted as boxes and arguments as circles. For the purpose of this discussion, we assume that all the premises are ordinary without distinguishing between different types of premises. Premises are shown as edges without arrowheads. Pro arguments are indicated by circle arrowheads while con arguments are shown with standard arrowheads. Argument a_1 asserts that the child knows what she wants and she wants to live with her mother, making it a pro argument for the statement "It is in the child's best interest that she lives with her mother". In contrast, argument a_2 argues that the mother is less wealthy than the father, serving as a con argument against that statement. In this scenario, a_1 is given priority over a_2. Consequently, according to the BA proof standard, the statement "It is in the child's best interest that she lives with her mother" is accepted.

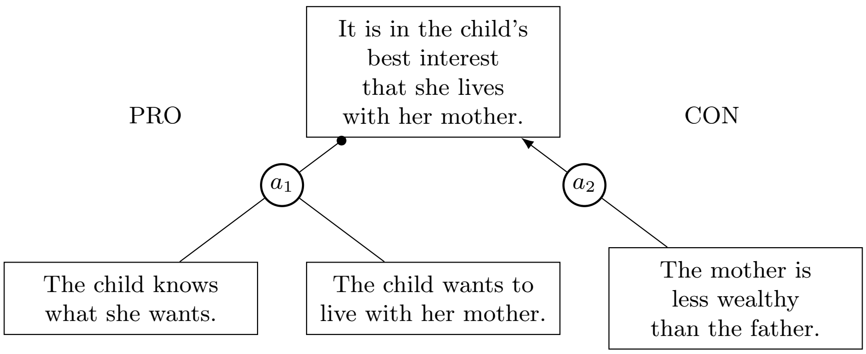

Fig. 1: Child custody case represented in Carneades argument graphs

Moreover, balancing extends beyond what Gordon proposed in Carneades, particularly in how the reason-based approach and balancing are gaining increasing importance in philosophical discussions. The fundamental metaphor of balancing is that reasons possess polarity (favoring or disfavoring an action) and weight (magnitude) [46], and these factors are aggregated to determine an act's deontic status, such as whether it is permissible, impermissible, or required. This metaphor is visualized as a balance scale [47], where reasons supporting an action push one side down while opposing reasons push the other side up, ultimately determining the normative outcome.

Particularly, there is a dispute on whether a reason's weight is context-sensitive. Atomists say no, while holists say yes. The former says that a reason in one context is a reason with the same polarity and weight in every other context; instead, the latter says that a reason in one context but, in some other context, has a different magnitude (weak vs strong) or polarity (for vs against) or is not a reason at all [13]. The moderate view [14] acknowledges that reasons have some context-independent weight, but their strength can be increased, reduced, or removed systematically through contextual considerations. Two key mechanisms shape this process: amplifiers (increase a reason's weight or attenuators that decrease it) and undercutters (nullify a reason's weight entirely), which modify the force of reasons, and two types of weights—the default weight, which represents a reason's initial strength, and the final weight, which accounts for contextual influences [46]. A major challenge in formal modeling is that existing frameworks fail to fully capture these nuances. A formal approach to this moderate view is proposed by Alcaraz et al. [1] and Streit [41] builds on weighted argumentation [3]. The model provides a normative structure that organizes the contextual considerations and their impact on the reasons. Example 2 illustrates the basic idea.

Example 2. Figure 2 represents the balancing scale model with undercutters, amplifiers, and attenuators graphically. The gray circles represent the two options for the child: living with the mother or the father, each placed on one side of the scale. Green and red circles stand for, respectively, positive and negative reasons. "Father is wealthy" is a positive reason for the child to live with the father, while "Father is busy" is a negative reason. "Child wants to live with the mother" is a positive reason for living with the mother. Blue, yellow, and purple circles represent, respectively, undercutters, attenuators, and amplifiers. "Father just went bankrupt" undercuts "Father is wealthy," nullifying its weight. "The Civil Code states that the judge must take into consideration the child's opinion" amplifies "Child wants to live with the mother," increasing its weight. "Public opinion states that ten-year-old children do not know what they want" attenuates "Child wants to live with the mother," reducing its weight.

There are more philosophical discussions on reasons and balancing. For example, Tucker distinguishes two kinds of reasons: justifying reasons and requiring reasons [46]. The difference can be illustrated by an example. There are two people trapped in a burning building. You can do nothing (save none), save one of them, or save both of them. If you save at least one, you will lose your legs. This scenario illustrates the tension between collective values (saving lives) and individual considerations (preserving one's health), and it raises the question of how to model and weigh these competing types of reasons. These two roles are distinct but complementary, highlighting the multidimensional nature of reasons. To model this interaction, Tucker introduces the Dual Scale model [46], which uses one scale to evaluate whether an action is permissible (based on the balance of justifying weight for the action and requiring weight for its alternative) and another scale to assess whether the alternative is impermissible.

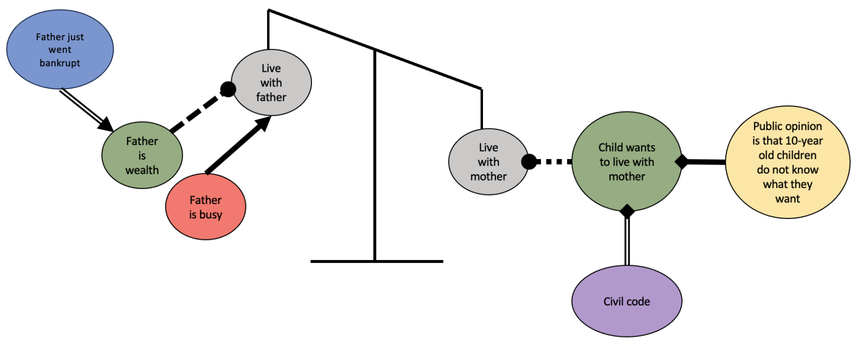

Fig. 2: A balancing scale model representing the child custody case with undercutters and modifiers

Interestingly, while "balancing" itself is not always explicitly mentioned, many theoretical approaches in formal argumentation often involve balancing, such as bipolar argumentation [11, 51, 53], weighted argumentation [8], and ranking-based argumentation [2]. Quantitative approaches in formal argumentation, often explored through extended abstract argumentation frameworks, bring a balancing perspective to reasoning. These models expand upon Dung's original framework by moving beyond binary notions of acceptance or rejection. Instead, they assign numerical scores, weights, or probabilities to arguments and their interactions. The question is: Which quantitative method in argumentation best serves the philosophical discussion of balancing?

2.2 Argumentation as Dialogue

Argumentation as dialogue conceptualizes argumentation as verbal interaction aimed at resolving opinion conflicts [37]. It involves defining argumentation protocols—rules governing the argumentation game—and strategic aspects that guide engagement. This approach draws on speech act theory, game theory, and both axiomatic and operational semantics. It is particularly suited for debates, chatbots, persuasion, and negotiation systems. We refer to the chapter by Black et al. for an overview of argumentation as dialogue [9]. For a comprehensive overview of argumentation as dialogue, we refer to the paper by Black et al. [9]. With recent advancements in LLMs and chatbot technology, this topic has gained renewed interest and presents promising opportunities for practical applications.

Chatbots are conversational software that seeks to understand input from human users and generate human-like responses [9]. In chatbot development, *questions* play a crucial role in enhancing the effectiveness of argumentation-based chatbots and building engaging conversations [27]. The use of *questions* allows chatbots to guide dialogues, challenge assertions, support critical thinking, and provide justifications. Below, we discuss how *question* mechanisms could be embedded into chatbots — justification-seeking questions defined by speech act theory.

Speech act theory, a subfield of pragmatics, examines how language conveys information and performs actions [5]. It has been formalized in the Foundation for Intelligent Physical Agents (FIPA) standards, widely used in computer science for agent communication [15]. A single utterance can serve multiple functions in dialogue [23], classified under basic types like assertions, questions, and commands, or complex ones like promises and declarations [39]. Unlike physical actions, speech acts primarily affect agents' mental and interactional states [45], shaping beliefs, desires, and intentions [17]. In human-like chatbots, speech acts structure interactions, with justification-seeking questions prompting explanations that enrich engagement and deepen understanding [21].

McBurney and Parsons [26] proposed an interaction protocol called Fatio comprising of five locutions for argumentation which can be considered as a set of speech acts.

F1: assert (P_i, ϕ)**:** Speaker P_i asserts a statement ϕ, creating a dialectical obligation to justify ϕ if requested.
F2: question (P_j, P_i, ϕ)**:** Speaker P_j questions a prior **assert**(P_i, ϕ), seeking justification. This act imposes no obligation on P_j.
F3: challenge (P_j, P_i, ϕ)**:** Speaker P_j challenges **assert**(P_i, ϕ), both seeking justification and assuming an obligation to justify not asserting ϕ, e.g., by providing a counterargument. Thus, **challenge** is stronger than **question**.
F4: justify $(P_i, \Phi \vdash \phi)$**:** Speaker P_i, having asserted ϕ and been questioned or challenged, provides a justification $\Phi \in A$ for ϕ.
F5: retract (P_i, ϕ)**:** Speaker P_i, having previously asserted or justified ϕ, withdraws it via **retract**(P_i, ϕ), removing any obligation to justify it.

Example 3 (Child custody dialogue). Alice and Lucy are talking about a divorce case, specifically whether it is in the child's best interest to live with her mother or with her father. They have the following dialogue.

Alice: It is in the ten-year-old child's best interest that she lives with her mother. (*assert*)
Lucy: Why? (*question*)
Alice: Because the child wants to live with her mother and the civil code states that the judge must take the child's opinion into account. (*justify*)
Lucy: A ten-year-old child does not know what she wants. (*challenge*)
Alice: Why? (*question*)
Lucy: Public opinion says that ten-year-old children do not know what they want. (*justify*)
Alice: Most ten-year-old children do know what they want. (*assert*)
Lucy: Why do you say that? (*question*)
Alice: Peter is a child psychologist, and Peter says that most ten-year-old children know what they want. (*justify*)

2.3 Argumentation as Inference

Argumentation as inference determines which conclusions can be drawn from an incomplete, inconsistent, or uncertain body of information. It defines a nonmonotonic notion of logical consequence through argument construction, attack, and evaluation, where arguments are structured as premises, conclusions, and inferences [37]. This approach employs formal methods such as nonmonotonic logic for commonsense reasoning, graph theory, computational logic, causal reasoning, and Bayesian reasoning. Its primary application is in knowledge representation and reasoning.

Most literature in formal argumentation focuses on argumentation as inference. Before Dung's 1995 paper, Pollock had already conducted extensive research on argument structure, defeasible reasons, the interplay between deductive and defeasible reasoning, argument strength, and different types of defeat [30–36]. He conceptualized reasoning as constructing arguments, where reasons serve as atomic links [31], and distinguished between defeasible (prima facie) and nondefeasible (conclusive) reasons [30]. While nondefeasible reasons logically entail their conclusions, defeasible reasons can be overridden by new information. He identified two types of defeaters: rebutting defeaters, which deny the conclusion, and undercutting defeaters, which challenge the link between reason and conclusion. To represent arguments, Pollock [31–33] introduced inference graphs, where nodes denote inference steps and arrows indicate defeasible inferences, deductive inferences, or defeat links [32].

Example 4 (Child custody in an inference graph). The dialogue between Alice and Lucy can also be illustrated in the format of Pollock's inference graph, as shown in Figure 3. Figure 3 illustrates two arguments rebutting the two opposite conclusions *"It is in the child's best interest that she lives with her father"*, and *"It is not in the child's best interest that she lives with her father"*. An undercutting argument, *"Public opinion is not reliable"*, defeats the argument *"Most ten-year-old children do not know what they want"*. In this figure, nondefeasible and defeasible inferences are visualized respectively with solid and dotted lines (without arrowheads). The arrows are defeat relations.

3 The Metamodel: Relations Between the BDI Models

A metamodel defines the structure and relationships between models, serving as a higher-level abstraction that standardizes their shared rules while allowing domain-specific variations. In software engineering, metamodels, e.g., Unified Modeling Language (UML) [29], enforce structural consistency by specifying syntax and constraints for models. In knowledge engineering, top-level ontologies, e.g., Basic Formal Ontology (BFO) [40], similarly ensure semantic consistency by providing foundational concepts and relationships for domain ontologies [42].

Table 1 in Section 2 might give the impression that the three approaches are distinct and that they have distinct application areas. We would like to

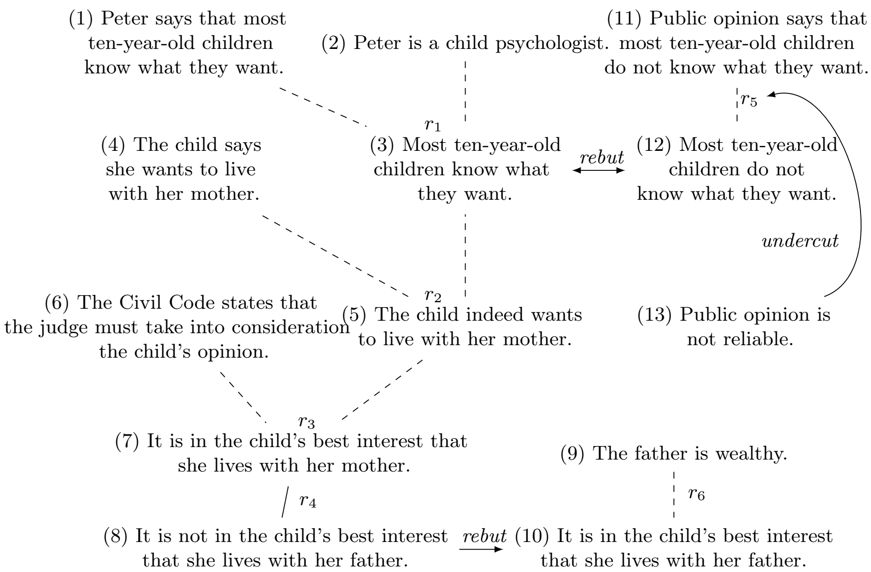

Fig. 3: The child custody case represented as a Pollock inference graph. The solid and dotted lines (without arrowheads) are nondefeasible and defeasible inferences respectively. The arrows are defeat relations.

point out that this is not the case. The approaches (or types) of argumentation are not mutually exclusive or even incompatible. You can switch from one to another if you want to look at the same problem or situation from different angles, highlight different aspects, or select a modeling approach that is more suitable for a particular purpose. This also means that complex application areas like the legal domain can make very good use of each approach. Indeed, legal reasoning often engages with each of the three models — argumentation as inference, dialogue, and balancing — across different contexts and legal roles. Judges and attorneys may rely on one form of argumentation more than another, depending on the nature of the case and their specific role in the legal process. For instance, inference is commonly used by judges, attorneys, actually any type of lawyer, when applying legal rules to facts or deriving conclusions from incomplete or inconsistent premises. Dialogue plays a central role in courtroom exchanges between opposing parties. The structure of a trial often resembles a dialogue: each party presents their arguments and responds to those of the other while the judge oversees the process to ensure it follows legal procedures. Balancing is typically the domain of judges as they weigh multiple factors, conflicting interests, or values, to determine the most appropriate outcome. This is particularly important in discretionary decision-making where the law, instead of trying to provide detailed rules, assigns special power to judges so that they can make decisions based on their own evaluations. In such cases, judges exercise their judicial discretion by carefully balancing competing considerations within the framework of legal principles to reach a fair and just decision. Hence, these different modes of reasoning can correspond to and interact with one another, creating a comprehensive tool set for legal reasoning and decision-making.

Some papers in formal argumentation attempt to translate these different models from one to the other, for instance, from Carneades to ASPIC+ [18]. As mentioned in Section 2, Carneades uses varying proof standards and burdens of proof, reflecting a balancing perspective. It is mapped into ASPIC+ by treating assumptions as necessary axioms and modeling exceptions as undercutters. Another example is between ASPIC+ and dialogue, that is, ASPIC+ can be seen as at a fixed moment with a single agent in a dialogue. For example, Modgil uses ASPIC+ to define argument structures dynamically throughout a dialogue [28], establishing correspondences between the dialectical status of dialogue moves and argument acceptability.

4 The A-BDI Metamodel and Dung style abstraction

From the perspective of the A-BDI metamodel, each model—balancing, dialogue, and inference can be formally related through Dung's abstract argumentation graph at a higher level. Dung's framework is extended with—such as argument strength, bipolar relations, and multi-agent—to accommodate the specific methods and concerns of each model. We plan to discuss the universality of attack in greater depth in a separate paper, but the central point here is that these extensions are essential to unify all three models within the A-BDI metamodel without reducing one to another.

From the diversity of natural language arguments, there is a corresponding diversity of formal representations. The modern stage of formal argumentation identifies the diversity of argumentation and reasoning approaches a common core: Dung's theory of abstract argumentation. This paradigm shift in formal argumentation shows, roughly, how many forms of reasoning can be characterised at an abstract level as an instance of graph reasoning. Formal argumentation has been used since the mid-nineties as a general framework to classify reasoning methods, besides non-monotonic logic and logic programming also, for example, instances of game theory and social choice.

While the central notion of Dung's theory is the acceptability or non-acceptability of arguments based on attack and defense, Dung shows that non-monotonic logic is a special form of argumentation (more details in Section 3.2). It can be visualized as the commutative diagram in Figure 4. There are two approaches to deriving conclusions from a knowledge base. The first is a direct approach where a given logic selects a set of rules with conclusions. The other is an indirect approach through argumentation, as shown in Figure 4 (2—4). Structured argumentation studies the process that adds the structure that turns collections of rules into arguments and assigns attack relations (2) among arguments. This gives us abstract argumentation frameworks — directed graphs where nodes represent arguments, and arrows represent attack relations. Then argumentation semantics (3) determine the acceptance status of arguments and their conclusions. To represent a given logic by structured argumentation, eventually the conclusions from both approaches must be the same.

$$\begin{tikzcd}
\text{knowledge base} \arrow[r, "(1)\ \textit{logic}"] \arrow[d, "(2)\ \textit{attack assignment}"'] & \text{conclusion extension} \\
\text{argumentation framework} \arrow[r, "(3)\ \textit{arg. semantics}"] & \text{accepted arguments} \arrow[u, "(4)\ \textit{conclusion extraction}"']
\end{tikzcd}$$

Fig. 4: Commutative diagram: two approaches to nonmonotonic inference: (1) logic systems; (2)–(4) argumentation systems. With appropriate choices on elements (2) and (3), one can obtain exactly the same conclusions as for the given logic (1).

Pre-existing ideas and methods, such as Pollock's argumentation-based inference, dialogue theories, and balancing, continue to persist and influence contemporary research. Rather than being rendered obsolete, these traditional theories are reinterpreted within the context of this new paradigm. On the one hand, Dung's theory provides a general framework that unifies diverse reasoning. On the other hand, it does not fully capture additional features important to dialogue, such as agents and strategies, or balancing, such as pros and cons with weights. Following this, there is a "reverse" diversity: various extended abstract argumentation frameworks and diverse semantics for (extended) abstract argumentation frameworks have been developed. For argumentation as dialogue, Abstract agent argumentation extends Dung's framework by associating arguments with agents [52]. Higher-order/meta-argumentation [10] introduces attacks on attack relations, reflecting real-life dialogues where arguments challenge other arguments and their relations. For instance, a lawyer may debate whether a suspect's argument effectively attacks another, or argue similarly about her own claims. For argumentation as balancing, examples are bipolar argumentation [11] that defines support and attack independently, weighted argumentation [8] that specifies a numeric value that indicates the relative strength of an attack, and value-based argumentation [4] defines values that are associated with an argument.

These examples illustrate how diverse extended argumentation frameworks build upon Dung's foundational concepts by introducing additional elements. We emphasize two points. The first is that the key reasoning ontology fundamentally relies on just two elements: arguments and attacks. Every utterance, be it a claim, an argument, or even an attack, can be modeled as an argument within these frameworks. Consequently, many of these extended frameworks can be *flattened* to basic ones, reinforcing the idea that attack graphs serve as a universal model of reasoning — much like how Turing machines serve as a universal model of computation. Second, while argumentation as inference has been extensively studied, the other two models have received relatively less attention.

5 Methodology

There is a gap between formal argumentation theories and applications of them. The main question is how to bridge these two. To address this, we need a "user guide" for argumentation that links theories to practical applications. However, creating such a user guide is challenging because of the diversity in natural, formal, and computational argumentation [54]. This diversity contrasts with the relative standardization found in fields like Answer Set Programming (ASP) and the Semantic Web. Understanding these differences is crucial to developing a "user guide" for argumentation.

First, we do not attempt to reduce one model to another. Instead, we assume that the three existing definitions or models exist for a reason, namely, they are developed for distinct concerns, and they represent distinct viewpoints on argumentation from stakeholders. To emphasize this point, we also call the models "conceptualizations". We do not exclude translation between these three models, but for our methodology, we believe it is too early to make commitments.

In Section 2, we take an example of child custody that seems to fit each conceptualization, and each conceptualization shows distinct aspects of the example. In this case, the judge must decide whether the child should stay with the father or the mother. This example is particularly illustrative because it involves inferential reasoning (e.g., drawing conclusions from legal and factual premises), structured dialogues among various parties (including parents, children, lawyers, and judges), and a final decision that cannot be strictly predetermined or automated. As the law emphasizes, the child's best interests must be balanced, making the decision process inherently complex and context-dependent.

The second step of the methodology we use in this paper is to compare the case study with other examples of reasoning in the literature. In Section 2, we focus on argumentation as balancing, as it is the least explored conceptualization within the COMMA community and the handbooks, despite extensive philosophical discussions. At the formal level, this means that we switch from qualitative approaches for inference and dialogue to quantitative approaches for balancing. Again, there is the problem of diversity, now the diversity of quantitative approaches, and the general problem of linking qualitative and quantitative approaches.

In the following, we present a preliminary step-by-step methodology for selecting a suitable model of formal argumentation. It is a structured process that considers the nature of the problem, the type of data and knowledge involved, and the stakeholders' objectives.

1. **Identify the problem context.** Determine the primary goal: Is the system intended to facilitate debates, persuasion, or negotiation? Is it meant to infer plausible conclusions from uncertain or conflicting information? Or is it balancing competing options and criteria?
2. **Define stakeholder requirements.** Elicit the needs and preferences of all parties involved. For instance, do stakeholders require a mechanism for real-time interaction, or do they need a computational approach to reason over information? This helps to narrow down which model aligns best with practical constraints (e.g., time, resources, domain expertise).
3. **Map requirements to conceptual features.** Based on the considerations in Step 1 and Step 2, compare the required features to the characteristics of each model:
   - *Dialogue-oriented features:* Turn-taking rules, strategic discourse planning, and protocol definitions for dispute resolution.
   - *Inference-oriented features:* Focus on constructing and evaluating arguments in a nonmonotonic or probabilistic setting, emphasizing the derivation of conclusions from incomplete or contradictory information.
   - *Balancing-oriented features:* Mechanisms for weighing pro and con arguments, evaluating multiple criteria, and integrating ethical or legal considerations.
4. **Assess formal and computational methods.** Evaluate whether the domain demands specific formal tools:
   - *Dialogue-based* approaches typically rely on protocol design, strategic game-theoretic analysis, and speech-act formalization.
   - *Inference-based* approaches often require nonmonotonic reasoning techniques, graph-based argumentation frameworks, or probabilistic logic.
   - *Balancing-based* approaches commonly employ methods from decision theory, ethics, and multi-criteria or case-based reasoning.

   Choose the model whose formal toolkit matches your technical and theoretical requirements.
5. **Evaluate Scalability and Practical Feasibility.** Each model may come with different computational overhead, data requirements, and integration constraints. Determine whether the targeted system must handle large-scale reasoning tasks or real-time interactions, and ensure that the chosen model can be implemented with available resources.
6. **Validate Through Pilot Studies or Prototyping.** If feasible, implement a small-scale prototype or conduct a pilot study to confirm that the chosen model meets the desired objectives. Gather feedback from end-users or experts to identify any alignment issues or missing features.

7. **Iterate and Refine.** Refine your selection by revisiting earlier steps in the light of new insights. Adjust the chosen conceptual framework or explore a hybrid approach if the system requirements cross over more than one model.

6 Hybrid AI: Motivating the Metamodel

AI broadly encompasses two major paradigms: sub-symbolic and symbolic AI. Sub-symbolic methods, notably LLMs, excel at pattern recognition and natural language processing by learning complex statistical patterns from extensive datasets. Symbolic approaches, in contrast, rely on formal knowledge representation and reasoning (KRR) to provide rigorous frameworks for inference, explanation, and rule-based compliance. Both paradigms are well-established and applied across various fields such as law, ethics, and finance, yet each possesses distinct strengths and limitations.

A key objective in human-centered AI is to integrate these two paradigms so that the resulting systems can leverage the flexibility of data-driven methods while benefiting from the interpretability and critical thinking mechanisms inherent in symbolic logic. Data-driven technologies often struggle with transparency and accountability, issues that are particularly pressing in domains where decisions have legal or ethical consequences. Argumentation-based methods can mitigate these issues by guiding AI systems—especially large language models—to generate more interpretable explanations and engage in structured dialogue. Through argumentation schemes, speech act theory, and dialogical protocols, AI systems can reason about agents' beliefs, preferences, and intentions in ways that promote empathy, contextual awareness, and critical questioning.

Recent research has begun to demonstrate how formal argumentation frameworks can be integrated with LLMs to meet requirements in automated decision-making. For example, abstract argumentation and discussion games have been combined with prompt engineering to ensure that AI-generated explanations remain grounded in formal constraints [25]. This enables crucial legal safeguards such as the right to explanation or the right to challenge, making automated decisions more transparent and open to revision. Likewise, other proposals that integrate argumentation into new-generation AI appear in Distributed Argumentation Technology [50], which incorporates argumentation reasoning within multi-agent systems using distributed ledger technology. An instantiation of such is the IHiBO (Intelligent Human-Input-Based Blockchain Oracle). By placing argumentation structures and interactions on a blockchain, IHiBO affords decentralized, traceable, and secure decision-making processes, directly addressing concerns such as legal compliance, trust, and auditability.

In light of these developments, the A-BDI metamodel provides a structured perspective on formal and computational argumentation, clarifying how different models—balancing, dialogue, and inference. Each of these views can be integrated with LLMs and other AI technologies, depending on the specific requirements of an application. This highlights the need for a methodology that guides when and where formal and computational argumentation should be adopted, ensuring that its use is well-suited to the domain and objectives at hand.

7 Related work

Methodologies from knowledge engineering, enterprise architecture, and formal logic offer valuable insights for the methodology of formal argumentation. For example, CommonKADS [48] emphasizes stakeholder analysis and knowledge modeling, which informs the identification of system users and the structuring of underlying knowledge in argumentation. LogiKEy [7], with its focus on logic-based knowledge engineering, highlights the role of formal verification, particularly in normative and legal reasoning. From an enterprise architecture perspective, ArchiMate [43], based on IEEE 1471-2000, demonstrates how diverse stakeholder concerns give rise to distinct viewpoints—an observation that aligns with the models of argumentation as inference, dialogue, or balancing. Similarly, in programming and decision theory, structured methodologies—ranging from test-driven development protocols to decision analysis frameworks [38, 22]—provide systematic ways to transition from abstract theories to concrete implementations, reinforcing the value of methodical approaches in designing and applying formal argumentation systems.

Among the three models, balancing has received the least attention, yet it raises challenges in bridging philosophical discussions and formal models. At the formal level, balancing shifts from qualitative approaches to quantitative ones raises the challenge of linking these approaches and addressing the diversity of quantitative methods. For instance, quantitative argumentation considers both the strength of an argument and the attack relation itself, using fixpoints and multiplication methods [3], whereas Tucker's model focuses purely on the attack and employs addition methods. Tucker's model seems to offer a more sophisticated treatment of binary relations, going beyond the fixed structure of Dung's framework. The challenge remains to explore how ranking and gradual semantics in argumentation can further develop context-sensitive weighting mechanisms.

The burning-building example, where saving lives (a collective value) conflicts with preserving one's own health (an individual concern), illustrates a broader issue central to argumentation. In social sciences, this tension is known as the micro-macro dichotomy [12]. Micro-level reasoning focuses on an agent's individual judgments and choices, while macro-level approaches consider interactions among multiple agents, often within ethical or legal frameworks. In multi-stakeholder settings like Jiminy's ethical governor [24], bridging these levels requires merging norms or perspectives into a unified system. The challenge is how to formalize the balance between individual and collective considerations using argumentation tools.

8 Summary and future work

In this paper, we have introduced the A-BDI metamodel, the first metamodel for formal and computational argumentation. It contains three models, which conceptualize argumentation as balancing, dialogue, and inference. Each model addresses distinct concerns and employs its own formal methods, yet they have been discussed separately by extending Dung's abstract argumentation. By leveraging extensions that include argument strength, bipolar relations, and multi-agent dynamics, the A-BDI metamodel goes beyond a simple classification, enabling deeper relations and connections among diverse argumentation approaches. In complex real-world applications like child custody decisions, multiple models, or even all three, may be required. This raises further questions about whether additional models should be incorporated and whether existing ones should be further refined, for instance, by distinguishing between different types of dialogues [49]. With the increasing interaction between argumentation and LLMs and Generative Artificial Intelligence (genAI), the need for a well-defined methodology of formal argumentation has become more pressing. This is not a topic of one paper, but a change of focus of the whole community.

A key conclusion of this paper is that, in the past, the emphasis has been on argumentation as inference, but in the future, the emphasis will be on argumentation as dialogue and argumentation as balancing. There are three reasons. First, the advancement of chatbots and LLMs moves the field's center of attention from formal and computational argumentation to human-like, interactive dialogue systems. There is renewed interest in systems like Fatio, prompted by these new technologies. Second, argumentation on the web continues to grow in importance [44], consistent with earlier predictions by Gordon [19]. Third, deeper engagement with philosophical and ethical aspects of argumentation leads to more interaction between reason-based models and argumentation-based models, as illustrated by recent work in ethics and related areas [47].

In future work, several directions can be pursued. A key challenge is the methodology used to bridge the gap between formal theories and their application. Drawing inspiration from established frameworks such as CommonKADS, ArchiMate, and LogiKEy, this methodology will include steps for identifying stakeholders, selecting appropriate conceptual models, and validating implementations iteratively, tailored specifically to argumentation. Additionally, existing methodologies from knowledge-based systems (KBS) and related fields need to be adapted to accommodate the integration of LLM-based agents. As LLMs become more prominent in AI applications, understanding how formal argumentation interacts with their reasoning and decision-making processes will be essential. Finally, as discussed in Section 4, a crucial aspect of future research is the universality of attack. The question is how this principle can help relate the three conceptual models—balancing, dialogue, and inference—without reducing one to another. Rather than forcing a translation between them, we take inspiration from architecture methodologies such as ArchiMate, where different perspectives are systematically related while maintaining their distinct roles.

References

1. Alcaraz, B., Knoks, A., Streit, D.: Estimating weights of reasons using metaheuristics: A hybrid approach to machine ethics. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. vol. 7, pp. 27–38 (2024)
2. Amgoud, L., Ben-Naim, J.: Ranking-based semantics for argumentation frameworks. In: International Conference on Scalable Uncertainty Management. pp. 134–147. Springer (2013)
3. Amgoud, L., Doder, D., Vesic, S.: Evaluation of argument strength in attack graphs: Foundations and semantics. Artificial Intelligence **302**, 103607 (2022)
4. Atkinson, K., Bench-Capon, T.: Value-based argumentation. Handbook of Formal Argumentation **2**, 397–441 (2021)
5. Austin, J.L.: How to do things with words. Harvard University Press (1975)
6. Baroni, P., Gabbay, D., Giacomin, M., van der Torre, L. (eds.): Handbook of Formal Argumentation, vol. 1. College Publications (2018)
7. Benzmüller, C., Parent, X., van der Torre, L.: Designing normative theories for ethical and legal reasoning: Logikey framework, methodology, and tool support. Artificial intelligence **287**, 103348 (2020)
8. Bistarelli, S., Santini, F., et al.: Weighted argumentation. Handbook of Formal Argumentation, Volume 2 (2021)
9. Black, E., Maudet, N., Parsons, S.: Argumentation-based dialogue. Handbook of Formal Argumentation, Volume 2 (2021)
10. Cayrol, C., Cohen, A., Lagasquie Schiex, M.C.: Higher-order interactions (bipolar or not) in abstract argumentation: a state of the art (2021)
11. Cayrol, C., Lagasquie-Schiex, M.C.: On the acceptability of arguments in bipolar argumentation frameworks. In: European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty. pp. 378–389. Springer (2005)
12. Coleman, J.S.: Micro foundations and macrosocial behavior. Angewandte Sozialforschung anc AIAS Informationen Wien **12**(1-2), 25–37 (1984)
13. Dancy, J.: Ethics without principles (2004)
14. Drai, D.: Reasons have no weight. The Philosophical Quarterly **68**(270), 60–76 (2018)
15. FIPA: Communicative act library specification. http://www. fipa. org/specs/-fipa00037 (2002)
16. Gabbay, D., Giacomin, M., Simari, G., Thimm, M. (eds.): Handbook of Formal Argumentation, vol. 2. College Publications (2021)
17. Georgeff, M., Pell, B., Pollack, M., Tambe, M., Wooldridge, M.: The belief-desire-intention model of agency. In: Intelligent Agents V: Agents Theories, Architectures, and Languages: 5th International Workshop, ATAL'98 Paris, France, July 4–7, 1998 Proceedings 5. pp. 1–10. Springer (1999)
18. van Gijzel, B., Prakken, H.: Relating carneades with abstract argumentation via the ASPIC+ framework for structured argumentation. Argument & Computation **3**(1), 21–47 (2012)
19. Gordon, T.F.: Towards requirements analysis for formal argumentation. Handbook of formal argumentation **1**, 145–156 (2018)
20. Gordon, T.F., Prakken, H., Walton, D.: The Carneades model of argument and burden of proof. Artificial Intelligence **171**(10-15), 875–896 (2007)
21. Hakim, F.Z.M., Indrayani, L.M., Amalia, R.M.: A dialogic analysis of compliment strategies employed by Replika chatbot. In: Third International conference of arts, language and culture (ICALC 2018). pp. 266–271. Atlantis Press (2019)

22. Keeney, R.L.: Decision analysis: an overview. Operations research **30**(5), 803–838 (1982)
23. Kissine, M.: Speech act classifications. Pragmatics of speech actions **173**, 202 (2013)
24. Liao, B., Pardo, P., Slavkovik, M., van der Torre, L.: The Jiminy advisor: Moral agreements among stakeholders based on norms and argumentation. Journal of Artificial Intelligence Research **77**, 737–792 (2023)
25. Liga, D., Markovich, R., Yu, L.: Addressing the right to explanation and the right to challenge through hybrid-ai: Symbolic constraints over large language models via prompt engineering (2025), submitted to ICAIL 2025
26. McBurney, P., Parsons, S.: Locutions for argumentation in agent interaction protocols. In: International Workshop on Agent Communication. pp. 209–225. Springer (2004)
27. McBurney, P., Parsons, S., et al.: Argument schemes and dialogue protocols: Doug walton's legacy in artificial intelligence. FLAP **8**(1), 263–290 (2021)
28. Modgil, S.: Towards a general framework for dialogues that accommodate reasoning about preferences. In: International Workshop on Theorie and Applications of Formal Argumentation. pp. 175–191. Springer (2017)
29. Pilone, D., Pitman, N.: UML 2.0 in a Nutshell. " O'Reilly Media, Inc." (2005)
30. Pollock, J.L.: Defeasible reasoning. Cognitive science **11**(4), 481–518 (1987)
31. Pollock, J.L.: How to reason defeasibly. Artificial Intelligence **57**(1), 1–42 (1992)
32. Pollock, J.L.: Justification and defeat. Artificial Intelligence **67**(2), 377–407 (1994)
33. Pollock, J.L.: Cognitive carpentry: A blueprint for how to build a person. Mit Press (1995)
34. Pollock, J.L.: Defeasible reasoning with variable degrees of justification. Artificial intelligence **133**(1-2), 233–282 (2001)
35. Pollock, J.L.: A recursive semantics for defeasible reasoning. In: Argumentation in artificial intelligence, pp. 173–197. Springer (2009)
36. Pollock, J.L.: Defeasible reasoning and degrees of justification. Argument and Computation **1**(1), 7–22 (2010)
37. Prakken, H.: Historical overview of formal argumentation. In: Handbook of formal argumentation, pp. 73–141. College Publications (2018)
38. Raiffa, H.: Decision Analysis: Introductory Lectures on Choices Under Uncertainty. Reading, MA (1968)
39. Searle, J.R.: Expression and meaning: Studies in the theory of speech acts. Cambridge University Press (1979)
40. Smith, B.: Bfo/medo: Basic formal ontology (2003), `https://basic-formal-ontology.org/`, online; accessed 16 February 2025
41. Streit, D.: From Metaethics to Machine Decisions: Formal Models of Normative Reasons and Their Application in Philosophy and Machine Ethics. Doctoral thesis, Université du Luxembourg (2024)
42. Sure, Y., Staab, S., Studer, R.: On-to-knowledge methodology (otkm). Handbook on ontologies pp. 117–132 (2004)
43. The Open Group: ArchiMate Forum. `https://www.opengroup.org/archimate-forum` (2025), accessed January 21, 2025
44. Thieyre, J., Beynier, A., Maudet, N., Vesic, S.: Reassessing the impact of reading behaviour in online debates under the lens of gradual semantics. In: Fifth International Workshop on Systems and Algorithms for Formal Argumentation (SAFA-24). vol. 3757, pp. 119–133 (2024)

45. Traum, D.R.: Speech acts for dialogue agents. In: Foundations of rational agency, pp. 169–201. Springer (1999)
46. Tucker, C.: The Weight of Reasons: A Framework for Ethics. Oxford University Press, New York (2025)
47. Tucker, C.: The weight of reasons: A framework for ethics (2024)
48. Waldner, J.B.: Principles of Computer-Integrated Manufacturing. John Wiley & Sons (September 1992)
49. Walton, D., Krabbe, E.C.: Commitment in dialogue: Basic concepts of interpersonal reasoning. State University of New York Press (1995)
50. Yu, L.: Distributed Argumentation Technology. Ph.D. thesis (2023)
51. Yu, L., Al Anaissy, C., Vesic, S., Li, X., van der Torre, L.: A principle-based analysis of bipolar argumentation semantics. In: European Conference on Logics in Artificial Intelligence. pp. 209–224. Springer (2023)
52. Yu, L., Chen, D., Qiao, L., Shen, Y., van der Torre, L.: A Principle-based Analysis of Abstract Agent Argumentation Semantics. In: Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning. pp. 629–639 (11 2021). https://doi.org/10.24963/kr.2021/60, `https://doi.org/10.24963/kr.2021/60`
53. Yu, L., Markovich, R., Van Der Torre, L.: Interpretations of support among arguments. In: Legal Knowledge and Information Systems, pp. 194–203. IOS Press (2020)
54. Yu, L., van der Torre, L., Markovich, R.: Thirteen challenges in formal and computational argumentation. In: Gabbay, D., Kern-Isberner, G., Simari, G.R., Thimm, M. (eds.) Handbook of Formal Argumentation, Volume 3, pp. 890–976. College Publications (2024)