



**HAL**  
open science

## Mining implicit arguments for reasoning: A survey

Ekaterina Sviridova, Elena Cabrio, Serena Villata

► **To cite this version:**

Ekaterina Sviridova, Elena Cabrio, Serena Villata. Mining implicit arguments for reasoning: A survey. *Argument and Computation*, 2025, <10.1177/19462174251344764>. <hal-05435118>

**HAL Id: hal-05435118**

**<https://hal.science/hal-05435118v1>**

Submitted on 30 Dec 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Mining implicit arguments for reasoning: A survey

Argument &amp; Computation

1–25

© The Author(s) 2025

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/19462174251344764

[journals.sagepub.com/home/arg](https://journals.sagepub.com/home/arg)

Ekaterina Sviridova<sup>1</sup> , Elena Cabrio<sup>1</sup>  and Serena Villata<sup>1</sup> 

## Abstract

Argumentation is the process of creating arguments for and against competing claims. Computational argumentation involves different ways of analyzing and reasoning upon arguments and their relations. More precisely, Argument Mining is the research field aiming at automatically identifying and classifying argument structures in text. The research field is mainly focussed on the extraction of explicit argument structures (i.e., claims and premises connected by support and attack relations). However, an even more challenging task consists in extracting implicit argument structures in text (e.g., enthymemes). These structures are particularly valuable to then address argument reasoning, e.g., on incomplete and uncertain information, to finally compute the set of acceptable arguments, i.e., argument justification and skepticism. In this paper, we present and compare current approaches and available datasets for the novel task of Implicit Argument Mining. Future work perspectives are discussed to pave the way to further studies in this direction.

## Keywords

Argument Mining, implicitness

Received: 12 July 2024; accepted: 8 May 2025

## 1 Introduction

Rooted in people's inherent ability and necessity to articulate opinions and thoughts, argumentation permeates everyday discourse, shaping discussions, justifying decisions, and fostering a sound exchange of ideas. At its core, argumentation involves constructing and articulating grounded reasoning through language.<sup>1</sup> It encompasses diverse ways in which individuals present claims, provide evidence, and engage in discourse to support their viewpoints.

In the last ten years, several works in Natural Language Processing (NLP) focussed on the computational modelling of human argumentation, exploring different tasks, such as Argument Mining,<sup>2,3</sup> and its recent developments for automatic Argument Assessment<sup>4</sup> and Generation.<sup>5</sup> More precisely, Argument Mining (AM) stands out as the automatic identification and classification of explicit argument components and structures embedded in natural language text.<sup>3,6</sup> Recent approaches in AM delve deeper into the comprehension of argument-based reasoning processes and the required knowledge, going beyond the explicit content to analyze hidden or implicit meanings.<sup>7</sup>

While in general humans do not encounter significant difficulties in understanding an (even implicit) argument, this task is hard for computational approaches which often lack commonsense and background knowledge to reason on. Implicit inference refers to understated arguments, hidden assumptions and relations, that extend beyond the explicit context of an argument. To understand implicitness, computational models of argumentation consider different steps, among which the exploration of argumentation models to unveil argumentation structure, the study of enthymemes- i.e., incomplete arguments, where some components are left implicit<sup>8</sup> – to be able to identify and reconstruct this incompleteness, and the exploration of possible ways to restore implicit argument components.

<sup>1</sup>Université Côte d'Azur, CNRS, Inria, I3S, France

### Corresponding author:

Ekaterina Sviridova, Université Côte d'Azur, CNRS, Inria, I3S, France.

Email: [sviridova@i3s.unice.fr](mailto:sviridova@i3s.unice.fr)

The goal of this survey paper is to provide an overview of the existing approaches in the literature that explore the topic of implicitness in natural language argumentation, focussing in particular on the challenges that should be addressed by current methods to gain a complete understanding of an argument structure. Specifically, we begin by outlining our methodology for gathering a collection of research works focussed on enthymeme analysis (Section 2). Then, we provide a high-level roadmap of enthymeme analysis (Section 3), giving a clear understanding of the workflow before diving into the details of models, components, and methodologies. After that, we present the main argumentation models (Section 4) and highlight those used in implicitness studies in natural language argumentation. In particular, we emphasize the significant role played by argumentative schemes,<sup>9</sup> i.e., informal logic argument-based reasoning schemes, utilized in modelling implicit inferences within argumentation. Subsequently, we investigate argument components that may be left implicit (Section 5), playing, however, a key role for argument comprehension. After establishing the taxonomy of argument components that may be missing, in Section 6 we investigate how a natural language argument can be defined implicit and what makes it explicit (i.e., what kind of knowledge is required for argument explicitation), to obtain a full and sound understanding of the argument.<sup>10</sup> Also, we analyze computational methods for enthymeme analysis (Section 7) and we provide an exhaustive list of existing datasets for the implicit argument mining task (Section 8). We conclude the paper with the discussion of some new future research lines that can stimulate argument research communities to pursue new directions towards the achievement of fully automated argument mining and reasoning frameworks (Section 9), and we outline main points of the paper in the conclusion (Section 10).

## 2 Methodology

This section briefly discusses the methodology for collecting literature for the survey. Currently, resources on enthymeme studies that integrate computational approaches and argumentation theories are scarce, as implicitness remains a challenging phenomenon for automatic identification and modelling. Therefore, our primary goal was to include as many existing and relevant works on the topic as possible.

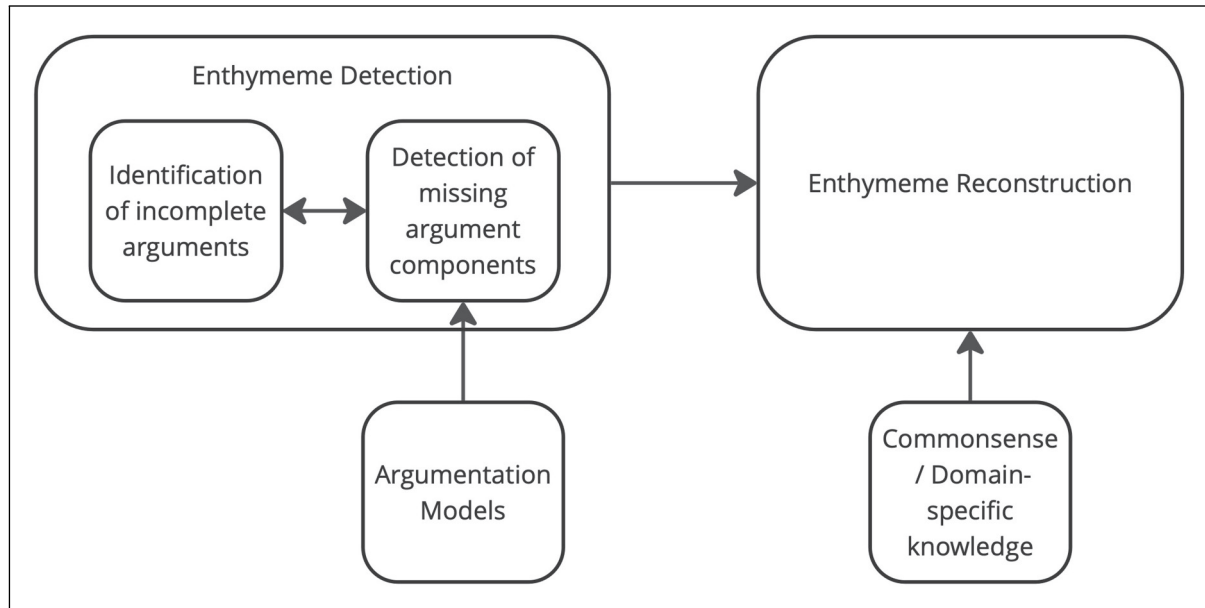
The search for relevant articles was conducted in three stages: (i) identifying key authors specializing in NLP, particularly in Argument Mining, implicitness in argumentation, hidden meanings, and figurative language, such as Lawrence, Reed, Walton, Boltužić, Šnajder, Stede etc.; (ii) searching by keywords relevant to the topic, including implicit argumentation, enthymeme, implicit premise, implicit warrant, implicit conclusion, enthymeme detection, and enthymeme reconstruction; (iii) using related papers found so far as benchmarks to discover additional relevant articles.

For the first stage, we referred to the programme committee members of the most recent workshop on AM (<https://argmining-org.github.io/2024/index.html#committee>) to retrieve author names and we conducted searches for their publications using platforms such as Google Scholar, ACL Anthology, and DBLP to find relevant works. At the second stage, we employed the same platforms, focussing on keyword searches to identify and evaluate titles, then, we reviewed the papers to select the most relevant ones. For the final stage, we utilized the Connected Papers resource (<https://www.connectedpapers.com>) which generates a graph of related works based on a selected paper or keyword. This graph visually represents the closeness between articles, highlights the most recent and most cited works, and provides metadata, such as authors, publication year, thus, it allows to select related and relevant articles.

The search was not restricted by time frame, as implicitness in argumentation is a relatively recent field within AM, with a limited number of publications. Similarly, no constraints were imposed on the type of publication. Our strategy aimed to include the majority of existing works on enthymemes and related topics, prioritizing their relevance to our research.

## 3 Enthymeme analysis pipeline

In this section we provide a high-level vision of the process of enthymeme exploration. To analyze and understand enthymemes, it is essential to follow key steps that allow to identify implicitness, to clarify reasons of omissions and to restore reasoning behind an incomplete argument. An overview of this process is represented in the pipeline (Figure 1). The process consists of two main steps or tasks: Enthymeme detection and enthymeme reconstruction.<sup>11,12</sup> The goal of the first task is to identify incomplete arguments and define missing components in them. Given an argument with (a) missing component(s), the goal of the second task is then to reconstruct the component(s) that fill the gap. In general, enthymeme detection is perceived as a single-step task, where identifying missing argument components automatically determines the argument as an enthymeme.<sup>10,13</sup> To effectively detect missing argument components, we need to apply a specific argumentation model to this step. However, we assume that identifying an argument as incomplete can be a distinct preliminary step, preceding the detection of its implicit components. This initial step involves evaluating argument quality, naturalness, and argumentativeness to determine incompleteness before analyzing the argumentative structure.<sup>12</sup> Such approach allows for a two-step process, (i) identifying incomplete arguments and (ii) detecting implicit argument



**Figure 1.** Enthymeme analysis pipeline.

components. While this flow is logical, it is still equally reasonable to detect missing components directly, since it allows to classify an argument as an enthymeme. At the same time, the second block of the pipeline, enthymeme reconstruction, nearly always requires extralinguistic information, such as commonsense knowledge<sup>14</sup> or domain-specific knowledge, as humans share background information for their reasoning. It is significant to note that some works address only the first module of the pipeline,<sup>11,15</sup> because enthymeme detection is a complex and versatile process in itself, while other works tackle only the second step as they start reconstruction given an enthymeme.<sup>14,16</sup> Addressing the complete pipeline, including both modules, is challenging, but some progress in this direction has also been made.<sup>12</sup> In the following, we will explore all the blocks of the presented pipeline, starting from the introduction of argumentation models employed for the other steps of enthymeme analysis.

## 4 Argumentation models

As it has been mentioned in various works,<sup>3,10,17</sup> understanding the argument structure first and foremost aids in enhancing comprehension and analysis of the underlying message in a natural language text. For example, decomposing argumentative units into distinct components allows us to attribute meanings to each component and trace the relations between them.<sup>18,19</sup> Therefore, this decomposition enables a deeper grasp of the interlocutor's reasoning that is nearly always hidden behind the text surface. As well as that, detection of argument components lets us better characterize arguments, for instance, to differentiate strong arguments from weak ones or complete arguments from incomplete ones. Moreover, argumentation schemes help us learn to construct various types of arguments in a proper way. Besides, employing argumentation schemes provides an undeniable advantage for argument mining. Unlike humans, computational models struggle to perceive the underlying reasoning in an argumentative text. By instructing them with prior knowledge of argument components and their relations, we address the challenge of automatically extracting arguments from free text,<sup>3</sup> thus, we take a step towards models that are able to understand human argumentation.

In the last years, a variety of AM approaches have been proposed, relying on different argumentation schemes and models.<sup>3</sup> Table 1 presents an extensive compilation of prominent Argumentation Models. It showcases contemporary argumentative approaches, beginning with those predominantly used in works exploring implicit argument components. The table provides insights into argument schemes and components, while strengths and drawbacks of each approach are discussed further in this chapter.

We will now explore specific details of the compilation. First, research on implicit inferences in argumentation mostly references the foundational models of Walton,<sup>8</sup> Toulmin,<sup>20</sup> and the argument structure approach of Freeman.<sup>21,22</sup> Studies of implicitness in argumentation typically focus on the components and relations within arguments, aiming to identify omitted argumentative discourse units<sup>18</sup> that need reconstruction. Walton, Toulmin, and Freeman's models are particularly valuable here, as they all centre on a core structure of argumentation – a conclusion (claim) supported or attacked by

Table 1. Argumentation models.

Argumentation models	Argument schemes	Argument components
<p><b>The Walton model</b><sup>8</sup> Used in Hulpus et al.,<sup>10</sup> Rajendran et al.,<sup>11</sup> Chakrabarty et al.,<sup>14</sup> Alshomary et al.,<sup>23</sup></p> <p><b>The Toulmin model of argumentation</b><sup>20</sup> Used in Hulpus et al.,<sup>10</sup> Habernal et al.,<sup>15</sup> Alshomary et al.,<sup>23</sup> Singh et al.,<sup>24</sup> Saadat-Yazdi et al.,<sup>33</sup></p> <p><b>Freeman's argument structure approach</b><sup>21,22,28</sup> Used in Becker et al.,<sup>16</sup> Alshomary et al.,<sup>23</sup> Hiday et al.,<sup>25</sup> Becker et al.,<sup>34</sup></p> <p>The New Rhetoric<sup>30</sup></p>	<p>Argument from position to know; argument from expert opinion; argument from witness testimony; argument from popular opinion; argument from popular practice; argument from example; argument from analogy; practical reasoning from analogy; argument from composition; argument from division; argument from opposition, etc.</p> <p>Reasoning from analogy; reasoning from generalisation; reasoning from sign; reasoning from cause; reasoning from authority; reasoning from dilemma; reasoning from classification; reasoning from opposites; reasoning from degree</p> <p>Defeasible <i>a priori</i> argument; defeasible <i>a posteriori</i> argument; conclusive <i>a priori</i> argument; conclusive <i>a posteriori</i> argument</p> <p><b>Arguments from association:</b> Quasi-logical arguments; relations establishing the structure of reality (establishment through particular case, reasoning by analogy); arguments based on the structure of reality (sequential relations, relations of coexistence, double hierarchy argument, differences of degree and order); <b>Arguments from dissociation</b></p> <p><b>Argument schemes using rules:</b> classification schemes, comparison schemes, opposition schemes, causal schemes; <b>Argument schemes establishing rules:</b> inductive argumentation from example; <b>Argument schemes using and establishing a rule:</b> illustrative argumentation from example, argumentation from analogy, argumentation from authority</p> <p>Symptomatic argumentation; argumentation based on similarities; instrumental argumentation</p> <p>Cause to effect; effect to cause; sign; sample to population; parallel case; analogy; population to sample; authority; ends-means</p> <p><b>Deductive argument schemes:</b> elementary deductive argument schemes, analytical arguments; <b>Probabilistic argument schemes:</b> pure probabilistic argument schemes (statistics, signs), impure probabilistic argument schemes (best explanation); <b>Practical argument schemes:</b> pure practical argument for pure evaluations, impure practical argument schemes (for justification of actions; justification of instruments), etc.</p>	<p>Conclusion, premise; major premise, etc.</p> <p>Claim, data, warrant, backing, qualifier, rebuttal</p> <p>Claim, premise, modality, rebuttal, counter-rebuttal</p> <p>Conclusion, premise; real objects of agreement / preferable objects of agreement</p> <p>Conclusion, premise</p> <p>Standpoint (claim), premise</p> <p>Conclusion, premise, warrant</p> <p>Thesis (claim), premise</p>
<p>Kienpointner's 'Alltagslogik'<sup>9</sup></p> <p>Pragma-Dialectics<sup>32</sup></p> <p>Grennan's Argument Typology<sup>9</sup></p> <p>Lumer and Dove's Argument Classification<sup>31</sup></p>		

evidence (premises) – though each model circumstantiates these components in a distinct way. To illustrate the usability of these three models in enthymeme detection and reconstruction, we will explore some examples (full set of papers that refer to one or another argumentation model is presented in the first column of the Table 1). The Walton model, for instance, is addressed by Rajendran et al.<sup>11</sup> and Chakrabarty et al.<sup>14</sup> for the detection and reconstruction of implicit premises, while Alshomary et al.<sup>23</sup> exploit it for detection and further generation of implicit claims. Several authors<sup>15,24</sup> use the model of Toulmin to explore implicit warrants. In this context, the Toulmin Model of argumentation serves the intended purpose due to its granularity. In the meantime, Becker et al.<sup>16</sup> resort to the argument structure approach of Freeman so as to annotate newly generated arguments with argument components and relations according to Freeman’s approach. Hidey et al.<sup>25</sup> reference Freeman’s approach as they apply semantic types of claims proposed in Freeman’s taxonomy<sup>26</sup> to their dataset. All in all, the research centred around implicitness in argumentation (i) utilizes only three major argumentation models, those of Walton, Toulmin and Freeman; (ii) exploits only argument components and relations, keeping aside argument schemes.

Although recent studies on enthymemes primarily focus on argument components of prominent argumentation models, argument schemes should not to be overlooked in this domain. They offer a pathway to easier detection of implicit components due to the fact that they classify arguments by their types and contexts and recognize their complete structure. Additionally, argument schemes may facilitate automatic reconstruction of incomplete arguments as they propose appropriate reasoning direction according to the specific type.<sup>8</sup> Therefore, we will discuss argument schemes from the Table 1 and we will elaborate on argument components characteristic of argumentation models where needed.

**The Walton model,**<sup>8</sup> first presented in 1995, encompasses 65 argument schemes that represent various types of everyday arguments as well as argument schemes tailored to specific contexts, such as legal or scientific (our compilation includes only a subset of these schemes due to space constraints). The model presents schemes for deductive, inductive and presumptive reasoning, making it an immense educational resource and a great foundation for the advancement of argument mining systems. The key contribution of this model is its emphasis on presumptive reasoning. Presumptive reasoning involves making plausible but defeasible inferences, meaning they can be overturned if new evidence arises. This reflects how people often reason in everyday situations, legal contexts, and scientific discussions, therefore, the model represents real-world argumentation. Regardless of the fact that the model describes the patterns of typical arguments, the vast number of schemes and their variations make the model difficult to utilize in practice, especially if there is a need to formalize the model for AI applications. Moreover, only a subset of schemes (8–10 schemes) find their real use, as many are too specific for everyday discourse or poorly understood. Argument components in this model vary in accordance with argument schemes, for instance, an argument may have a major premise apart from common conclusion and premise. Despite this, a conclusion and a premise are always present.

**The Toulmin model of argumentation,**<sup>20</sup> introduced in 1984, proposes 9 argument schemes centred on the function of warrants. Being a component of an argumentative structure, the warrant serves as the link between data (premises) and a claim. This perspective shifts the understanding of arguments, emphasizing the warrant as the key element in guiding the conclusion, rather than logic being compulsory. According to Toulmin, the warrant is what enables the conclusion to be drawn. However, the function of warrants is not a single characteristic of this classification, as it also employs other criteria. Some schemes, for instance, are based on types of reasoning, such as generalization, sign, or analogy; others are defined by rules of inference, like dilemma or opposites; and the rest of schemes are categorized by the content of the argument, such as authority, classification, cause, or degree. Apart from data, claims and warrants, the argumentative structure also includes backing, qualifier and rebuttal. Backing is an element that helps to verify the validity of the warrant by proposing new information or even new arguments in favour of the warrant. Qualifiers in their turn allow us to evaluate the strength of support provided by the warrant towards the claim. If we can undoubtedly accept the claim due to the warrant given the data, we classify the claim as ‘definitely’ true. However, if we are uncertain about the quality of the warrant, we classify the claim as ‘possibly’ true. Rebuttals indicate the situations where the warrant does not apply, therefore, they may invalidate the claim. Due to such granularity the model allows to follow the reasoning of argumentation step by step detecting all the stages of decision making process. However, Toulmin’s approach has faced significant criticism, as the distinction between premises and warrants is often vague, making warrants difficult to identify, therefore, warrants do not clarify our understanding of argumentation.<sup>27</sup> As well as that, the presence of warrants does not always seem realistic, since humans do not need to communicate explicitly each detail of their reasoning.

**Freeman’s argument structure approach,**<sup>21</sup> presented in 1991 and revised in 2011, classifies arguments according to their conclusive power and the way the warrants are backed.<sup>28</sup> An argument is conclusive if its warrants are conclusive, meaning unconditionally valid, valid without any exceptions. Warrants that allow exceptions are defeasible, thus making their argument defeasible. At the same time, both types of warrants may be backed *a priori* and *a posteriori*. Backing *a priori* refers to utterances that are true either by virtue of their logical form or by semantic meaning, while backing *a*

posteriori means that utterances require sense experience. Given all these divisions, there are four possible types of arguments: defeasible a priori, defeasible a posteriori, conclusive a priori and conclusive a posteriori. This argument typology could be particularly useful for studying enthymemes. On the one hand, real-world arguments are often enthymematic because they follow pragmatic principles, such as Grice's Maxim of Quantity,<sup>29</sup> which prioritizes concise communication. However, authors might deliberately omit certain statements in an argument to conceal their weak or defeasible nature.<sup>13</sup> Particularly in these cases determining whether an argument is defeasible or conclusive can aid in detecting enthymemes and can contribute to their subsequent classification. As for argumentative structure, Freeman's approach preserves two major argument components, that are premise and claim, with all other components being optional. Freeman's first novelty lies in replacing Toulmin's Qualifiers with Modalities. This shift from qualifiers to modalities reflects the transition from premise to conclusion, indicating how strongly the premise supports the conclusion. In contrast, Toulmin's model treats qualifiers rather as a property of the conclusion (the claim (conclusion) is classified as 'definitely' true, 'possibly' true etc.), which is not totally accurate. Freeman retains the concept of rebuttals that we saw in Toulmin's model, defining also two types of rebuttals: Rebutting defeaters and undercutting defeaters. A rebutting defeater (R) is an element detrimentally related to the conclusion, while undercutting defeater (U) is an element questioning the reliability of entailment between the premise and the conclusion. An introduction of a counter-rebuttal in its turn makes an argument more realistic providing arguers with a means to defend their claims when faced with a rebuttal. It seems that Toulmin has already considered a possibility to have an answer to a rebuttal, but never presented a counter-rebuttal as an independent component.<sup>22</sup> Counter-rebuttal is also further elaborated: It may directly indicate that the rebuttal is false or it may only undercut the rebuttal. Opposite to Toulmin, Freeman does not define warrant separately as a component given the criticism towards warrants, the difficulty to distinguish between premise and warrant and an assertion that warrants are artificial. Regardless of the fact that Freeman's approach looks better elaborated, as the author takes into consideration drawbacks of previous models, it still lacks significant details. The main one is that rebuttals and counter-rebuttals have to be further revised as they do not include all their possible types.<sup>22</sup>

**The new Rhetoric model,**<sup>30</sup> first introduced in 1958, translated to English in 1969, presents argument schemes that rely on the mechanisms of association and dissociation that allow to pass from premises to a conclusion. Arguments from Association are based on the association between concepts (e.g., the association between an act and a person who established the act) and include three classes: Quasi-Logical Arguments, Relations Establishing the Structure of Reality and Arguments based on the Structure of the Reality. Quasi-Logical type in its turn comprises techniques of Contradiction and Incompatibility, techniques of Identity and Definition, Arguments of Reciprocity, Arguments of Transitivity etc. Relations Establishing the Structure of Reality schemes are further subdivided into Establishment through Particular Case and Reasoning by Analogy, while Arguments based on the Structure of the Reality schemes are subdivided into Sequential Relations, Relations of Coexistence, Double Hierarchy Argument and Differences of Degree and Order. Arguments from Dissociation make a separate class that is based on the dissociation of concepts (e.g., the appearance of an object is incompatible with the reality, it is illusive). This system of schemes encompasses most possible associative and dissociative arguments. However, being based on various criteria (e.g., association, structure of reality), it lacks a unified framework, as the relationship between these criteria is not clearly defined. As for the argument structure, the New Rhetoric model identifies premise and conclusion as its core argument components. What sets this model apart from the models discussed so far is its interpretation of premises. The model emphasizes that premises are chosen specifically to reach the audience's agreement, making the agreement the major criterion for selecting appropriate premises. Premises are further divided into two classes based on the object of agreement: Real and preferable. This classification reflects again the fundamental aim of premises – to guide the audience toward agreement. The first class – real – comprises facts, truths and presumptions, while the second class – preferable – includes values, hierarchies and lines of argument relating to the preferable. Such modelling of argumentative structure represents the reality of our reasoning by capturing both objective elements, like facts and truths, and subjective elements, like values and hierarchies, therefore, providing frameworks for reaching agreement either through logical appeals or resonating with the audience's beliefs and preferences when necessary.

**Kienpointner's 'Alltagslogik',**<sup>9</sup> introduced in 1992, defines 21 main argument schemes divided into three classes: argument schemes using a rule, argument schemes establishing a rule and argument schemes both using and establishing a rule. The classes are built upon the type of inference: Each class depends on the nature of the reasoning process, whether the inference presumes an established logical rule, works to create one, or operates in both capacities. In every class the approach distinguishes schemes according to the epistemic nature of the premises, meaning that it accounts for the truthfulness and reliability of premises or at least the mere possibility of premises, whereby each scheme in the model has both, true and fictive (possible) variations. Furthermore, Kienpointner points out the significance of the dialectical and pragmatic function of the conclusion that can play a role in establishing argument schemes: Every scheme in the classification can support or oppose a certain assumption (dialectical function) and can have descriptive or normative conclusion (pragmatic function). One major advantage of Kienpointner's model is that all the schemes are either deductively valid or potentially

deductively valid.<sup>31</sup> This means the premises logically entail the conclusion, making it impossible for the premises to be true while the conclusion is false. By this criterion the model addresses a key argumentative issue: Many well-structured and persuasive arguments often fail to meet the standard of deductive validity. However, the same advantage appears to be a drawback in some cases. For instance, Kienpointner aims to resolve deductive invalidity of arguments either by strengthening premises or by weakening claims, which leads to creation of false premises and useless conclusions.<sup>31</sup> Another major problem of Kienpointner's model is its empiric approach of collecting argument schemes. The model lacks deep understanding of theoretical foundation of argumentation as it is based on the analysis of the vast corpora of everyday arguments. Together with that, it does not include various contexts apart from everyday argumentation. As for argument components, Kienpointner's 'Alltagslogik' keeps premise and conclusion not introducing new proposals.

**Pragma-Dialectics approach,**<sup>32</sup> fully elaborated in 1992, focuses on argumentation designed to resolve differences of opinion through logical reasoning, valid arguments, and rational explanations. This model does not accept persuasive techniques that are based on emotional appeal or fallacious reasoning. Furthermore, it aims to evaluate arguments not just for their logical validity but also for their pragmatic effectiveness in achieving resolution. Pragma-Dialectics proposes the classification of argument schemes grounded in two main criteria: causality and analogy, which form the basis for three general classes. The first class, symptomatic argumentation, presupposes that 'something is symptomatic of something else' in an argument, therefore a premise in this argument is a sign or symptom of what is stated in the conclusion. Symptomatic argumentation unifies the category of causal argumentation and analogical pattern. The second class, argumentation based on similarities, is based on analogical relations between a premise and its conclusion. The third class, instrumental argumentation, introduces the relations of causality between a premise and its conclusion. Other subclasses of argument schemes fall into one of these three classes. A major advantage of the logical approach is its wide range of applications: The model can be used to analyze and improve real-world practical argumentation, particularly in contexts like political debates, legal reasoning, and other domains where fallacious evidence is unacceptable. At the same time, this approach may be complex to fully understand and may require deep research of its theoretical and meta-theoretical principles. Analyzing argument components we notice that dialectic grounds of the model influence the components. Claim is referred to as 'standpoint' because this term better represents a position that one party adopts in a discussion or debate, and this position may be challenged or defended. Such perspective also highlights the role of conclusions in the interaction between participants, while claims seem to be static components, standpoints possess a certain flexibility when rationally questioned. Apart from claim in the form of standpoint, Pragma-Dialectics also defines premise as the second essential argument component.

**Grennan's argument typology,**<sup>9</sup> elaborated in 1997, is grounded in inductive validity of an argument, proposing that conclusions are drawn based on the likelihood or probability derived from premises. Unlike deductive validity (where the conclusion necessarily follows from the premises), for instance, in Kienpointner's 'Alltagslogik', inductive validity deals with arguments where the conclusion is supported by the premises with a degree of probability. In this regard, Grennan defines 9 argument schemes that derive from 9 warrant types. This classification is based on warrants because in inductive argumentation warrants are used to justify why the premises make the conclusion plausible, reflecting inductive validity. Moreover, all types of warrants present logical justification of inference as Grennan's model does not reach out to any other (e.g., emotional) kind of reasoning. We will now take a closer look at each argument scheme within the typology. Cause to Effect scheme represents the relations between a premise and a claim, where the occurrence described in the premise produces the phenomenon stated in the claim. Effect to Cause means that the reason of a premise is its claim. Sign is about symptomatic relations between a premise and a claim: the phenomenon stated in the premise is symptomatic of the one in the claim. Sample to Population is based on the assumption that the truth for a sample of a multitude is true for any other sample of this multitude, while Population to Sample represents the opposite: the truth for some known elements of a multitude is also true for the other exact element of this multitude. Parallel Case shows the truths that are shared by several elements in parallel. Analogy explains analogical relations of elements within argument components: B1 is to B2 in a claim as A1 to A2 in a premise. Authority points out a reliable source for a claim. Ends-Means states that the action of a claim achieves its end state in a premise. This argumentation model, which builds upon the typology of warrants, is further enriched by the classification of conclusions, making it highly detailed.<sup>9</sup> As already mentioned, argument components of Grennan's Typology include premise, conclusion and warrant.

**Lumer and Dove's argument classification,**<sup>31</sup> presented in 2011, suggests an alternative to previous approaches, combining logical reasoning and pragmatic purpose as its foundational criteria for argument schemes. The system categorizes argument schemes into three major classes: Deductive, Probabilistic, and Practical schemes, each further divided into detailed subclasses. The first class, Deductive argument schemes, includes elementary deductive arguments and analytical arguments. The second class, Probabilistic schemes, is split into pure probabilistic arguments, like those based on statistics or signs, and impure probabilistic arguments, such as those employing the best explanation. The third class, Practical schemes, characterized by schemes' pragmatic aim to recommend actions, encompasses a wide variety of arguments,

including those for evaluations, welfare-ethical judgments, and justification of actions or instruments. This approach is advantageous due to its ability to address the complexity of real-world argumentation, as each subclass is defined according to logic and pragmatic classifications. However, the reliance on mixed criteria can also be seen as a limitation, as it introduces heterogeneity that might complicate creating a unified system. It is significant to mention that Lumer and Dove retain ‘premise’ as an argument component, but adopt the term ‘thesis’ instead of ‘claim’. This choice reflects the broader and more nuanced goals of argumentation. While a claim or conclusion typically serves as the endpoint of an argument, a thesis represents a position that can be justified, evaluated, defended, or even revised. Furthermore, authors propose that argumentative validity, situational adequacy, rationality of the addressee, convincability, rational acceptance of the reasons and intelligibility of the argument are necessary conditions for a perfect argument.

All in all, it is important to highlight that argument components from argumentation models play an essential role in unveiling the structure of an argument for argument mining and analysis, while being particularly useful for enthymeme studies. Since the first step in enthymeme detection is to determine which elements in an argument are missing, these models provide the foundation for identifying and classifying implicit components.

## 5 Implicit argument components

In this section, we define the argument components that may be left implicit and describe the associated challenges. The idea that premises are most often the implicit component in arguments can be traced back to early works on enthymemes, such as the paper of Rajendran et al.,<sup>11</sup> which, influenced by Walton’s research,<sup>8</sup> defined enthymemes primarily as arguments with implicit premises. This early perspective introduced a bias in the literature, overlooking the fact that enthymemes may also define arguments with implicit claims, warrants, or other components. While it is true that premises are frequently left implicit,<sup>14</sup> as humans can infer and fill reasoning gaps using their background knowledge, the analysis of various types and genres of argumentative texts shows that other components can be left implicit to a similar extent. As such, some research works explore implicit warrants,<sup>15,24</sup> while others state that conclusions that are self-evident may be left implicit.<sup>23</sup>

Table 2 illustrates argument components that, although may be left implicit, are nevertheless essential for a comprehensive understanding of an argument. The identification of these implicit argument components is a challenging yet crucial task for automated systems. The table also outlines the NLP tasks that have been defined to tackle such challenge, as well as the benchmarks created to evaluate these tasks. Additionally, it includes details about the manual or automatic annotations proposed by the authors of the tasks. In the following, we describe each task, discuss the proposed annotations, and examine the similarities and differences in the approaches.

### 5.1 Implicit premise

To begin with, Fill-in-the-Gap Task or Claim Matching Task<sup>35</sup> is meant to develop methods able to fill the gap between two claims, specifically, a user’s claim on a certain topic and a major claim unifying all users’ claims on this topic. This task follows the assumption that arguments on a debatable topic discussed in social media share one main claim. However, in practice, real users’ claims do not correspond to this main claim. Consider an example:<sup>35</sup>

(1) *Main Claim: Legalization of marijuana causes crime.*

*User Claim: It would be loads of empathy and joy for about 6 hours, then irrational, stimulant-induced paranoia. If we can expect the former to bring about peace on Earth, the latter would surely bring about WWII.*

The topic of this debate is *Marijuana* with a negative stance label (against marijuana). The main claim of the debate is clear; however, the user’s claim presented in the example is markedly different: while the main claim represents a general statement that calls things by their right names, the user’s claim describes a person’s specific condition after using drugs, defining neither an actor nor a reason of this state. The user’s claim remains figurative and understanding its link to the major claim requires a lot of background knowledge. Therefore, filling this gap between the main and the user’s claim is necessary for argument understanding and future reconstruction of the reasoning, especially for automatic argumentative analysis. The claim matching task<sup>35</sup> involves manually reconstructing fill-in-the-gap premises to link new users’ claims to existing ones. While this approach aids the matching process, the task presents several complexities that are currently difficult to address automatically. First, the size of the gap between the main claim and the user’s claim can vary significantly, making it challenging for language models (LMs) to handle larger gaps effectively. Second, while humans can easily reconstruct implicit premises between two claims, they often use entirely different premises for the same set of claims, posing a challenge for models to learn consistent patterns of reconstruction. Finally, the task itself is

**Table 2.** Implicit components of an argument.

Implicit AC	Tasks	Data	Annotations	M / A
Premise between claims Premise	Fill-in-the-gap (claim matching) task <sup>35</sup> Enthymeme detection <sup>11,12</sup>	Online Debate Forum <sup>35</sup> ArguAna corpus <sup>36</sup> ICLEv3 corpus <sup>12,37</sup> ART, <sup>38</sup> Room for Debate (NYT), <sup>15</sup> Online Debate Forum <sup>35</sup> Microtext Corpus <sup>39</sup>	Main claims, fill-in-the-gap premises Explicit / neutral / implicit opinions Enthymematic arguments Premises	M M A A
Warrant	Enthymeme reconstruction <sup>12,14</sup> Argument reasoning comprehension task <sup>15,40</sup>	ICLEv3 corpus <sup>12,37</sup> Room for Debate (NYT) <sup>15</sup>	Enthymematic arguments Warrants, alternative warrants	A M
Warrant, premise	Evidence detection <sup>24</sup>	Stanford NLI (SNLI), <sup>41</sup> Multi NLI (MNL) <sup>42</sup> Context Dependent Evidence Detection, <sup>43</sup> Room for Debate (NYT) <sup>15</sup>	Correct warrants from two options Warrants extracted from existing corpora	A A
Claim / conclusion	Target inference <sup>23</sup> Modelling implicit argumentation by explicit stances <sup>47</sup> Enthymeme detection and reconstruction <sup>12</sup>	Claim Stance Dataset, <sup>44</sup> iDebate, <sup>45</sup> Argument Essays V2 <sup>46</sup> Twitter at SemEval 2016 Task 6 <sup>48</sup> ICLEv3 corpus <sup>12,37</sup>	Targets of implicit conclusions Explicit targets of debates, stances towards targets Enthymematic arguments	A M A

Note. (AC=argument components, M=manual ann., A=automatic ann.)

very specific, as it assumes the presence of main claims in the argumentation, which is not always the case unless the debate has a well-defined topic where opinions are explicitly related to that topic.

The tasks of Enthymeme Detection<sup>11</sup> and Enthymeme Reconstruction<sup>14</sup> focus on identifying and reconstructing implicit premises in arguments. Effective reconstruction of implicit elements first requires identifying missing elements. To do that, the ArguAna corpus<sup>36</sup> used by Rajendran et. al.<sup>11</sup> is manually annotated with the labels ‘explicit opinions’, ‘neutral opinions’ and ‘implicit opinions’ that allow to progress in automatic detection of enthymemes. The authors train a binary classifier on this annotated dataset to distinguish between implicit and explicit opinions automatically. This method demonstrates its utility and serves as an initial step toward enthymeme detection in natural language texts. On the other hand, Stahl et al.<sup>12</sup> take a step back to explore the creation of enthymematic dataset from scratch so as to automatically detect and reconstruct arguments with implicit components at a later stage. Their approach is based on the ICLEv3 corpus,<sup>37</sup> where they systematically remove argument components (claims or premises) to construct enthymematic arguments. This method ensures an extensive set of high-quality enthymemes, as the process of removal can be controlled at each step. The creation of such a dataset offers several advantages. First, the controlled removal ensures that enthymemes remain natural, sound, sufficient and maintain logical coherence. Second, it provides a diverse training resource for models, allowing exploration of different implicit argument components. Finally, this method facilitates evaluation of detection and reconstruction models, since the removed components serve as a gold standard for assessing performance. Following that, the authors propose baseline approaches for enthymeme detection and reconstruction that prove, on the one hand, the usability of the dataset, and the ability of computational approaches to learn from data and accomplish the task, on the other.

When comparing these two approaches of enthymeme detection, we observe that both have practical advantages, but some limitations are also present. The method proposed by Rajendran et al.<sup>11</sup> is grounded in real-world argumentative texts, ensuring that it reflects the authenticity and veracity of natural arguments. Additionally, the manual annotations provide high-quality metadata, offering a strong foundation for model training. In contrast, the corpus created by Stahl et al.<sup>12</sup> is more artificial, as its construction involves removing argument components to create enthymemes, which may not fully capture the natural occurrence of implicitness in real-world contexts. At the same time, controlled creation of the second (Stahl et al.<sup>12</sup>) ensures higher data quality and offers the advantage of removed components serving as a gold standard for reconstruction – an opportunity that the approach of Rajendran et al.<sup>11</sup> does not provide. Furthermore, both methods demonstrate immediate applicability for the automatic detection of implicitness in argumentation, as they rely on training binary classifiers on their respective datasets. However, the applicability of these methods may be limited to specific domains. The first approach, using the ArguAna corpus, focuses on opinions, which may not encompass the full range of implicit argumentation phenomena. Meanwhile, the second approach, based on the ICLEv3 corpus, is better suited for applications like teaching students to write argumentative essays, which could restrict its broader applicability.

The task of Enthymeme Reconstruction in its turn represents the process of generating an implicit premise (or claim, as discussed later in this section) for an incomplete argument. As previously mentioned, reconstruction cannot be accomplished without a preceding detection step, making these tasks complementary. For instance, the work of Stahl et al.,<sup>12</sup> which we presented just before in the context of corpus creation and automatic enthymeme detection, naturally includes the reconstruction step. The authors frame enthymeme reconstruction as a generation task, where the model is provided with an enthymematic argument and a special mask token indicating the position of the implicit element. This approach demonstrates both the utility of the created corpus and the effectiveness of automatic generation, even at a baseline level. Meanwhile, Chakrabarty et al.<sup>14</sup> consider enthymeme reconstruction as a stand-alone task, assuming the argument to be incomplete. They also frame enthymeme reconstruction as a generation task but demonstrate that fine-tuning a baseline model on enthymematic data alone is insufficient for effective reconstruction of implicitness. To address this, they propose enhancing generation with commonsense knowledge, resulting in a more advanced reconstruction approach that achieves improved results. In general, the task of Enthymeme Reconstruction is inherently more complicated than that of Enthymeme Detection. Enthymeme detection is typically framed as a binary classification task (implicit vs. explicit), which is relatively straightforward for machine learning models. In contrast, reconstruction requires generating the missing component (a premise in our context). This involves creating meaningful, contextually appropriate, and logically coherent text, which is far more complex. Additionally, while detection often relies on surface-level linguistic cues or knowledge of argument structure, reconstruction demands a deeper understanding of the argument’s context and underlying reasoning. Furthermore, reconstruction frequently depends on integrating commonsense or domain-specific knowledge, adding an extra layer of complexity to the task. Despite these challenges, the necessity of enthymeme reconstruction is undeniable. Reconstructed premises provide valuable insights into deeper human reasoning behind arguments. Also, they enable experimentation with alternative methods for generating argument components of higher quality.<sup>14</sup>

## 5.2 Implicit warrant

Argument Reasoning Comprehension Task<sup>15,40</sup> is meant to fill in the gap between a premise and its claim in order to restore the reasoning. It is based on the assumption that to reach complete comprehension of an argument or to frame a new argument, humans apply common knowledge and reasoning skills that basically remain tacit. This common knowledge constitutes a warrant – an argument component between a claim and a premise that justifies the fact that the premise entails the claim, as in Example 2:<sup>15</sup>

- (2) *Premise: College students have the best chance of knowing history.*  
*Claim: College students' votes do matter in an election.*  
*Warrant: Knowing history means that we won't repeat it.*

Here, the reconstructed warrant allows us to restore the reasoning behind the argument consisting of a premise and a claim: *College students have the best chance of knowing history. Since knowing history means that we won't repeat it, therefore, College students' votes do matter in an election.* Without this warrant humans are able to comprehend the relations between the premise and the claim, but their guesses may also vary according to their education, main values and even age. In this context, the Room for Debate dataset<sup>15</sup> has been annotated with manually reconstructed warrants and alternative warrants for opposite claims. The underlying intuition is that an alternative warrant reconstructed for an opposite claim (a claim of a twisted, opposite stance) can guide the reconstruction of the original claim by applying the same reasoning process with minimal modifications. Although this approach looks complex at the first glance, it provides valuable insights. Computational results indicate that current models used in argumentation, such as neural models, are not particularly effective for this task, often performing poorly when selecting the correct warrant from two options (the warrant for the original claim and the alternative warrant). This poor performance is likely due to the high lexical similarity between the original and alternative warrants. Since alternative warrants often differ only in negation, models struggle to distinguish them accurately. Nevertheless, this task proposes new challenges and a high-quality dataset for the future evolution of automatic generation of implicit warrants and argumentative analysis of incomplete arguments. SemEval-2018 Task 12<sup>40</sup> advances the automatic reconstruction of implicit warrants within the Argument Reasoning Comprehension Task. The datasets used in the task are enhanced with automatically selected correct warrants from available options — options that originally lead to contradicting claims. To identify the correct warrant, the authors leverage transfer learning from the Natural Language Inference task. This approach is valuable as it demonstrates the potential to address the challenge of distinguishing between two lexically similar stances (or lexically similar warrants, in this context). Moreover, the annotation and methodology provide a foundation for developing automatic systems to analyze implicit warrants. Both approaches to the Argument Reasoning Comprehension task<sup>15,40</sup> point into the same direction: Advancing the automatic reconstruction of implicit warrants by first proposing the selection of a correct warrant from two possible options. Nevertheless, the approaches are elaborated differently: The former relies on manual annotations, which add complexity and presumably enhance the quality of the data, while the latter employs a transfer learning method that demonstrates better performance compared to other existing systems.

The task of Evidence Detection<sup>24</sup> has implicit warrants and premises in its focus of attention. The objective is to improve the detection of correct evidence (premise) for a claim by leveraging warrants. In order to explore the reasoning behind an argument, it is significant to restore both, warrants and premises. Thus, the authors of the task take a step back to shed light on both implicit components, assuming that reconstructed warrants may improve the identification of correct premises. Their approach involves first automatic extraction of implicit warrants given a claim and a premise from other existing corpora, to then verify that these warrants may help to correctly select a premise (now implicit) for a claim. This method resembles the previously described approaches to implicit warrant reconstruction, as it also relies on selecting a suitable component. In this specific case, extracting warrants from existing data ensures that the reasoning aligns with real-world arguments avoiding artificial noises that can arise during generation. Furthermore, by using data that reflects how humans naturally construct arguments, this approach enables LMs to learn reasoning patterns that are close to human logic. As a result, the approach holds significant potential for automatic reconstruction of arguments similar or identical to authentic human argumentation.

## 5.3 Implicit claim/conclusion

The last group of tasks concerns claims or conclusions that are left implicit in an argument. Target Inference task<sup>23</sup> proposes to reconstruct implicit conclusions with the help of restored targets of the conclusions. It is grounded in the assumption that targets explicitly stated in premises correspond to conclusion targets of the same argument. In this way,

it is possible to infer conclusion targets from premise targets. Then, conclusion targets may be further leveraged in automatic generation of implicit conclusions, so that this generated conclusion is close to an authentic or a human-proposed conclusion. Considering an example:<sup>23</sup>

- (3) *Premise targets: Relocating to the best universities; Improving the pool of students; Online courses; Stanford University's online course on Artificial Intelligence*  
*Conclusion target inferences: Online courses; Distance-learning*  
*Conclusion target ground-truth: Online courses*

Example 3 represents premise targets automatically extracted from the dataset with conclusion targets automatically inferred from them. As can be seen, one inferred conclusion target perfectly matches with the ground truth, while the second one is semantically close to the ground truth.<sup>23</sup> This novel approach to implicit conclusion reconstruction is valuable and practical due to several reasons. First it ensures that generated conclusions remain consistent with the argument's actual content, reducing the likelihood of irrelevant outputs. Second, using inferred targets aligns conclusions with premises, therefore enhancing the coherence of argumentation. Additionally, exploiting explicit information within an argument may avoid the need for external knowledge, while also limiting risks associated with purely generative techniques, such as hallucinations and biases. Overall, the authors propose a self-contained methodology that successfully generates meaningful conclusion targets. This approach has the potential to produce coherent conclusions and can also be combined with other methods to further enhance generation.

Modelling Implicit Argumentation by Explicit Stances task<sup>47</sup> proposes to leverage an explicit stance of an argument and overall explicit debate stances to reconstruct implicit claims. The authors hypothesize that humans can infer an argument given a stance, which represents a standpoint supporting or contradicting a target. Consequently, using explicit stances to model complete arguments is a reasonable approach. To support this hypothesis, the authors first semi-automatically select debate targets that define positions on the debate topic. Next, stances toward these targets are manually annotated. In a subsequent step, the authors verify whether targets and stances can be assigned automatically, demonstrating the potential for automatic detection of debate stances. This approach is promising not only for detecting stances but also for improving the reconstruction of implicit conclusions, and it shares similarities with the previously discussed Target Inference task. As in the previous task, here the authors explore how to exploit information contained within arguments, avoiding the need for external resources and maximizing the use of the information already available. Therefore, the advantages of this methods, if further used for implicit claim reconstruction, are clear: It ensures the relevance and coherence of reconstructed claims, it minimizes the risk of generative errors and biases, and it offers broad applicability across various domains of argumentation.

The last task addressing implicit claims, Enthymeme Detection and Reconstruction,<sup>12</sup> has been extensively discussed as a task that deals with implicit premises in argumentation. However, the authors of this approach focus on both argument components – premises and claims – by creating a corpus of enthymematic arguments containing either implicit premises or implicit claims. The main advantage of this approach to implicit claim detection and reconstruction is that the task is divided into two blocks (detection and reconstruction), allowing each block to be used independently or improved separately. In comparison, the Target Inference<sup>23</sup> and Modelling Implicit Argumentation by Explicit Stances<sup>47</sup> tasks presuppose that claims are left implicit, focussing exclusively on reconstructing the missing claims. As a result, these tasks have to be specifically adapted for datasets where the first step involves detecting whether implicit claims are present in the argument at all. Another advantage of the Enthymeme Detection and Reconstruction approach by Stahl et al.<sup>12</sup> is that it accounts for the interplay between claims and premises. In contrast, all the other tasks presented focus on a single argument component that may be left implicit. Therefore, this approach represents a significant step forward in the detection and reconstruction of enthymemes.

So far, we have primarily explored argument components that may be left implicit, as they represent a challenge for computational approaches. Nevertheless, the ability to properly recognize relations connecting different arguments plays a major role in understanding the argumentation. Relations of support or attack can hold between argument components, such as premise–claim, premise–premise, claim–claim. The difficulty of their automatic detection lies in their implicit nature: Quite often argument components are connected to each other via implicit inferences that may be retrieved only with the help of external knowledge.<sup>49</sup> However, the literature on implicit argumentative relations identification is sparse. Argumentative Relation Classification task<sup>49</sup> demonstrates how injected commonsense knowledge improves argumentative relations detection on two datasets: Student Essays version 2<sup>46</sup> and a dataset from Debatepedia.<sup>49</sup> Saadat-Yazdi et al.<sup>33</sup> also assume that the majority of argumentative relations are implicit and the detection of implicit argumentative relations is highly dependent on commonsense knowledge. In particular, authors highlight that commonsense knowledge in an argumentative unit is expressed in warrants that are often left implicit. Therefore, it is necessary to explore and reconstruct

implicit warrants. Authors use Student Essays corpus,<sup>50</sup> Debatepedia<sup>49</sup> and M-Arg Presidential Debate corpus.<sup>51</sup> All in all, implicit argumentative relations still require extensive and thorough exploration, particularly in understanding the role of external information and the importance of human reasoning for the complete comprehension of argumentative flow.

## 5.4 Discussion

Although main argument components that can be left implicit are already addressed in various studies, many gaps still persist and the defined tasks remain challenging for automatic analysis. First, almost all the research works described here represent either initial or preliminary studies of enthymeme detection and/or reconstruction, meaning that future works are indispensable. Some of these primary studies already set further goals, while proposing first and necessary steps to approach them,<sup>11,23</sup> others create valuable datasets that will facilitate detection and reconstruction tasks and improve evaluation.<sup>12</sup> Majorly, they test baselines so as to set an initial threshold for the tasks. Therefore, we can claim that there is still an open perspective of in-depth approaches addressing different implicit argument components. Beyond that, some of the methods of enthymeme reconstruction take a step towards employing extralinguistic knowledge to improve generation quality and versatility,<sup>14</sup> but such strategies are underexplored. It is essential to push the research in this direction in order to explore how various linguistic characteristics, sociological features or external knowledge may aid not only in enthymeme reconstruction, but in the detection phase as well, and not only for premises, but for all argument components. Additionally, the studies discussed in this section explore implicitness in argumentation across various fields of application, such as debates, social media posts, essays, and learners' argumentation etc., however, none address the generalizability of the proposed techniques. It might be practical to apply each method to different fields to determine whether argument components vary significantly between domains or if the same approaches can be effectively utilized across multiple domains. Such a study would reveal if it is mandatory to elaborate different strategies for each domain or it is possible to find and enhance the only one applicable to all fields. Furthermore, this investigation could identify the most challenging domains for enthymeme analysis, opening new directions for scientific exploration.

## 6 Argument explication

This section focuses on the examination of established methodologies for restoring implicit elements, a process referred to as argument explication.<sup>10</sup> This investigation will entail an exploration of the knowledge sources employed in argument reconstruction, as well as a detailed examination of the existing techniques facilitating the transition of an argument from an implicit to an explicit state.

### 6.1 Required knowledge and proposed methods

The exploration of argument explication unfolds along two research directions: (i) Some works assert that the contextual information surrounding an argument may be sufficient to restore its implicit components, facilitating argument understanding; (ii) in contrast, alternative studies emphasize the necessity of additional knowledge beyond the argument context. This additional knowledge encompasses both universally shared commonsense knowledge and domain-specific knowledge inherent to a particular field. As mentioned by Lauscher et al.,<sup>7</sup> argument reasoning is highly dependent on commonsense and domain-specific knowledge, therefore, this approach seems sound. While these two directions are shortly described by Becker et al.,<sup>34</sup> we propose a thorough analysis in the following (see [Table 3](#)).

Works pertaining to the first group consider local context of an argument as sufficient for implicit components identification and reconstruction. More precisely, it is possible to reconstruct an enthymeme using information within the argument to detect missing components and filling these gaps with the help of similar or related arguments. For example, Rajendran et al.<sup>11</sup> propose analyzing opinions from an online hotel reviews dataset to facilitate the detection and reconstruction of implicit argument components within the dataset, using opinion analysis as a foundation. The authors hypothesize that opinions about a hotel or its specific aspects, classified as implicit or explicit, indicate whether an argument is complete or enthymematic. To test this hypothesis, they annotate the stances of opinions as either implicit or explicit. In addition to classifying opinions, they suggest extracting sentiments from these opinions to further support the reconstruction of incomplete arguments. Retrieved sentiments (positive or negative) are then used to reconstruct predefined conclusions for the reviews. Thus, the authors already address cases where the implicit component of an enthymeme is a conclusion. This approach utilizes both metadata and the dataset itself as sources of essential knowledge, aiding in the detection and reconstruction of enthymemes. Following the same direction, Alshomary et al.<sup>23</sup> propose using the target of an implicit conclusion as a basis for its reconstruction. To identify the target of the conclusion, they first determine the target of an explicit premise within the same argument. The authors hypothesize that the target of the premise typically corresponds

**Table 3.** Argument explicitation knowledge and methodologies.

Explicitation knowledge		Methods	Research papers
Local context		Leverage sentiment stance classification of opinions and other explicit opinions within the same dataset in order to facilitate enthymeme detection and reconstruction	Contextual stance classification of opinions: A step towards enthymeme reconstruction in online reviews <sup>11</sup>
		Define a target of a premise and exploring other arguments within the same dataset help to define a target of a conclusion and reconstruct it	Target Inference in Argument Conclusion Generation <sup>23</sup>
		Define an overall stance of a debate (being in favour or against a defined target) to reconstruct implicit claim	Stance-based AM-Modelling Implicit Argumentation Using Stance <sup>47</sup>
Knowledge enhancement	Shared knowledge	Annotate semantic clause types of handcrafted implicit argument components for enthymemes detection and reconstruction + Annotate commonsense knowledge relations (ConceptNet)	Enriching Argumentative Texts with Implicit Knowledge <sup>34</sup> ; Implicit Knowledge in Argumentative Texts: An Annotated Corpus <sup>16</sup>
		Employ discourse-aware commonsense knowledge model to generate implicit components + Apply abductive reasoning for incomplete arguments	Implicit Premise Generation with Discourse-aware Commonsense Knowledge Models <sup>14</sup>
		Use similar and large datasets to introduce missing commonsense knowledge for automatic reconstruction of implicit components	GIST (SemEval-2018 Task 12): A network transferring inference knowledge to Argument Reasoning Compreh. task <sup>40</sup>
	Apply commonsense knowledge relations (ConceptNet) to improve argumentative relation classification + Utilize normative commonsense knowledge model to generate implicit warrants that assist in the automatic detection of implicit argumentative relations	Argumentative Relation Classification with Background Knowledge <sup>49</sup> ; Uncovering Implicit Inferences for Improved Relational AM <sup>33</sup>	
	Domain-specific knowledge	Train annotators with domain-specific knowledge + Use domain-specific knowledge for the reconstruction of implicit components	casiMedicos <sup>52</sup>

to the target of the conclusion, as all components of an argument collectively work toward addressing a single subject. Consequently, they reconstruct the target of the conclusion based on the identified premise target, which then serves as a foundation for reconstructing the conclusion itself. Once again, this method emphasizes leveraging the inherent knowledge and relationships present within the dataset under study. A work of Wojatzki and Zesch<sup>47</sup> is grounded in the correlation between the explicit overall stance of a debate and an argument within that debate that contains implicit components. Specifically, they assert that it is possible to infer an implicit claim of an argument by considering the explicit stance of that argument and the explicit overall stance of the debate. Consider Example 4:

- (4) *Premise: As the Bible says that infidels are going to hell!*  
*Debate stance: being against atheism*  
*Argument stances: being in favour of Christianity; being in favour of existence of hell*

Having as explicit premise ‘*As the Bible says that infidels are going to hell!*’ we are unable to restore a claim. However, taking into consideration the stances of this incomplete argument, that are ‘*being in favor of Christianity*’ and ‘*being in favor of existence of hell*’, and the overall explicit stance of the debate, we can reconstruct the claim ‘*I am against atheism*’. The complete argument will be: ‘*As the Bible says that infidels are going to hell! I am against atheism*’.

The second facet of argument explicitation presented in Table 3 pertains to knowledge enhancement. It encompasses harnessing shared knowledge or employing domain-specific knowledge for the purpose of enthymeme reconstruction and implicit argumentative relations detection. Shared knowledge utilization is based on the assumption that in order to restore an enthymeme or implicit argumentative relations, it is necessary to apply shared knowledge outside the scope of the argument. Reconstructing implicit elements often requires understanding unstated assumptions or connections that are commonly accepted or understood within a broader context. Shared knowledge bridges this gap, providing the necessary context to make sense of implicit components that are not explicitly stated in the argument itself. There are several explored methodologies of introducing common knowledge to facilitate the reconstruction of incomplete arguments. Primarily, such knowledge may be incorporated in the form of commonsense knowledge relations. In this context, ConceptNet<sup>16</sup> serves as a graph repository of commonsense facts about the world interconnected by relations. According to this method, commonsense knowledge is universally shared, thus, humans do not include it in the discourse. Intuitively, we need to apply to commonsense knowledge, so as to restore implicit argument components. Additionally, authors claim that the inclusion of semantic clause types proves beneficial in the process of enthymeme reconstruction. Being supplementary information about an argumentative text, semantic clause types potentially facilitate argument explicitation.<sup>16,34</sup> Another method of incorporating commonsense knowledge introduced by Chakrabarty et al.<sup>14</sup> is using a discourse-aware commonsense knowledge model. PARA-COMET is a knowledge model that is able to generate commonsense inferences and relations on the basis of its own knowledge graph (KG) of commonsense relations and given texts from the dataset. As commonsense knowledge forms the foundation for implicit argument components, leveraging generated knowledge has the potential to enhance the reconstruction of these components. As an auxiliary approach, authors propose to appeal to abductive reasoning implying that humans reach out to this kind of reasoning to comprehend incomplete arguments. Consider an example:

(5) *Reason: Vaccinations save lives.*

*Claim: Vaccination should be mandatory for all children.*

*Reconstructed premise: Vaccinations are the best way to prevent childhood diseases.*

In Example 5, the implicit premise generated by the model with PARA-COMET commonsense knowledge serves as an inference between the stated claim and its reason. Without this connection, it is not clear why the author mentions only children in the claim, as the reason acknowledges the benefit of vaccination in general. Yet another methodology to incorporate common knowledge is to use similar corpus that may provide missing information and relations. Transferring knowledge from other datasets has the potential to improve the reconstruction of implicit argument components.<sup>40</sup> This claim is based on the assumption that other large datasets can fulfil unstated essential knowledge by providing numerous texts. Therefore, the authors use this knowledge to address argument reasoning comprehension task, in particular, to select a correct warrant out of two possibilities automatically. Last but not least, recent papers consider that implicit argumentative relations are also highly dependent on commonsense knowledge.<sup>33,49</sup> Authors of these researches propose incorporating ConceptNet commonsense knowledge relations<sup>49</sup> and generating missing knowledge with the Commonsense Transformer (COMET) that is able to produce necessary warrants in a form of chains of inferences.<sup>33</sup> The second approach has better performances on the same data, presumably due to the fact that COMET varies according to cultures and beliefs while ConceptNet is a static KG providing universal knowledge.

Despite the diversity in methods for integrating commonsense or shared knowledge into argument explicitation, there remains an underrepresentation of both commonsense and domain-specific knowledge. For instance, in a comprehensive survey of knowledge use in computational argumentation, researchers join both categories of knowledge, underlining the persistent challenge of harnessing this extensive domain of knowledge.<sup>7</sup> Moreover, state-of-the-art works focus mostly on exploiting commonsense knowledge while domain-specific knowledge remains unexplored. It is clear that domain-specific focus narrows the generalizability and practical application, however, it might be needed for specific tasks in specific spheres. In this regards, we would like to highlight that using domain-specific knowledge for enthymeme reconstruction in such spheres as medical or legal domain is significant for improvement of manual annotations and development of automatic identification with further reconstruction of incomplete arguments. In the medical domain, *casiMedicos*<sup>52</sup> is a publicly available dataset of medical questions with possible answers to the questions and explanations of a correct answer which constitute an argument. Analyzing it, we noticed that clinicians tend to omit pieces of evidence that are clear to medical community or replace medical terms with jargon producing miscomprehension by general public. Considering an example:

(6) *Premises: A 6 months-old infant presenting to the emergency with axillary temperature 37.2°C, respiratory rate 40 rpm, heart rate 160 bpm, blood pressure 90/45 mmHg, SatO2 95% on room air. He shows moderate respiratory*

*distress with intercostal and subcostal retraction. Pulmonary auscultation: scattered expiratory rhonchi, elongated expiration and slight decrease in air entry in both lung fields. Cardiac auscultation: no murmurs.*  
*Claim: The patient probably presents with bronchiolitis.*

In Example 6, in order to claim the diagnosis, it is necessary to know which symptoms should be taken into account as relevant. Here the reasoning behind the choice of the disease is unstated: It is not clear even for a human which information provided in the premise leads to the conclusion. Manual analysis of such implicit elements reveals that commonsense knowledge is not sufficient for their reconstruction. Therefore, it might be beneficial to incorporate domain-specific knowledge from biomedical ontologies (e.g., Human Phenotype Ontology<sup>53</sup>) and biomedical vocabulary (e.g., Unified Medical Language System (UMLS)<sup>54</sup>) to be able to identify and reconstruct enthymemes in medical domain.

## 6.2 Discussion

In our exploration of argument explicitation methodologies, we observed an interesting trend. A subset of studies adopt an argumentation model as the foundational framework for further research. This argumentation model serves as a cornerstone for identifying implicit argument components. Subsequently, researchers employ diverse knowledge and apply various techniques within the framework of this model to reconstruct implicit argument components. As a result, the entire process of exploring and reconstructing argument reasoning is centralized around the major role of the argumentation model and its components.<sup>11,14,15,23,24,35,40,47</sup> These approaches can be referred to as Model-based Explicitation or Explicitation Based on Enthymeme Reconstruction, depending on whether the emphasis is placed on the reconstruction of enthymemes or on the fact that the arguments under study represent a certain argumentative model.<sup>10</sup>

Another group of studies highlights that the argument reasoning comprehension relies on background commonsense knowledge, universally shared among humans, along with reasoning skills. As shared knowledge is generally evident to most individuals, it needs not be explicitly articulated. Human reasoning skills, a component hidden between the lines, also requires deeper study. With these integral elements, humans effortlessly achieve a comprehensive understanding of arguments. However, computational models meant to analyze natural language are equipped with neither background knowledge nor with reasoning skills. Thus, research within this direction brings forward the necessity to reconstruct commonsense knowledge that is majorly left implicit.<sup>16,33,34,49</sup> This approach can be referred as Knowledge Enhancement-based Explicitation of enthymemes.<sup>10</sup>

The above mentioned directions, however, are not mutually exclusive. In Model-based Explicitation, authors underline the importance of argumentative structure for the exploration of implicitness. Nevertheless, they do not obviate methods to incorporate commonsense knowledge to improve enthymemes detection/reconstruction. Similarly, Knowledge Enhancement-based Explicitation methods do not dismiss the presence of model-based representation of arguments, they rather accentuate the imperative of having commonsense knowledge integrated in automatic analysis of argumentation.

It is also noteworthy to mention that the identification of the best argument explicitation method is currently challenging due to several reasons. First, the scope of existing research is still limited to allow for meaningful comparisons of different approaches on the same task. For example, it is unreasonable to compare and rate approaches for explicitation of different components (e.g., implicit claims reconstruction and implicit premise reconstruction) as the detalization and depth vary, while different components presumably require distinct knowledge for reconstruction. Therefore, it is needed to have more studies on each component, which is not yet the case. Second, factors such as limited resources and privacy concerns (e.g., clinical data), restrict the range of methods that researchers can test on certain enthymematic datasets. Additionally, the diversity of datasets and domains further complicates the identification of a universal approach, as different methods may perform variably depending on the type of data or the nature of the implicit components. For instance, approaches that work well for reconstructing implicit premises in short argumentative texts may not generalize effectively to long, domain-specific discourses. Moreover, the absence of standard evaluation metrics for enthymeme reconstruction adds another layer of complexity, making it difficult to directly compare the effectiveness of various techniques.

## 7 Computational approaches towards enthymeme detection and reconstruction

So far, we have discussed the steps involved in enthymeme analysis, introduced argumentation models that support this process, explored tasks designed to detect and reconstruct implicit argument components and analyzed key techniques for argument explicitation. In this section, we will describe computational approaches predominantly used for enthymeme detection and reconstruction, taking into account main ways to introduce universally shared knowledge into automatic enthymeme analysis.

The foundation for developing robust computational approaches to study implicitness in argumentation begins with manual annotations. Early research in enthymeme analysis relied heavily on manual annotation of argument

components,<sup>25,35</sup> stances in opinions,<sup>11</sup> stances towards targets,<sup>47</sup> reasons,<sup>15</sup> and commonsense relations.<sup>16</sup> Also, before the adoption of automatic techniques, researchers focussed on preliminary manual approaches to simulate automatic analysis. These included manual classification of argument components into semantic types,<sup>25</sup> manual reconstruction of implicit knowledge,<sup>16</sup> manual reconstruction of argument components such as warrants and alternative warrants,<sup>15</sup> human similarity judgment (e.g., semantic similarity between a claim and a major claim).<sup>35</sup> Such foundational steps are crucial for several reasons. First, human-annotated data enables researchers to explore the nature of implicitness, uncovering patterns and features that characterize implicit premises, claims, and warrants, while also providing valuable insights into human reasoning. Additionally, manual analysis serves as the ground truth for evaluating automatic approaches. Finally, it demonstrates the feasibility of tasks related to implicitness and helps identify potential challenges and pitfalls for future automatic analysis.

With the foundational groundwork established, we now turn to the computational approaches used for enthymeme analysis. The two core tasks – enthymeme detection and reconstruction – shape two distinct groups of approaches, which in turn define two categories of computational models.

**Enthymeme detection.** In NLP terms, the detection task can be framed as *detection*, *classification* or *prediction* with each method focussing on binary classification of arguments into implicit or explicit, identifying argument components or argumentative relations, predicting labels for components, or predicting correct components and their relations. For instance, prediction of a correct warrant from a pair of two lexically close warrants that lead to contradicting claims was proposed by Choi and Lee,<sup>40</sup> while Singh et al.<sup>24</sup> test automatic detection of (implicit) relations between a claim and an evidence (premise) to further extract warrants for these relations so as to improve evidence reconstruction. As for computational models used within this first direction, two major paths are distinguishable. Initially, researchers relied on simple linear classifiers as baselines and neural networks, even highlighting that the development of more sophisticated methods was postponed for future work. Among linear classifiers, Logistic Regression and SVMs have been predominantly used.<sup>11,35,49</sup> In neural approaches, the most prominent model is the bidirectional LSTM, often combined with various attention mechanisms and parameters. This model has been utilized for encoding argument components,<sup>40</sup> predicting correct warrants<sup>15,40</sup> and classifying argumentative relations.<sup>49</sup> With the emergence of more advanced architectures, particularly transformer-based LMs,<sup>55</sup> the focus shifted towards using such LMs for detection and classification tasks. Among these, BERT and its enhanced variants, such as DeBERTa and RoBERTa, have become the most prominent models. As such, Stahl et al.<sup>12</sup> utilize DeBERTa for the enthymeme detection task, while Delas et al.<sup>56</sup> implement argument scheme classification using 21 BERT-based models, each model corresponding to one scheme.

**Enthymeme reconstruction.** The reconstruction task, in turn, is formulated in computational terms as either *extraction* or *generation*. *Extraction* involves retrieving argument components, argumentative relations or commonsense knowledge from external sources (e.g., ConceptNet commonsense knowledge resource<sup>57</sup>). In contrast, *generation* focuses on automatic creation of meaningful and coherent implicit components or implicit knowledge. As with the detection task, the initial steps in reconstruction – particularly the extraction tasks – were not automated and relied on manual retrieval of relevant information. First automatic approaches to extraction and generation utilized neural models, among which the most used were LSTM-type models. For instance, LSTMs were employed to extract commonsense knowledge from ConceptNet<sup>49</sup> or generate conclusion targets for the following reconstruction of implicit conclusions.<sup>23</sup> Again, the advent of transformers and specifically Large Language Models (LLMs) introduced the capability to generate natural language text without relying on templates or predefined rules. This development made it feasible to generate missing argument components entirely from scratch or to combine the generative process with commonsense KGs or repositories for more relevant results. Various LLMs have been utilized for generating implicit argument components and implicit knowledge, demonstrating relatively strong performance. For instance, sequence-to-sequence model BART was employed for implicit premises and claims generation,<sup>12,14</sup> COMET<sup>33</sup> was utilized for warrants generation, while autoregressive GPT-2 and bidirectional autoregressive XLNet restored implicit commonsense knowledge.<sup>58</sup>

Additionally, several approaches to enthymeme detection and reconstruction focus on incorporating commonsense knowledge into the classification or generation process. These techniques typically involve either using a static structured KG<sup>49</sup> or leveraging a specialized knowledge model capable of generating commonsense knowledge based on a KG and the given discourse.<sup>14,33</sup> In the first approach, encoded commonsense relations from the KG can be directly injected into a neural model to define relations between arguments or into a transformer model to generate missing information, such as implicit argument components. In the second approach, the transformer-based knowledge model COMET dynamically generates novel knowledge by combining the commonsense relations of the KG it was trained on with the current discourse. This newly generated knowledge can then be injected into a transformer model to generate implicit argument components or identify the relations between them. According to the performances presented in Chakrabarty et al.,<sup>14</sup> Paul

et al.,<sup>49</sup> Saadat-Yazdi et al.,<sup>33</sup> dynamic generation of knowledge proves to be more meaningful and coherent, as it considers specific arguments it is applied to, ensuring consistency with the context. In contrast, the static approach lacks this adaptability.

For all the approaches discussed, computational models can be provided with either pairs of sentences (e.g., a premise and a claim) or individual sentences, with or without accompanying metadata. While LLMs are capable of processing longer text sequences with reasonable performance, research on enthymemes has not yet fully explored this direction. This hesitation stems from the challenges of the task: Within a pair of sentences, it is difficult enough to detect missing information, and it is even more complicated to generate meaningful argument components that fit the overall context without duplicating explicit parts. Additionally, processing larger texts, whether annotated or not, reduces control over the reconstruction process. As a result, biases and inconsistencies in generated components are harder to trace back to their source, making it difficult to identify and reduce potential quality issues. Lastly, detecting exact locations of missing information, maintaining logical consistency throughout a full-text argument, and aligning generated components with the nuances of the argument's domain require advanced modelling techniques and fine-tuning. Therefore, addressing these challenges will be critical for the future advancement of automatic systems.

Evaluating and comparing computational approaches to enthymeme analysis is challenging for several reasons. First and foremost, the field of enthymeme studies is not yet extensive enough to allow for meaningful comparisons between approaches addressing the same task. Instead, most approaches explore unique directions, making it difficult to identify a single best method. Moreover, it is not reasonable to directly compare early approaches with more recent ones, as the former were appropriate and innovative for their time, while the latter benefit from advancements in technology and methodologies. Furthermore, certain areas of research within implicit argumentation remain underexplored, which limits our ability to fully evaluate the best technologies. Nevertheless, it has to be noted that both neural networks and transformer-based LMs are predominantly used and have demonstrated success in the tasks discussed. Transformer-based models, in particular, are gaining preference due to their scalability and consistently strong performance across various domains and tasks in implicit argumentation.

## 8 Data for argument reasoning studies

This section presents a compilation of datasets used in implicit argumentation studies (see Table 4), selected for their relevance to AM and reasoning. The table provides details about each dataset's size, source, original annotations (whether argument components and/or relations are labelled), and additional annotations introduced in subsequent works, highlighting the methodology used for annotation collection (manual or automatic). As well as that, the datasets are grouped according to their application domains: education, web-based discussions, public opinions.

Handcrafted educational datasets have been created to teach students to improve their reasoning<sup>39,46</sup> and persuasiveness of their arguments,<sup>46</sup> and to explore abductive reasoning in natural language texts.<sup>38</sup> Argument Essays V2<sup>46</sup> and Microtext corpus<sup>39</sup> consist of high-quality argumentative texts in monologue form, containing up to 40 and 5 sentences per text, respectively. These datasets have several advantages: each document has a predefined topic, thus, the documents can be analyzed and compared more effectively, each document has a homogeneous reasoning flow due to single authorship, and they all have self-contained arguments as they are manually written answers to specific themes. Both datasets were originally annotated manually with argument components and their relations. They are well-suited for implicitness detection and subsequent generation tasks, as humans often omit evident information or weak arguments. Moreover, provided annotations enable researchers to intentionally remove specific components for automatic generation experiments.

Another handcrafted dataset ART<sup>38</sup> consists of manually created 5-sentence commonsense narratives and their explanations based on abductive reasoning. Therefore, it represents a set of short argumentative texts that are also suitable for enthymeme detection and reconstruction tasks, since gaps in human reasoning are quite probable. The original version of this dataset includes neither argument component annotations nor relation classifications; however, it was later enriched with automatically generated premises.<sup>14</sup>

Datasets based on web discussions include debates on various topics (politics, abortions, human and minority rights etc.), comments on controversial issues and persuasive arguments in monologue and dialogue forms. iDebate dataset<sup>45</sup> includes 676 short argumentative texts on controversial topics from an online debate platform *idebate.org*. Each discussion consists of a central claim that represents a positive or negative position towards the topic and arguments (premises) that support this position. In total, there are 2,259 claims and 17,359 premises. Although this dataset was initially created to develop methods for concise and informative summarization of argumentative texts, it is also well-suited for elaborating techniques for implicitness detection and reconstruction. For instance, the clear and well-defined conclusions of each argument in the dataset can be omitted to test the feasibility of reconstructing them based on the premises.<sup>23</sup> The evaluation

**Table 4.** Datasets in implicit argumentation studies grouped by their application domain.

	Dataset	Size	Source	AC	RC	New Ann.	M / A
Education	Argument Essays V2 <sup>46</sup>	402 essays 1552 ACs	Handcrafted	✓	✓	Implicit conclusion targets <sup>23</sup>	A
	Microtext corpus <sup>39</sup>	112 texts, 576 segments	Handcrafted	✓	✓	Implicit knowledge, semantic clause types, commonsense relations <sup>16,34</sup> ; Generated premises <sup>14</sup>	M+A
	Abd. Reasoning (ART) <sup>38</sup>	20k narratives 200k expl.	Handcrafted	✗	✗	Generated premises <sup>14</sup>	A
Web-based Discussions	iDebate <sup>45</sup>	676 debates 19,618 ACs	Online debates <i>idebate.org</i>	✓	✗	Implicit conclusion targets <sup>23</sup>	A
	Online Debate Forum <sup>35</sup>	494 enthym. 500 claim pairs	Debates <i>createdebate.com</i>	✓	✗	Generated premises <sup>14</sup>	A
	Room for Debate <sup>15</sup>	1654 triples (C+P+W)	Online debates <i>NYT</i>	✓	✗	Generated premises <sup>14</sup>	A
	Change My View <sup>59</sup>	21k discussions	Civil opinions <i>CMV</i>	✓	✗	ACs, semantic clause types <sup>25</sup>	M
	Claim Stance Dataset <sup>44</sup>	55 topics, 2394 claims	<i>idebate.org</i> <i>Wikipedia</i>	✓	✗	Implicit conclusion targets <sup>23</sup>	A
Public Opinions	ArguAna <sup>36</sup>	2100 reviews 24,5k product features, 31k statements	Hotel reviews <i>TripAdvisor.com</i>	✗	✗	Explicit / implicit opinions <sup>11</sup>	A

Note. (AC=argument components, RC=relation classification, M=manual ann, A=automatic ann).

of these reconstructed conclusions is straightforward, since the original conclusions serve as gold-standard references for the debate topics.

The Online Debate Forum dataset,<sup>35</sup> extracted from *createdebate.com*, contains arguments on four topics: Marijuana, Gay Rights, Abortion, and Obama. Each text’s main position (for or against the topic) is labelled, and every sentence is manually matched to a single major claim corresponding to the topic. Originally, this dataset was designed for ‘fill-the-gap’ task, where a gap between a user claim and a manually matched major claim represents reasoning that humans can infer. Therefore, the idea was to bridge this gap by manually annotated premises. In total, after the annotations the dataset contains 500 claim pairs (user claim + major claim) and 3977 fill-the-gap premises, meaning that each argumentative text has 8 premises in average. Subsequently, the dataset was updated with automatically generated premises, not to bridge the gap between two claims, but to enable the correct inference of conclusions and to complete enthymemes. This means that the dataset is suitable for both, specific tasks, such as detecting or reconstructing components in a specific position (e.g., a premise between two claims), and more general tasks, such as fully reconstructing an enthymeme, regardless of the specific position of implicit components.

Another debate dataset, created from scratch and claiming higher quality argumentative texts compared to those extracted from debate platforms, is Room for Debate.<sup>15</sup> This dataset comprises a collection of arguments from *The New York Times*, which, due to editorial oversight and moderation, can be considered more credible. The dataset consists of 1,654 argumentative triples, each containing a claim, a premise, and a warrant. It is well-suited for the detection and reconstruction of implicit warrants, premises, or claims; however, its strict triplet structure could be a potential limitation. Real-world argumentation is often more complex, as arguments frequently involve more than three components. For example, a single claim might be supported by several premises or warrants, thus, the triplet structure oversimplifies these dynamics, potentially excluding valuable information and overlooking multi-level relations between argument components. As well as that, such structure may restrict flexibility of the analysis, meaning that arguments with a different set of components may be poorly represented.

Change My View (CMV) dataset<sup>59</sup> comprises 21,000 civil discourse texts. The dataset is structured as follows: The initiator of a discussion creates a title for their post, which represents the major claim of their argument, and then describes the reasons for their belief. These reasons can include both claims and premises. Other participants respond in an attempt to change the initiator’s view. Their argumentative speech also includes claims and premises. If successful, the initiator indicates that their perspective has shifted. Furthermore, each discussion tree in this dataset is divided into separate

dialogues, with each argumentative text featuring one initiator and one respondent attempting to challenge the initiator's opinion. Such update allows to trace the reasoning flow and consistency of arguments, making it easier to analyze how the initiator's opinion evolves in response to challenges. It also enables a focussed examination of one-to-one argumentative interactions, providing insights into persuasive strategies. Additionally, this structure facilitates the identification of implicit components within a single dialogue, improving the dataset's utility for tasks like argument reconstruction and implicitness detection. Compared to the other datasets discussed in this section, the CMV dataset offers the richest arguments in terms of both, quantity and quality of data. Each argumentative text represents a complete discussion, featuring dynamic exchanges between the initiator and the respondent. Also, as stated in the table, the CMV dataset is annotated with argument components.

Claim Stance Dataset<sup>44</sup> introduces arguments addressing 55 topics randomly chosen from *idebate.org*. Each argument includes claims and premises that were manually extracted from *Wikipedia* articles. Furthermore, each argumentative text is enhanced with manually annotated stances for the claims, indicating whether they take a pro or contra position on the topic. The targets of claims are also highlighted. This labelling is particularly useful for implicit claim reconstruction, as the target of an argument helps define the context of the claim.<sup>23</sup> However, this dataset has a characteristic that may be a limitation for certain studies. Since each argument follows a strict structure, containing only one claim and its associated set of premises, this may limit the depth of analysis by restricting the exploration of more complex argumentative relationships.

The last group of datasets is based on public opinions and includes ArguAna dataset.<sup>36</sup> This dataset is a collection of hotel reviews extracted from *TripAdvisor.com* platform. The dataset consists of 2100 review texts with ratings of 1850 hotels across over 60 locations. In addition to ratings, each review includes metadata, sentiment scores about various aspects of the hotels (e.g., cleanliness or service), and manual annotations of hotel features and amenities. These specific annotations enable fine-grained analysis of argumentation, allowing researchers to examine how sentiment and specific hotel attributes influence the reasoning behind user ratings. Moreover, subsequent work on this dataset<sup>11</sup> demonstrates that users' opinions about hotels and their amenities are expressed both implicitly and explicitly, making this dataset particularly valuable for enthymeme studies. Thus, the dataset was further utilized for the automatic classification of implicit and explicit opinions, facilitating enthymeme analysis. However, it is important to note that the dataset is not labelled with argument components at any stage, which could be seen as both a limitation and an opportunity for future work.

**Discussion.** Despite the variety of available datasets, several gaps and challenges remain in the detection and analysis of implicit arguments. First, only two datasets include annotations for argument relations; moreover, these relations have not been thoroughly analyzed yet. This represents a significant opportunity for future research, as relations between argument components are often implicit, reflecting the natural tendency of humans to leave parts of their reasoning unexplained. Second, most datasets focus on short argumentative texts or analyze implicitness between specific components (e.g., a warrant for a set of one claim and one premise), which limits their applicability to long, real-world discourses. The CMV dataset<sup>59</sup> stands out as an exception, offering long, real-world arguments with opinion exchanges and persuasive techniques; however, realistic argumentative discussions remain underrepresented overall. Finally, there is a lack of domain-specific argumentative datasets for enthymeme analysis. For example, medical texts, which often rely heavily on professional knowledge and are thus inherently enthymematic, could provide invaluable resources for exploring implicit argumentation in highly specialized contexts. All in all, the diversity of datasets, spanning various domains such as debates, reviews etc., emphasizes the significance of comprehending reasoning in argumentation and highlights the necessity to examine implicitness in argumentation across different subjects. These datasets not only provide valuable resources for exploring implicit argument components and ways to address them, but also reveal challenges of dealing with incomplete reasoning and diverse argument structures.

## 9 Perspectives in implicit AM

In advancing the areas of argument reasoning and AM, several ambitious directions emerge, each offering unique opportunities for further development of implicit inferences exploration. These directions, when integrated and expanded upon, hold the potential to significantly contribute to the advancement of computational argumentation.

**Synthesizing explicitation methodologies.** One promising avenue involves the unification of existing methodologies of explicitation, namely Model-based Explicitation, Explicitation Based on Enthymeme Reconstruction, and Knowledge Enhancement-based Explicitation (Section 6). Although these methodologies have not been explored completely in isolation, they were focussed on their separate aspects of argument explicitation, whether that be the argumentation model or enthymeme reconstruction or knowledge incorporation. We suggest that there is considerable potential in synthesizing their strengths at equal scale. Our proposal involves coordinating a comprehensive analysis of an argument to discern its

underlying argumentation model together with enthymeme detection, followed by a subsequent phase of reconstruction. Next, we anticipate that harnessing commonsense or domain-specific knowledge to facilitate enthymeme detection and reconstruction is a relevant step. Therefore, it is imperative that this step attains equal significance to the preceding two. Incorporating insights from each methodology, we aim to create a more holistic approach that addresses the nuances of implicit argument components effectively.

**Creating implicitness typology.** Another direction involves a thorough exploration of various types of implicit inferences within argumentation. As it has been already noticed,<sup>10</sup> the intention of an implicit inference may be different, thus changing the purpose of its reconstruction. For instance, an argument may be based on factual information to make a solid support of an opinion, therefore, any missing components would be true facts so as to continue the flow of argumentation. On the other hand, an argument may represent a subjective text that is meant to give rise to an emotional response. In this case, the reconstruction of missing components might include emotional triggers. By delving into the intricacies of implicit reasoning, we seek to identify and categorize different forms of implicit inferences. This exploration serves as the foundation for developing novel approaches dedicated to the detection and reconstruction of implicit components.

**Knowledge integration.** In the pursuit of advanced AM, there is a compelling need to explore new methodologies for incorporating domain-specific and commonsense knowledge. Despite existing efforts, the current state of knowledge integration remains insufficient. We aim at stimulating the research community to deeply explore innovative approaches that transcend the limitations of previous attempts, aiming to enrich argumentative analyses with relevant knowledge base.

**Argument mining with LLMs.** Last but not least, in the realm of AM, particularly for the detection of enthymemes, traditional methodologies have typically been segmented into several distinct NLP subtasks. These subtasks often include: Span identification and/or enthymeme detection (identifying the parts of the text that constitute an argument and/or an enthymeme),<sup>11</sup> argument component classification, argumentative relation classification, commonsense knowledge encoding (integrating external knowledge to explicitate implicit entities),<sup>14</sup> enthymeme reconstruction.<sup>14</sup> While dividing the process into these subtasks has its advantages, it also introduces several challenges. Error propagation is a significant issue. Errors made in earlier stages can cascade through the subsequent stages, compounding the overall error rate.<sup>60</sup> For instance, incorrect span identification can lead to misclassification of components. Biases constitute another challenge. Biases can stem from training data, model design, or task-specific constraints, thus, they can accumulate throughout the subtasks producing impact over the final results. Moreover, reproducibility issues arise due to the number of tuning parameters required for each subtask. Different implementations might yield varying results due to differences in parameter settings, even with the same initial data and objectives. Given these limitations, one promising future direction is the broader application of LLMs to argument mining and especially enthymeme reconstruction.<sup>33,58,61</sup> Although these models have launched debate by shifting some research focus toward prompt engineering, it is important to acknowledge the benefits they retain and the growing interest they have garnered.<sup>62</sup> As discussed in [Section 7](#), LLMs have already found their use in enthymeme reconstruction, specifically for generating implicit components, and have demonstrated competitive results. LLMs offer an approach by handling the entire process of enthymeme analysis as a generation task, therefore, all the usual subtasks can be transformed into a cohesive process. This reduces the need for separate steps and minimizes error and bias propagation. Also, by using a single model to handle all aspects of argument mining, the complexity and variability introduced by multiple tuning parameters are significantly reduced. The ability of LLMs to infer missing information makes them well-suited for enthymeme detection and reconstruction, while their rapid development will make this approach more feasible and the results more sound. However, current limitations of LLMs still need to be taken into account. First, these models are originally multitask, meaning that they were not trained on only argument component classification or implicitness reconstruction. Thus, the quality of results might be unexpectedly good or bad. Second, LLMs are restrained over sensitive topics. For example, if our goal is to disclose implicitness in hateful argumentation, LLMs will introduce biases due to their regulations over certain themes or they will not be able to provide us with proper hateful content. Despite these constraints, as these models continue to evolve, they will likely become quite effective at handling the nuances of implicit reasoning. Future research should fully explore the potential of LLMs in this context, investigating their strengths and how they can be effectively combined with other supervised approaches.

These four perspectives comprise a significant step forward in the maturation of advanced reasoning methods.

## 10 Conclusion

In this survey paper, we explored the process of enthymeme analysis, from the initial steps of identifying incompleteness in arguments to the advanced stages of restoring implicit elements. We began by defining the pipeline for enthymeme analysis, outlining its key steps, and conducting an in-depth exploration of the argumentation models used in enthymeme studies. Our findings revealed that only three argumentation models are commonly employed in this field, and current studies focus exclusively on argument components from these models. Despite their potential value, argument schemes remain

overlooked due to the probable complexity of their integration in the process of implicitness analysis. To address this gap, we provided detailed descriptions of argument schemes within argumentation models to encourage further research and facilitate informed selection.

Next, we introduced argument components that are frequently left implicit, based on the current state of research. As argumentation model selection for enthymeme analysis may be considered as a preliminary or even an optional step (Section 6), identifying missing argument components remains a critical and indispensable part of the two-step pipeline for analyzing implicitness in argumentation. Our representation of implicit argument components includes the associated tasks designed to address implicitness detection and/or reconstruction. Additionally, we expanded on the annotations proposed by task authors to tackle these tasks, highlighting the strengths and limitations of their approaches.

We then examined which additional knowledge (if any) is needed for the successful explicitation of arguments with implicit components. Our analysis focussed on two primary approaches to enthymeme explicitation: Relying solely on the local context and the information within the arguments themselves, and incorporating external knowledge to aid in reconstruction. For the latter, we discussed the integration of universally shared knowledge and domain-specific knowledge, tailored to the type and subject of the discourse. We also highlighted that neither approach to knowledge integration is sufficiently explored, with only a few studies addressing the use of universally shared knowledge and no research yet delving into the application of domain-specific knowledge for enthymeme explicitation.

Furthermore, we reviewed computational approaches to enthymeme detection and reconstruction, presenting the tasks as they are formulated for automatic methods. We described the development of models for automatic enthymeme analysis and attempted to evaluate their relative performance.

Subsequently, we examined the most representative and utilized datasets for exploring implicitness in argumentation. We categorized these datasets into three groups to facilitate their selection, visualized their annotations and types in a table, and discussed their respective advantages.

Finally, we discussed the future perspectives of implicitness studies in argumentation, particularly within the context of Argument Mining. Our findings emphasize the need for more extensive research into integrating diverse knowledge sources, utilizing advanced computational models tailored to the complexities of enthymeme analysis, and adopting versatile datasets to encompass the diverse domains where enthymemes may occur.

## Funding

The authors received the following financial support for the research, authorship, and/or publication of this article: This work has been supported by the French government, through the 3IA Cote d'Azur Investments in the project managed by the National Research Agency (ANR) with the reference number ANR-23-IACL-0001.

## Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## ORCID iDs

Ekatereina Sviridova  <https://orcid.org/0009-0008-0272-3911>

Elena Cabrio  <https://orcid.org/0000-0001-9374-7872>

Serena Villata  <https://orcid.org/0000-0003-3495-493X>

## References

1. Atkinson K, Baroni P, Giacomin M, et al. Towards artificial argumentation. *AI Maga* 2017; 38: 25–36.
2. Cabrio E and Villata S. Five years of argument mining: a data-driven analysis. In: *Proceedings of the twenty-seventh international joint conference on artificial intelligence, IJCAI-18*, 2018, pp.5427–5433. International Joint Conferences on Artificial Intelligence Organization.
3. Lawrence J and Reed C. Argument mining: a survey. *Comput Linguist* 2019; 45: 765–818.
4. Wachsmuth H, Naderi N, Hou Y, et al. Computational argumentation quality assessment in natural language. In: *Proceedings of the 15th conference of the European chapter of the Association for Computational Linguistics: volume 1, Long Papers*, (eds Lapata M, Blunsom P and Koller A), 2017, pp.176–187. Valencia, Spain: Association for Computational Linguistics, <https://aclanthology.org/E17-1017/>.
5. Wang X, Cabrio E and Villata S. Argument and counter-argument generation: a critical survey. In: *International conference on applications of natural language to information systems*, 2023, pp.500–510. Springer.
6. Lawrence J and Reed C. Argument mining using argumentation scheme structures. In: *COMMA*, 2016.
7. Lauscher A, Wachsmuth H, Gurevych I, et al. Scientia potentia est – on the role of knowledge in computational argumentation. *Trans Assoc Comput Linguist* 2022; 10: 1392–1422.

8. Walton D, Reed C and Macagno F. *Argumentation schemes*. Cambridge University Press, 2008.
9. Macagno F, Walton D and Reed C. Argumentation schemes. History, classifications, and computational applications. *J Log Appl* 2017; 4: 2493–2556.
10. Hulpus I, Kobbe J, Meilicke C, et al. Towards explaining natural language arguments with background knowledge. In: *PROFILES/SEMEX@ISWC*, 2019.
11. Rajendran P, Bollegala D and Parsons S. Contextual stance classification of opinions: a step towards enthymeme reconstruction in online reviews. In: *Proceedings of the third workshop on argument mining*, 2016, pp.31–39. Berlin, Germany: Association for Computational Linguistics.
12. Stahl M, Düsterhus N, Chen M-H, et al. Mind the gap: automated corpus creation for enthymeme detection and reconstruction in learner arguments. In: Bouamor H, Pino J and Bali K (eds) *Findings of the Association for Computational Linguistics: EMNLP 2023*. Singapore: Association for Computational Linguistics, 2023, pp.4703–4717.
13. Razuvayevskaya O and Teufel S. Finding enthymemes in real-world texts: a feasibility study. *Argum Comput* 2017; 8: 1–17.
14. Chakrabarty T, Trivedi A and Muresan S. Implicit premise generation with discourse-aware commonsense knowledge models. In: *Proceedings of the 2021 conference on empirical methods in natural language processing* (eds Moens M-F, Huang X, Specia L and Yih SW-t), 2021, pp.6247–6252. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.
15. Habernal I, Wachsmuth H, Gurevych I, et al. The argument reasoning comprehension task: identification and reconstruction of implicit warrants. In: *Proceedings of the 2018 conference of the NAACL, Volume 1 (Long Papers)*, 2018, pp.1930–1940. New Orleans, Louisiana: Association for Computational Linguistics.
16. Becker M, Korfhage K and Frank A. Implicit knowledge in argumentative texts: an annotated corpus. In: *Proceedings of the twelfth language resources and evaluation conference* (eds Calzolari N, Béchet F, Blache P, Choukri K, Cieri C, Declerck T, Goggi S, Isahara H, Maegaard B, Mariani J, Mazo H, Moreno A, Odijk J and Piperidis S), 2020, pp.2316–2324. Marseille, France: European Language Resources Association.
17. Cabrio E, Tonelli S and Villata S. From discourse analysis to argumentation schemes and back: relations and differences. In: Leite J, Son TC, Torroni P, van der Torre L and Woltran S (eds) *Computational logic in multi-agent systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp.1–17.
18. Peldszus A and Stede M. From argument diagrams to argumentation mining in texts: a survey. *Int J Cognit Inform Natl Intell* 2013; 7: 1–31.
19. Stede M, Afantenos S, Peldszus A, et al. Parallel discourse annotations on a corpus of short texts. In: *Proceedings of the tenth international conference on language resources and evaluation (LREC'16)*, 2016, pp.1051–1058. Portorož, Slovenia: ELRA.
20. Toulmin S. *The uses of argument*. Cambridge University Press, 2003.
21. Freeman J. *Argument structure: representation and theory*. Argumentation Library. Springer Netherlands, 2011.
22. Wang J. On freeman's argument structure approach. In: *Chinese conference on logic and argumentation*, 2016, <https://api.semanticscholar.org/CorpusID:8638296>.
23. Alshomary M, Syed S, Potthast M, et al. Target inference in argument conclusion generation. In: *Proceedings of the 58th Annual meeting of the Association for Computational Linguistics* (eds Jurafsky D, Chai J, Schluter N and Tetreault J), 2020, pp.4334–4345. Association for Computational Linguistics.
24. Singh K, Reiser P, Inoue N, et al. Improving evidence detection by leveraging warrants. In: *Proceedings of the second workshop on fact extraction and verification*, 2019, pp.57–62. Hong Kong, China.
25. Hidey C, Musi E, Hwang A, et al. Analyzing the semantic types of claims and premises in an online persuasive forum. In: *Proceedings of the 4th workshop on argument mining*, 2017, pp.11–21. Copenhagen, Denmark: Association for Computational Linguistics.
26. Freeman J. What types of statements are there? *Argumentation* 2000; 14: 135–157.
27. Keith W and Beard D. Toulmin's rhetorical logic: what's the warrant for warrants? *Philos Rhetoric* 2008; 41: 22–50.
28. Freeman JB. What types of arguments are there? *OSSA Conf Arch* 2013; 51: 1–15.
29. Grice P. Logic and conversation. *Synt Semant* 1975; 3: 41–58.
30. Perelman C and Olbrechts-Tyteca L. *The new rhetoric: a treatise on argumentation*. University of Notre Dame Press, 1991.
31. Lumer C and Dove IJ. Argument schemes – an epistemological approach. *OSSA Conf Archive* 2011; 17: 1–32.
32. van Eemeren F, Grootendorst R and Grootendorst R. *A systematic theory of argumentation: the pragma-dialectical approach*. Cambridge University Press, 2004.
33. Saadat-Yazdi A, Pan JZ and Kokciyan N. Uncovering implicit inferences for improved relational argument mining. In: *Proceedings of the 17th conference of the European chapter of the Association for Computational Linguistics* (eds Vlachos A and Augenstein I), 2023, pp.2484–2495. Dubrovnik, Croatia: Association for Computational Linguistics.
34. Becker M, Staniek M, Nastase V, et al. Enriching argumentative texts with implicit knowledge. In: *Natural language processing and information systems – 22nd international conference on applications of natural language to information systems, NLDB 2017*,

- Liège, Belgium, proceedings*, Volume 10260, *Lecture Notes in Computer Science* (eds Frasinca F, Ittoo A, Nguyen LM and Métais E), 2017, pp.84–96. Springer.
35. Boltužić F and Šnajder J. Fill the gap! Analyzing implicit premises between claims from online debates. In: *Proceedings of the third workshop on argument mining*, 2016, pp.124–133. Berlin, Germany: Association for Computational Linguistics.
  36. Wachsmuth H, Trenkmann M, Stein B, et al. A review corpus for argumentation analysis. In: *Conference on intelligent text processing and computational linguistics*, 2014, pp.115–127. Berlin, Heidelberg: Springer Berlin Heidelberg.
  37. Granger S, Dupont M, Meunier F, et al. *International corpus of learner English. Version 3*. Louvain-la-Neuve: Presses universitaires de Louvain, 2020.
  38. Bhagavatula C, Bras RL, Malaviya C, et al. Abductive commonsense reasoning. In: *8th International conference on learning representations, ICLR 2020*, 2020, Addis Ababa, Ethiopia. OpenReview.net.
  39. Peldszus A. An annotated corpus of argumentative microtexts. In: *First European conference on argumentation: argumentation and reasoned action*, 2015, Lisbon, Portugal.
  40. Choi H and Lee H. GIST at SemEval-2018 task 12: a network transferring inference knowledge to argument reasoning comprehension task. In: *Proceedings of the 12th international workshop on semantic evaluation*. (eds Apidianaki M, Mohammad S M, May J, Shutova E, Bethard S and Carpuat M), 2018, pp.773–777. New Orleans, Louisiana: Association for Computational Linguistics.
  41. Bowman SR, Angeli G, Potts C, et al. A large annotated corpus for learning natural language inference. In: *Proceedings of the 2015 conference on empirical methods in natural language processing* (eds Màrquez L, Callison-Burch C and Su J), 2015, pp.632–642. Lisbon, Portugal: Association for Computational Linguistics.
  42. Williams A, Nangia N and Bowman S. A broad-coverage challenge corpus for sentence understanding through inference. In: *Proceedings of the 2018 conference of the North American chapter of the Association for Computational Linguistics: human language technologies, Volume 1 (Long Papers)* (eds Walker M, Ji H and Stent A), 2018, pp.1112–1122. New Orleans, Louisiana: Association for Computational Linguistics. DOI: [10.18653/v1/N18-1101](https://doi.org/10.18653/v1/N18-1101). <https://aclanthology.org/N18-1101/>.
  43. Rinott R, Dankin L, Alzate Perez C, et al. Show me your evidence – an automatic method for context dependent evidence detection. In: *Proceedings of the 2015 conference on empirical methods in natural language processing* (eds Màrquez L, Callison-Burch C and Su J), 2015, pp.440–450. Lisbon, Portugal: Association for Computational Linguistics.
  44. Bar-Haim R, Bhattacharya I, Dinuzzo F, et al. Stance classification of context-dependent claims. In: *Proceedings of the 15th conference of the European chapter of the Association for Computational Linguistics: volume 1, Long Papers* (eds Lapata M, Blunsom P and Koller A), 2017, pp.251–261. Valencia, Spain: Association for Computational Linguistics.
  45. Wang L and Ling W. Neural network-based abstract generation for opinions and arguments. In: *Proceedings of the 2016 conference of the North American chapter of the Association for Computational Linguistics: human language technologies* (eds Knight K, Nenkova A and Rambow O), 2016, pp.47–57. San Diego, CA: ACL. <https://aclanthology.org/N16-1007>.
  46. Stab C and Gurevych I. Identifying argumentative discourse structures in persuasive essays. In: *Proceedings of the 2014 conference on empirical methods in natural language processing*, 2014, pp.46–56. Doha, Qatar.
  47. Wojatzki M and Zesch T. Stance-based argument mining – modeling implicit argumentation using stance. In: *Conference on natural language processing*, 2016. <https://api.semanticscholar.org/CorpusID:85555944>.
  48. Mohammad S, Kiritchenko S, Sobhani P, et al. SemEval-2016 task 6: detecting stance in tweets. In: *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)* (eds Bethard S, Carpuat M, Cer D, Jurgens D, Nakov P and Zesch T), 2016, pp.31–41. San Diego, California: Association for Computational Linguistics.
  49. Paul D, Opitz J, Becker M, et al. Argumentative relation classification with background knowledge. In: *COMMA*, 2020.
  50. Opitz J and Frank A. Dissecting content and context in argumentative relation analysis. In: *Proceedings of the 6th workshop on argument mining* (eds Stein B and Wachsmuth H), 2019, pp.25–34. Florence, Italy: Association for Computational Linguistics.
  51. Mestre R, Milicin R, Middleton SE, et al. M-arg: multimodal argument mining dataset for political debates with audio and transcripts. In: *Proceedings of the 8th workshop on argument mining*, 2021, pp.78–88. Punta Cana: Dominican Republic.
  52. Sviridova E, Yeginbergen A, Estarrona A, et al. CasiMedicos-arg: a medical question answering dataset annotated with explanatory argumentative structures. In: *Proceedings of the 2024 conference on empirical methods in natural language processing* (eds Al-Onaizan Y, Bansal M and Chen Y-N), 2024, pp.18463–18475. Miami, Florida, USA: Association for Computational Linguistics.
  53. Köhler S, Gargano M, Matentzoglou N, et al. The human phenotype ontology in 2021. *Nucleic Acids Res* 2021; 49: D1207–D1217.
  54. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004; 32: D267–D270.
  55. Vaswani A, Shazeer NM, Parmar N, et al. Attention is all you need. In: *Neural information processing systems (NIPS)*, 2017, Long Beach, CA, USA.
  56. Delas Z, Plüss B and Ruiz-Dolz R. An argumentation scheme-based framework for automatic reconstruction of natural language enthymemes. In: *COMMA, Frontiers in Artificial Intelligence and Applications*, 2024, pp.61–72.
  57. Speer R and Havasi C. Representing general relational knowledge in ConceptNet 5. In: *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)* (eds Calzolari N, Choukri K, Declerck T, Doğan MU, Maegaard B,

- Mariani J, Moreno A, Odijk J and Piperidis S), 2012, pp.3679–3686. Istanbul, Turkey: European Language Resources Association (ELRA).
58. Becker M, Liang S and Frank A. Reconstructing implicit knowledge with language models. In: *Proceedings of deep learning inside out (DeeLIO): the 2nd workshop on knowledge extraction and integration for deep learning architectures*. (eds Agirre E, Apidianaki M and Vulić I), 2021, pp.11–24. Association for Computational Linguistics.
  59. Tan C, Niculae V, Danescu-Niculescu-Mizil C, et al. Winning arguments: interaction dynamics and persuasion strategies in good-faith online discussions. In: *Proceedings of the 25th international conference on world wide web*, 2016.
  60. Kawarada M, Hirao T, Uchida W, et al. Argument mining as a text-to-text generation task. In: *Proceedings of the 18th conference of the European chapter of the Association for Computational Linguistics (Volume 1: Long Papers)* (eds Graham Y and Purver M), 2024, pp.2002–2014. St. Julian's, Malta: Association for Computational Linguistics.
  61. Alshomary M and Wachsmuth H. Conclusion-based counter-argument generation. In: *Proceedings of the 17th conference of the European chapter of the Association for Computational Linguistics*, (eds Vlachos A and Augenstein I), 2023, pp.957–967. Dubrovnik, Croatia: Association for Computational Linguistics.
  62. Chiang W, Zheng L, Sheng Y, et al. Chatbot arena: an open platform for evaluating llms by human preference. *CoRR* abs/2403.04132, 2024.