

Is There Anything More Deceptive Than an Obvious Fact? Investigating Implicitness in User-Generated Argumentative Text

Ekaterina Sviridova, Elena Cabrio, Serena Villata

Université Côte d'Azur, Inria, CNRS, I3S

Nice, France

sviridova@i3s.unice.fr, elena.cabrio@univ-cotedazur.fr, villata@i3s.unice.fr

Abstract

While various attempts towards unveiling implicitness in argumentation have been made, particularly towards improving automatic detection and reconstruction of implicit components and background knowledge, the task remains overly challenging. In this paper, we present, to the best of our knowledge, the first fine-grained typology of implicitness in argumentation, distinguishing among *implicature*, *ambiguity*, and *presupposition*. Applying this typology, we annotate 78 full-length discussions from the *Change My View* forum, building the largest publicly available dataset of real-world *enthymemes* with implicitness types labeled. For comparison, we additionally annotate 112 short argumentative texts from the *Microtext* corpus to examine how text length and complexity influence the automatic analysis of natural arguments. Leveraging these datasets, we establish strong baselines for two tasks: (i) enthymeme detection and (ii) fine-grained implicitness classification, with both encoder-only and large language models, highlighting the challenge of modeling implicit reasoning in long, unstructured discourse.

Keywords: Argumentation, Enthymeme

1. Introduction

Interest in comprehensive understanding of human argumentation has intensified, spurred in part by the progress of Argument Mining (AM) methods (Lawrence and Reed, 2019). Nevertheless, natural-language arguments are rarely complete, meaning that humans omit information that is presumed to be well known or not necessary for understanding. Though these unstated implicit elements are effortlessly recovered by humans, they remain difficult for current systems, which are constrained by limited reasoning capabilities (Choi and Lee, 2018; Saadat-Yazdi et al., 2023).

Despite various attempts to understand, automatically identify, and reconstruct incomplete arguments (referred to as *enthymemes* (Walton et al., 2008; Rajendran et al., 2016)), a gap in our understanding of the intrinsic characteristics of implicitness remains. While argument component detection has already been widely tackled across different domains in AM (Stab and Gurevych, 2017; Persing and Ng, 2020; Habernal et al., 2023; Sviridova et al., 2024), implicitness has mostly been examined in more restricted settings, e.g., selecting the component from two options (Choi and Lee, 2018), or predicting the implicit component or relation between explicit ones (Singh et al., 2019).

We still know little about whether implicitness functions similarly across different domains, whether distinct types of implicitness can be found in natural language arguments, and whether different reconstruction strategies are required for each. Hulpus et al. (2019) highlight that the purpose and the type of implicitness in argumentation may vary,

so should its reconstruction. Although possible categories of implicitness have already been outlined in pragmatics and formal argumentation (Grice, 1975; Sperber and Wilson, 1995; Levinson, 2000; Lombardi Vallauri et al., 2022), their behavior in real-world discourse remains underexplored. Accordingly, we address two main research questions:

- *what kinds of implicitness occur in natural language arguments, and how do they function within real-world discourse?*
- *to what extent can current language-model architectures automatically capture variations of implicitness in long, real-world argumentative discourse?*

We address these research questions providing two key contributions: (i) we introduce a novel, pragmatically motivated fine-grained typology of implicitness in argumentation, together with two new linguistic resources — one consisting of real-world argumentative discourse drawn from the Change My View forum (Tan et al., 2016), and the other comprising short, well-structured argumentative texts from the Microtext corpus (Peldszus and Stede, 2013) — both manually annotated with fine-grained implicitness types; and (ii) we establish strong baselines for enthymeme detection and fine-grained implicitness classification on both datasets, comparing encoder-based with large language models¹.

¹The annotated corpora, the accompanying annotation guidelines and the software are available at: <https://github.com/ekaterinasviridova4/Investigating-implicitness-in-user-generated-argumentative-text>

Our approach is the first, to the best of our knowledge, to address enthymeme detection without restrictions on text length, demonstrating that it is feasible to detect implicit argumentation in long, unstructured discourse. Moreover, we are the first to recognize that previously proposed strategies for implicitness detection may not generalize effectively to real-world discourse, where arguments tend to be longer and structurally more complex. In this context, we examine how argument length and complexity influence both the annotation process and automatic classification performance. These outcomes lay the groundwork for accurate, type-aware reconstruction of implicit content in natural language argumentation.

The paper unfolds as follows. Section 2 outlines existing research directions in the field. Section 3 presents the datasets selected for in-depth enthymeme analysis, introduces our fine-grained typology of implicitness, and details the annotation process through which we obtain a novel resource. Section 4 describes our experimental setup and baseline models, while Section 5 reports and discusses the empirical results. Finally, we conclude by summarizing our findings (Section 6) and elaborate on the current limitations (Section 7).

2. Related Work

In this section, we review existing research on implicitness in argumentation, covering both implicit argument components and implicit commonsense knowledge, and examine relevant datasets for mining such arguments. Despite growing interest, this line of research remains relatively limited, as understanding implicitness in argumentation requires extralinguistic knowledge (e.g., cultural differences, generation gaps), reasoning skills, domain expertise, and other competencies that algorithms either lack or possess only to a limited degree.

Broadly, two main research directions have emerged for addressing incomplete arguments, i.e., enthymemes. One focuses on analyzing argument components that are left implicit: implicit premises and/or warrants (Boltužić and Šnajder, 2016; Chakrabarty et al., 2021; Stahl et al., 2023; Habernal et al., 2018; Choi and Lee, 2018; Singh et al., 2019), or enthymemes involving implicit conclusions (Wojatzki and Zesch, 2016; Alshomary et al., 2020). In parallel, progress has been made in addressing the knowledge missing from arguments. Approaches aiming to transform enthymemes into fully explicit arguments suggest that incorporating external knowledge is essential for argument reconstruction. As such, Becker et al. (2017, 2020) leverage commonsense knowledge from ConceptNet (without explicitly addressing argument components), while Chakrabarty et al. (2021) recon-

struct premises using the discourse-aware commonsense knowledge model PARA-COMET. Other works focus on reconstructing warrants through knowledge transfer from external datasets (Choi and Lee, 2018), or on modeling argumentative relations using ConceptNet and the commonsense transformer COMET (Paul et al., 2020; Saadat-Yazdi et al., 2023).

Research on implicitness in argumentation has therefore evolved around two sequential tasks: **enthymeme detection** (e.g., identification of incomplete arguments) and **enthymeme reconstruction** (e.g., generative process to restore arguments, mostly via commonsense knowledge injection). Early studies on detection relied on manually annotated data capturing argument components (Boltužić and Šnajder, 2016; Hidey et al., 2017), reasons (Habernal et al., 2018), and commonsense relations (Becker et al., 2020). This stage was essential for demonstrating the feasibility of the detection task (Razuvayevskaya and Teufel, 2017) and for further development of automatic detection models, as high-quality annotated data are critical for both training and evaluation. In terms of computational methods for enthymemes detection, BERT-based models have proved particularly effective (Stahl et al., 2023; Delas et al., 2024). For enthymeme reconstruction, research similarly began with manually curated argument completions (Becker et al., 2020; Habernal et al., 2018), later advancing to sequence-to-sequence models such as BART (Chakrabarty et al., 2021; Stahl et al., 2023) and generative language models including COMET (Saadat-Yazdi et al., 2023), GPT-2, and XLNet (Becker et al., 2021).

Alongside the growing interest in enthymeme analysis, several datasets with high-quality annotations have been developed. Among standard Argument Mining resources annotated with argument components and relations, a few have been further enriched to capture implicit aspects of argumentation. As such, Alshomary et al. (2020) extended *Argument Essays V2* (Stab and Gurevych, 2014), *iDebate* (Wang and Ling, 2016) and *Claim Stance Dataset* (Bar-Haim et al., 2017) with targets of implicit conclusions. Similarly, the *Microtext corpus* (Peldszus and Stede, 2013; Becker et al., 2020; Chakrabarty et al., 2021) has been augmented with implicit knowledge and automatically generated premises, while Chakrabarty et al. (2021) also expanded *Online Debate Forum* (Boltužić and Šnajder, 2016) and *Room for Debate* (Habernal et al., 2018) with generated premises. Finally, the *Change My View (CMV)* dataset (Tan et al., 2016; Hidey et al., 2017) includes semantic clause-type annotations, offering a more discourse-oriented perspective. Although these datasets include an extensive variety of discourse types and topics, all but

one (the *CMV* dataset) consist of short argumentative texts (3–5 sentences long). This feature allows for better control over the detection and reconstruction of incomplete arguments but does not reflect how arguments occur in real-world contexts. In natural discourse, arguments rarely appear in isolation — they unfold as part of a broader conversation and intertwine with contextual cues. Therefore, it is essential to study arguments in their full discourse context, accounting for their natural complexity, ambiguity, and interdependence.

3. Implicitness Annotation

In the following, we introduce the two datasets selected for our study, we detail the typology of implicit arguments we propose, and we then discuss the annotation process and the obtained resources.

3.1. Data

Since the focus of the current study is to analyze complete, real-world argumentative discourse, we select the *Change My View* dataset, extracted from a discussion forum on reddit.com designed to encourage users to challenge one another's opinions on a wide range of topics (Tan et al., 2016). The dataset comprises 78 full civil discourse texts presented in dialogical format (annotated with argument components): the original poster introduces a topic by stating their opinion and providing supporting reasons, while other users attempt to challenge that viewpoint. If the opposition is successful, the original poster explicitly states the change of opinion with a delta symbol. To ensure clarity and facilitate analysis, the dataset has been preprocessed into dialogues, each involving one original poster and one respondent. This structure allows for a transparent analysis of reasoning patterns and opinion change within the flow of natural conversation. Each text is stored in a separate XML file, annotated with claims and premises using XML markup. Annotations follow the structure:

```
<premise type="logos_pathos" id = "3j"> the
belief eventually becomes the driving force and
incentive for the action in later generations.
</premise> <claim type="evaluation_rational"
id="3k"> You are probably right that cultural norms
are extremely influential in the creation of a religion
</claim>
```

To compare the challenges of analyzing long, real-world discourse with those of short, clearly structured argumentative texts, we additionally select the *Microtext corpus* (Peldszus and Stede, 2013) as a reference dataset. This corpus con-

sists of 112 short (3–5 sentence) argumentative texts, professionally translated from German into English, each annotated with argumentation structures, i.e., argument discourse units (ADUs) and their interrelations. The discussion topics in both datasets range widely — from politics to everyday matters.

So as to illustrate the complexity of full-text argument analysis, Figure 1 shows a relation graph of one *CMV* dialogue titled “*Religion is not violent or not violent, its followers are*”. In the graph, nodes represent argument components, Premise (P) and Claim (C), while edges denote argumentative relations (support or attack). To preserve the original sequence, we number argument components accordingly. For comparison, Figure 2 shows a graph from the *Microtext* corpus, in which nodes and edges similarly represent argument components and their relations. Due to space constraints, we do not include the textual content of the components in the graphs; however, the complete versions are available in the annotation guidelines².

This comparison reveals the complexity of full-text argument analysis. The structure of the *Microtext* corpus example is relatively transparent, with straightforward relationships between components confined to just five sentences. This limited scope makes it comparatively easier to identify what is implicit. By contrast, the *CMV* dialogue exhibits significantly greater complexity. A respondent may reference earlier statements or return to the initial claim (e.g., the discussion title) at the end of the exchange. This level of complexity requires an automatic system to retain the full context of the discussion and attend to all parts of the dialogue equally in order to accurately detect implicit content. Also, unlike short texts, full-text arguments often contain non-linear reasoning structures, such as rhetorical questions, or shifts in stance, which conceal the location and nature of implicit components. This makes it more difficult for both annotators and (hypothetically) models to trace argument flow and pinpoint what is left unsaid, especially when implicit content is distributed across distant parts of the text.

3.2. Annotation Guidelines

The necessity to characterize implicitness in argumentation for in-depth analysis and enthymeme reconstruction has been emphasized in prior work (Becker et al., 2020; Saadat-Yazdi et al., 2023). Since the enthymeme reconstruction process may differ depending on whether the implicit content is factual or subjective (Hulpus et al., 2019), we propose to explore a possible typology of implicitness

²<https://github.com/ekaterinasviridova4/Investigating-implicitness-in-user-generated-argumentative-text>

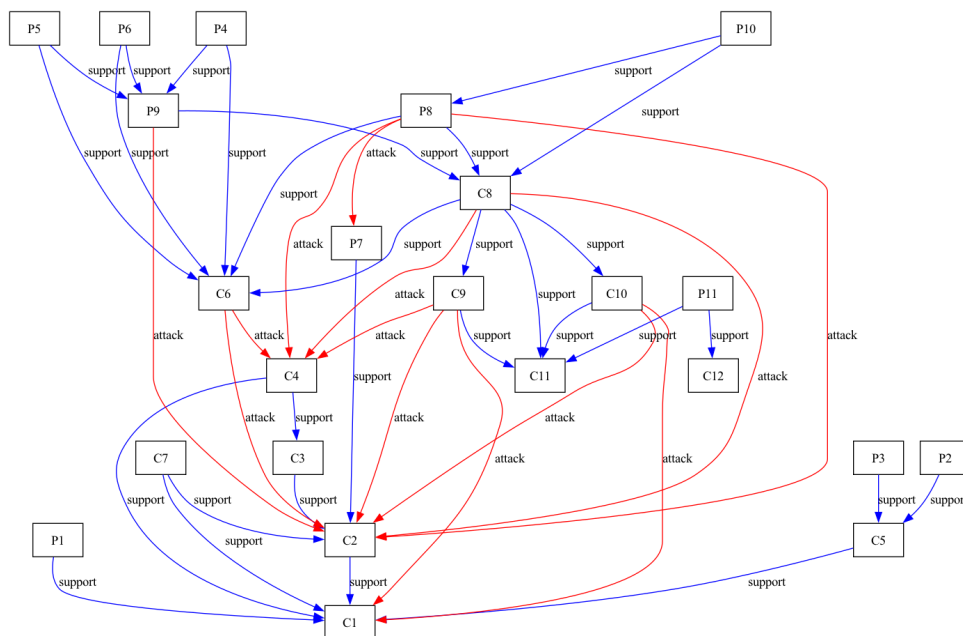


Figure 1: Relation graph of a CMV argumentative text.

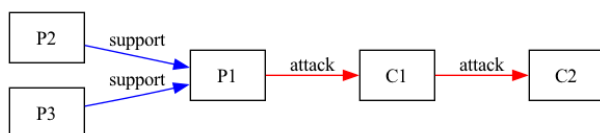


Figure 2: Relation graph of a Microtext argument.

in argumentation, examining the pragmatic dimension of implicit statements in argumentation, alongside computational approaches and philosophical perspectives. A common form of implicitness occurs when individuals intentionally omit information that is universally known or easily inferred from context, thus, reducing cognitive load for the audience. This principle aligns with Grice's Maxim of Quantity (Grice, 1975). Lombardi Vallauri (2016); Lombardi Vallauri et al. (2022) describe a related type of implicitness – implicatures. We adopt this term and extend its definition as follows: **Implicature** refers to content that reflects universal knowledge, or that can be inferred from the utterance and its context, though not explicitly stated. To illustrate this type, we provide a real example from the CMV dataset (title of the discussion: "Religion is not violent or not violent, its followers are"):

1. [**claim:** I point out that many Christians follow the bible which has numerous examples of sexism, but in application, there are numerous branches of Christianity that are no more sexist than secular groups.] [**premise:** For example, Congregationalists and Universalists.] So, [**claim:** my belief is that while religion can inform the views of people, it is far more likely

that religion will be used to justify actions that would have been executed any way.]

Example 1 contains implicitness in the premise, which implies that these groups are religious and they do convey sexism, but not more than secular ones. It is an **Implicature** because the intended meaning relies on contextual inference, though knowledge on Christians supports interpretation.

Another common type of implicitness involves vague or ambiguous expressions. Building on Lombardi Vallauri (2016); Lombardi Vallauri et al. (2022), and on Razuvayevskaya and Teufel (2017) that respectively describe these as expressions that require additional information or subjective interpretation, and as information the lack of which arouses cognitive interest of readers and creates ambiguity, we establish the following type of implicitness: **Ambiguity** refers to content that may denote multiple entities or states of affairs and cannot be resolved from context alone, requiring further precision. This type significantly differs from implicature by its main feature: a single meaning cannot be grasped from the context, and different kinds of audience may interpret the same text differently. In Example 2 (title: "Surplus Value Theorem is definitive proof that capitalism is an inefficient system" in CMV), while the broader context (partially omitted for the sake of space) provides relevant background, parts of this passage remain ambiguous: the phrase "value is created at each stage" does not have a single interpretation; therefore, we class it as **Ambiguity**.

2. [**claim:** This is one of those views that likes to view production, capital, labor, and all of economics as a zero sum game.] [**claim:** It's

not.] **[claim:** *Money circulates multiple times in a system, and value is created at each stage.*
] **[claim:** *I.e. you're simplifying it down to a level where it doesn't show all of the uses of the money that actually happen.]*

The third type, adapted from Lombardi Vallauri (2016), reflects the speaker's implicit assumption of shared understanding and can function as a persuasive device, as the speaker controls how information is framed: **Presupposition** refers to presenting a notion as already shared by the author and their addressee(s). Unlike implicatures or ambiguities, which partially obscure meaning – whether to minimize cognitive load or influence the audience - presuppositions convey information openly, assuming that it is already known and accepted. As a result, the likelihood of the content being questioned is reduced. Presuppositions can also reduce cognitive effort by omitting information presumed to be familiar. In Example 3 ("*Freedom of speech is being taken too far*", CMV):

3. **[claim** *I certainly value our free speech]* **but** *to me [claim there is a clear line between exercising your first amendment right ([premise "President Obama sucks!"]) and doing things that are known to be offensive to other cultures.] ([premise Satirical cartoons of prophets, assassinating leaders, etc.])*

alongside an **Implicature** related to the "First Amendment right," we also find a **Presupposition** embedded in the phrase "*President Obama sucks!*". The statement is phrased not to question its truthfulness, and presupposes that the Obama administration undermined freedom of speech and that these actions may have had serious consequences. Rather than implying that the speaker wants the addressee to adopt a belief, presuppositions create a communicative scenario in which the speaker assumes that the addressee already knows and agrees with the given information.

A significant challenge in annotating implicitness types is category overlap: a single span may have features of two categories. To ensure consistency and limit cognitive load, annotators are instructed to assign the most appropriate label to each span. Although multiple labels would allow for more nuanced analysis, they would further increase task complexity and complicate subsequent automatic processing. When a span of a different category is embedded within a larger one (e.g., a presupposition inserted in brackets within an ambiguous sentence), each segment has to be annotated with its respective label.

Non-implicit sentences are labeled as **Explicit**.

Although the implicitness categories proposed in our guidelines have roots in pragmatics and philosophy, the novelty of our work lies in its empirical

grounding and practical application. We systematically adapted and refined these categories for argumentative discourse, and to our knowledge, this is the first empirical study to apply them to full-text, real-world dialogues, and to compare their complexity with that of short-text, monologue arguments, thereby laying the groundwork for training and evaluating computational models.

3.3. Annotation Process and Statistics

The annotation process included four stages: *i)* during training, two expert-annotators studied the guidelines and annotated two full-text CMV samples and five Microtext samples with implicit and explicit labels, and as a second layer they annotated implicit instances with their finegrained categories; *ii)* at the reconciliation phase, the guidelines were discussed and refined with new examples for three implicitness labels; *iii)* after the reconciliation, the annotators received five full-text CMV arguments so as to calculate the preliminary IAA; *iv)* following this calibration, the annotators proceeded to the annotation of the entire datasets. Label Studio annotation tool was used for the labeling process³.

As an output of the annotation process of the CMV dataset, we produced two *.conll*-formatted files (one for implicit / explicit annotations and one for implicature / ambiguity / presupposition / explicit annotation). Each file contains 66463 lines, with individual argumentative texts separated by line breaks. Arguments are notably long, averaging 851 tokens per argument (ranging from 244 to 6663 tokens), which reflects the complexity of real-world discourse. The total count of tokens is 66386.

The annotation of the Microtext followed the same processing. However, given that the original data include only ADUs, in the case the annotators had to also label argument components (premises and claims). Therefore, we got an output of three *.conll*-formatted files. Each file contains 8869 lines, each argumentative text contains 78 tokens in average (ten times shorter than an average argument of the CMV), and the total token count being 8758 (seven times shorter than the CMV).

Table 1 shows the distribution of both standard argumentative components and our novel annotation of implicitness on the CMV and Microtext data. While the CMV dataset is moderate in size, it offers substantial depth: it provides full-length, dialogical arguments annotated for multiple types of implicitness. This makes it suited for exploring enthymemes in detail and testing automatic methods on real, non-simplified discourse. Importantly, the dataset is comparable in scale to prior resources used in enthymeme research: Argumentative Essays V2 (Stab and Gurevych, 2014) in-

³<https://labelstud.io>

cludes 402 short essays with 1552 components; Room for Debate (Habernal et al., 2018) contains 1654 argumentative triples (claim, premise, warrant). At the same time, the CMV dataset offers not only a competitive number of annotated elements, but also richer, long-form dialogical structures and a novel focus on implicitness. For comparison, the Microtext corpus comprises shorter, self-contained arguments characterized by clear argumentative structure. The dataset maintains a relatively balanced distribution between implicit (44%) and explicit (56%) instances. Given that this is a smaller dataset, certain categories remain underrepresented. Nevertheless, both datasets demonstrate a similar distribution trend among the implicitness types: Implicature is the most frequent, while Ambiguity remains the least represented. Despite differences in argument length and structural complexity, these two datasets together enable us to examine how model performance varies with text length and complexity, and to assess whether the same architectures can effectively classify both short and long argumentative texts.

3.4. Inter-Annotator Agreement (IAA)

The IAA has been calculated on a random batch of 10% of both datasets (i.e., 8 CMV conversations and 11 Microtext). For the Microtext corpus, we obtained substantial agreement, with Cohen’s $\kappa=0.71$ (exact matches) for the fine-grained classification task. To account for minor boundary mismatches, we also calculated token-level agreement, which allows partial overlap between annotated spans; this metric reached 0.85. For the CMV dataset, Cohen’s κ reached 0.52. While this result reflects only moderate agreement, several aspects of our annotation process must be considered. First, the task involves four non-obvious labels (Implicature, Ambiguity, Presupposition, Explicit) and requires sequence annotation on long texts, which inherently complicates the process. Moreover, Cohen’s κ measures strict instance-level agreement, counting only exact matches. Given the length of the annotated texts and the length of each span to annotate, annotators may disagree on which elements to include in a span. Thus, we also compute token-level agreement on CMV annotations. This metric accounts for 0.72, indicating a reasonable level of agreement for such a complex task.

Analyzing the agreement patterns across both datasets, we find that the majority of disagreements arise when fine-grained implicit instances are labeled as explicit by another annotator. In other words, the primary challenge lies in distinguishing between implicit and explicit content. At the same time, no strong systematic confusion is found among the implicitness types themselves.

4. Experimental Setup

In this section, we first describe the process of enthymeme detection using encoder-only models, and then compare their performance with that of generative large language models (LLMs). Although prior work tends to use encoder-based models for detection tasks due to the risk of hallucination in decoder models, we test this assumption by evaluating whether LLMs are indeed incapable of accurate detection. To the best of our knowledge, this is the first study to evaluate language model performance on enthymeme detection using full-length, real-world argumentative discourse, and to establish a baseline for the fine-grained classification of implicitness types.

To conduct an exhaustive analysis, we explore three levels of classification: (1) detection of argument components (claim and premise), (2) binary classification of implicit vs. explicit content (i.e., enthymeme detection), and (3) fine-grained classification of implicitness — Implicature, Ambiguity, Presupposition, and Explicit. As stated before, we test the three tasks on the two datasets previously described, i.e., CMV and Microtext corpus. The tasks are formulated as sequence labeling.

Sequence Labeling with Encoders. We experiment with BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019) on the three-level classification task using both the CMV dataset and the combined CMV + Microtext dataset. Given the length of CMV’s input texts, we also train Longformer, a transformer model specifically designed to handle long sequences (Beltagy et al., 2020). For Longformer, we conduct experiments following the same setup. The combination of two datasets has been necessary, as the Microtext alone is relatively small to train any model reliably. However, this setup allows us to test two hypotheses: (1) enthymemes and their argumentative structure are easier to recognize in shorter argumentative texts, and (2) our fine-grained classification approach generalizes well across datasets. To this end, we merge and shuffle both datasets using a random seed of 42 for each model and experimental setting. Given the input length limitations of BERT and RoBERTa (512 tokens), longer CMV texts are preprocessed into shorter segments of 400 tokens. This ensures that no tokens are lost during tokenization and that the data fit the models’ constraints.

Since encoder-based models operate directly on .conll files, no further modifications of the data are needed. Both the CMV dataset and the combined CMV + Microtext dataset are split into training (80%), validation (10%), and test (10%) sets. All models are trained for 8 epochs with a learning rate of $2e-5$. For consistency, each experiment is run three times with different random seeds, and

Labels	CMV			Microtext		
	# Tokens	# Sequences	%	# Tokens	# Sequences	%
Premise	28008	1591	58%	6303	416	73%
Claim	18137	1171	42%	2107	155	27%
Implicit	23531	1035	37%	4065	249	44%
Explicit	36416	1781	63%	4690	316	56%
Implicature	11737	515	19%	1800	112	20%
Ambiguity	5252	234	8%	435	30	5%
Presupposition	6542	286	10%	1830	107	19%
Explicit	36416	1781	63%	4690	316	56%

Table 1: Label distribution over the entire CMV and Microtext datasets on the token level (number of tokens) and the sequence level (number of sequences).

Model	AC		Enthymeme		Fine-grained	
	CMV	CMV+Mic.	CMV	CMV+Mic.	CMV	CMV+Mic.
BERT	0.77 (0.02)	0.75 (0.01)	0.77 (0.01)	0.75 (0.01)	0.71 (0.04)	0.56 (0.02)
RoBERTa	0.78 (0.03)	0.79 (0.02)	0.78 (0.02)	0.62 (0.04)	0.67 (0.04)	0.62 (0.03)
Longformer	0.72 (0.01)	0.75 (0.01)	0.75 (0.02)	0.79 (0.01)	0.56 (0.06)	0.58 (0.02)
	CMV	Mic.	CMV	Mic.	CMV	Mic.
Mistral 7B zero-shot	0.28	0.70	0.17	0.42	0.10	0.23
Llama 8B zero-shot	0.42	0.76	0.23	0.46	0.16	0.18
Mistral 24B zero-shot	0.43	0.76	0.24	0.46	0.18	0.23
GPT-4o zero-shot	0.60	0.65	0.55	0.56	0.25	0.27
	CMV	CMV+Mic.	CMV	CMV+Mic.	CMV	CMV+Mic.
Mistral 7B fine-tune	0.30 J / 0.35 B	0.29 J / 0.34 B	0.26 J / 0.30 B	0.22 J / 0.24 B	0.10 J / 0.13 B	0.12 J / 0.14 B
Llama 8B fine-tune	0.30 J / 0.34 B	0.26 J / 0.29 B	0.31 J / 0.34 B	0.29 J / 0.32 B	0.13 J / 0.15 B	0.16 J / 0.17 B
Mistral 24B fine-tune	0.39 J / 0.42 B	0.32 J / 0.36 B	0.36 J / 0.38 B	0.29 J / 0.32 B	0.13 J / 0.14 B	0.14 J / 0.15 B

Table 2: Macro-avg F1-scores for (1) argument component (AC) detection, (2) enthymeme detection, and (3) fine-grained implicitness classification for CMV vs CMV + Microtext and CMV vs Microtext settings. Standard dev. is reported for BERT, RoBERTa and Longformer in brackets. Fine-tuned models' results are reported according to Jaccard (J) and BERT (B) similarity scores. **Bold** indicates best overall result.

we report the mean and standard deviation of performance scores. All experiments are conducted using an NVIDIA A100-SXM4-40GB GPU.

Sequence Labeling with LLMs. We experiment with Mistral 7B (Jiang et al., 2023), Llama 3 8B (Grattafiori et al., 2024), Mistral Small 3 24B ⁴ and GPT-4o (Achiam et al., 2024). All models except GPT-4o are evaluated on our three-level detection task under two experimental settings: zero-shot and fine-tuning. GPT-4o is tested only in the zero-shot setting. Few-shot prompting was excluded from the experiments, as it proved to be an ineffective strategy: across all detection levels it altered the F1 score by no more than 0.02. Models are fine-tuned for three epochs using LoRA (Low-Rank Adaptation) to reduce the number of trainable parameters and prevent overfitting. The experiments are conducted on an NVIDIA H100-94GB GPU.

Zero-shot experiments allow us to evaluate LLMs on CMV and Microtext separately, regardless of Microtext's limited size, thus directly testing our hypothesis about the impact of text length on model

performance. Fine-tuning experiments mirror the setup of the encoders: models are trained on CMV alone as well as on the combined CMV + Microtext dataset, enabling a consistent comparison across architectures.

It is also important to note that some preprocessing is required before feeding data to the LLMs. While encoder-based models allow for .conll files as an input, LLMs require natural language input. Therefore, we convert the datasets into textual form where either *i*) each sentence is enclosed within tags (e.g., <Implicit> sentence </Implicit>) or *ii*) each sentence is followed by a JSON-style label (e.g., "sentence": "...", "type": "Implicit"). Both formats are tested, and prompts for the task are constructed accordingly. To mitigate hallucinations and improve generation quality, we experimented with multiple prompt formats and varying levels of detail. The best-performing prompts are provided in Appendix A.

⁴<https://mistral.ai/news/mistral-small-3-1>

5. Results and Error Analysis

Table 2 reports on the obtained results. Among encoder-only models, RoBERTa achieves the highest performance across most experimental settings. When comparing results on CMV alone to those on the combined CMV + Microtext dataset, we observe a slight performance drop for BERT and RoBERTa, while Longformer shows a modest improvement when the Microtext data are included. This variability can be attributed to the structural differences between the two datasets: although Microtext contains shorter and more straightforward arguments, it is considerably smaller in size, making its examples underrepresented and limiting their overall influence on model learning. Since adding Microtext to the training data increases imbalance for the enthymeme and fine-grained classification tasks, we experimented with downsampling the majority class. However, given that this adjustment yielded no meaningful improvement, thus, we kept our original data distribution.

Regarding task complexity, we observe that the encoders perform well on enthymeme detection, although this task appears more challenging than argument component classification. We also observe that the models succeed in capturing distinct characteristics of each implicitness category, achieving relatively strong performance under the fine-grained (four-label) classification setting. Table 3 presents the performance of encoder models across the individual implicitness subclasses on the CMV dataset and the Microtext corpus. Among these, Implicature, being the most frequent category, achieves the highest scores across all three models, while Presupposition emerges as the most difficult type to detect. The reason for that is the linguistic nature of presupposition: it conveys information that is not hidden but rather presupposed, taken for granted as part of the common ground between interlocutors. Consequently, such information seems to be expressed explicitly, making the boundary between explicit and presupposed content particularly prone to confusion. Furthermore, when examining model confusion patterns, we observe that misclassifications most often involve implicit instances and their subtypes being labeled as explicit, which mirrors the analysis of annotator disagreement.

Compared to the encoder-based models, LLMs' performance is lower across all settings and tasks. On the CMV dataset, results for binary classification (AC and enthymeme detection) are close to random, with GPT-4o performing better, but still falling short of the encoder baselines. All models achieve higher scores in zero-shot settings on the Microtext corpus than on CMV, indicating that shorter and more structurally constrained arguments are easier

for LLMs to process and classify. Nevertheless, zero-shot results on the fine-grained Microtext task remain far below those of the encoders. Fine-tuning does not yield any significant improvement across settings, leading to even lower performance compared to zero-shot evaluation in some cases. We test various prompts for all the tasks to obtain more reliable results. Finally, even argument component detection, which is the simplest of the three tasks, appears challenging for LLMs when applied to long, real-world argumentative texts. Therefore, we do not report performance of LLMs across implicitness types as we do for the encoder-based models, as these results do not offer any meaningful insights.

To better understand the low performance of LLMs, we first apply different similarity metrics to mitigate mismatches between model outputs and gold data. We test Jaccard and BERT-based similarity measures. Although the scores vary slightly across metrics, the overall trends remain consistent (see Table 2). Second, we manually inspect all model outputs for formatting and content quality. While no formatting errors are detected, several content-related issues persist. Some models, particularly Mistral and Llama, tend to output correctly structured outputs with empty content fields, either partially (toward the end of a text) or entirely for some samples. This means that the models fail to generate any text that can be evaluated against the gold annotations. Moreover, careful inspection reveals hallucinations in well-formed outputs, e.g., assigning the same label to every sentence. As for the GPT, it often corrects orthographical mistakes present in original texts. All the issues negatively impact evaluation results. Several examples of LLM outputs are provided in Appendix B.

Model	Implicature		Ambiguity		Presupposition	
	CMV	Mic.	CMV	Mic.	CMV	Mic.
BERT	0.69	0.49	0.64	0.37	0.55	0.44
RoBERTa	0.62	0.52	0.52	0.45	0.53	0.41
Longformer	0.50	0.49	0.47	0.39	0.35	0.25

Table 3: Macro-avg F1-scores for fine-grained types of implicitness on CMV and CMV + Microtext.

Observing the obtained results overall, enthymeme detection appears to be a computationally feasible task. Despite its inherent complexity and subjectivity, encoder-only models demonstrate the ability to identify implicit and explicit elements in real-world argumentative discourse to a reasonable extent. Additionally, the argumentative texts in our datasets cover a wide range of topics, from gaming to politics, which increases the task difficulty. As for the performance of LLMs on the classification task, the results are not promising, but not entirely surprising either. Although LLMs have been successfully applied to other classification tasks

(e.g., AC detection in medical domain (Sviridova et al., 2024)), they struggle to distinguish between implicit and explicit in long contexts and to understand the nuances of implicitness across different text lengths.

Finally, our hypothesis that shorter texts with more straightforward argumentative structures are easier to analyze is confirmed, particularly in zero-shot settings, where we could directly compare model performance on long, real-world CMV discussions versus short, well-structured arguments from the Microtext corpus.

6. Conclusion

This paper presents a novel fine-grained linguistic resource annotated for implicit content in real-world argumentative discourse, enabling an in-depth analysis of enthymemes. The annotations provide a foundation for identifying specific types of implicitness, each of which shapes the language and strategy required for effective reconstruction. Annotations were carried on two datasets, the CMV dataset containing 78 full-text dialogues and the Microtext corpus containing 112 self-contained argumentative texts. While testing encoder-models and LLMs, we set a solid baseline for enthymeme detection in extended settings, evaluating models on real-world discussions covering a wide range of topics. Comparing model performance on long, complex discourse to that on shorter arguments confirms that the automatic analysis of extended argumentative texts remains a major challenge, whereas the processing of short, well-structured arguments is already feasible. Although challenging, the proposed task is essential for advancing a deeper understanding of how humans convey and interpret implicit meaning in arguments. Progress in this direction not only contributes to the study of argument mining and reasoning, but also lays the foundation for developing computational systems capable of recognizing and reconstructing hidden argumentative content. These findings point toward a promising direction for future research on enhancing models' ability to reason over implicit content in complex discourse.

7. Limitations

We identify three main limitations in our work, that should be addressed in future research. The first is related to the data quantity. High-quality datasets of full-text argumentative discourse are extremely rare, yet they are essential for the automatic analysis of real-world argumentation. The dataset that we used in this study, Change My View, is one of the very few of its kind, but it remains relatively small for more extensive investigation.

Second, current language models are still far from a full understanding of nuanced implicit meanings in argumentative discourse. Although some models already demonstrate promising results (BERT, RoBERTa), their performance is still influenced by various factors such as input length, data quality, domain diversity, and data structure. As for LLMs, their main limitation is the difficulty to generate an entire text and without any corrections. In case of GPT, the model always tries to correct mistakes that are inherent to natural language discourse and should be kept to save the original data. As for Mistral and Llama, their pitfall is early stopping. Even using specific settings to avoid early stopping we could only mitigate the problem.

The last limitation is the subjectivity of the task, even though we achieved satisfiable inter-annotator agreement for such a challenging task. While we cannot fully explain how language models interpret implicitness or which internal logic guide their predictions, studying implicitness and its possible types remains crucial for achieving a deeper understanding of human discourse and enabling accurate enthymeme reconstruction.

Acknowledgments

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. Experiments presented in this paper were carried out using the Grid'5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations.

8. Bibliographical References

- Josh Achiam, Steven Adler, and OpenAI. 2024. [Gpt-4 technical report](#). ArXiv:2303.08774 [cs.CL].
- Milad Alshomary, Shahbaz Syed, Martin Potthast, and Henning Wachsmuth. 2020. [Target inference in argument conclusion generation](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4334–4345, Online. Association for Computational Linguistics.
- Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. 2017. [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*, pages 251–261, Valencia, Spain. Association for Computational Linguistics.
- Maria Becker, Katharina Korfhage, and Anette Frank. 2020. [Implicit knowledge in argumentative texts: An annotated corpus](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2316–2324, Marseille, France. European Language Resources Association.
- Maria Becker, Siting Liang, and Anette Frank. 2021. [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*, pages 11–24, Online. Association for Computational Linguistics.
- Maria Becker, Michael Staniek, Vivi Nastase, and Anette Frank. 2017. [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, Proceedings*, volume 10260 of *Lecture Notes in Computer Science*, pages 84–96, Liège, Belgium. Springer.
- Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. [Longformer: The long-document transformer](#).
- Filip Boltužić and Jan Šnajder. 2016. [Fill the gap! analyzing implicit premises between claims from online debates](#). In *Proceedings of the Third Workshop on Argument Mining*, pages 124–133, Berlin, Germany. Association for Computational Linguistics.
- Tuhin Chakrabarty, Aadit Trivedi, and Smaranda Muresan. 2021. [Implicit premise generation with discourse-aware commonsense knowledge models](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6247–6252, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- HongSeok Choi and Hyunju Lee. 2018. [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*, pages 773–777, New Orleans, Louisiana. Association for Computational Linguistics.
- Zvonimir Delas, Brian Plüss, and Ramon Ruiz-Dolz. 2024. [An argumentation scheme-based framework for automatic reconstruction of natural language enthymemes](#). In *COMMA*, Frontiers in Artificial Intelligence and Applications, pages 61–72.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, et al. 2024. [The llama 3 herd of models](#). ArXiv:2407.21783 [cs.AI].
- Paul Grice. 1975. Logic and conversation. *Syntax and Semantics*, 3: Speech Acts:41–58.
- Ivan Habernal, Daniel Faber, Nicola Recchia, Sebastian Bretthauer, Iryna Gurevych, Indra Spiecker genannt Döhmann, and Christoph Burchard. 2023. [Mining legal arguments in court decisions](#). *Artificial Intelligence and Law*, 32(3):1–38.
- Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. 2018. [The argument reasoning comprehension task: Identification and reconstruction of implicit warrants](#). In *Proceedings of the 2018 Conference of the NAACL, Volume 1 (Long Papers)*, pages 1930–1940, New Orleans, Louisiana. Association for Computational Linguistics.
- Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown. 2017. [Analyzing the semantic types of claims and premises in an online persuasive forum](#). In *Proceedings of the 4th Workshop on Argument Mining*, pages 11–21, Copenhagen, Denmark. Association for Computational Linguistics.
- Ioana Hulpus, Jonathan Kobbe, Christian Meilicke, Heiner Stuckenschmidt, Maria Becker, Juri Opitz, Vivi Nastase, and Anette Frank. 2019. [Towards explaining natural language arguments with background knowledge](#). In *PROFILES/SEMEX@ISWC*.
- Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.
- John Lawrence and Chris Reed. 2019. [Argument mining: A survey](#). *Computational Linguistics*, 45(4):765–818.

- Stephen C. Levinson. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Language, Speech, And Communication. Bradford Books.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. *Roberta: A robustly optimized bert pre-training approach*.
- Edoardo Lombardi Vallauri. 2016. *The “exaptation” of linguistic implicit strategies*. *SpringerPlus*, 5.
- Edoardo Lombardi Vallauri, Federica Cominetti, and Viviana Masia. 2022. *The persuasive and manipulative power of implicit communication*. *Journal of Pragmatics*, 197:1–7.
- Debjit Paul, Juri Opitz, Maria Becker, Jonathan Kobbe, Graeme Hirst, and Anette Frank. 2020. *Argumentative relation classification with background knowledge*. In *COMMA*.
- Andreas Peldszus and Manfred Stede. 2013. *From argument diagrams to argumentation mining in texts: A survey*. *International Journal of Cognitive Informatics and Natural Intelligence*, 7:1–31.
- Isaac Persing and Vincent Ng. 2020. *Unsupervised argumentation mining in student essays*. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6795–6803, Marseille, France. European Language Resources Association.
- Pavithra Rajendran, Danushka Bollegala, and Simon Parsons. 2016. *Contextual stance classification of opinions: A step towards enthymeme reconstruction in online reviews*. In *Proceedings of the Third Workshop on Argument Mining*, pages 31–39, Berlin, Germany. Association for Computational Linguistics.
- Olesya Razuvayevskaya and Simone Teufel. 2017. *Finding enthymemes in real-world texts: A feasibility study*. *Argument Computation*, 8:1–17.
- Ameer Saadat-Yazdi, Jeff Z. Pan, and Nadin Kocciyan. 2023. *Uncovering implicit inferences for improved relational argument mining*. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2484–2495, Dubrovnik, Croatia. Association for Computational Linguistics.
- Keshav Singh, Paul Reisert, Naoya Inoue, Pride Kavumba, and Kentaro Inui. 2019. *Improving evidence detection by leveraging warrants*. In *Proceedings of the Second Workshop on Fact Extraction and VERification*, pages 57–62, Hong Kong, China.
- Dan Sperber and Deirdre Wilson. 1995. *Relevance: Communication and Cognition*. Blackwell Publishers.
- Christian Stab and Iryna Gurevych. 2014. *Identifying argumentative discourse structures in persuasive essays*. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 46–56, Doha, Qatar.
- Christian Stab and Iryna Gurevych. 2017. *Parsing argumentation structures in persuasive essays*. *Computational Linguistics*, 43(3):619–659.
- Maja Stahl, Nick Düsterhus, Mei-Hua Chen, and Henning Wachsmuth. 2023. *Mind the gap: Automated corpus creation for enthymeme detection and reconstruction in learner arguments*. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 4703–4717, Singapore. Association for Computational Linguistics.
- Ekaterina Sviridova, Anar Yeginbergen, Ainara Estarrona, Elena Cabrio, Serena Villata, and Rodrigo Agerri. 2024. *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*, pages 18463–18475, Miami, Florida, USA. Association for Computational Linguistics.
- Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. *Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions*. *Proceedings of the 25th International Conference on World Wide Web*.
- Douglas Walton, Chris Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press.
- Lu Wang and Wang Ling. 2016. *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*, pages 47–57, San Diego, California. ACL.
- Michael Wojatzki and Torsten Zesch. 2016. *Stance-based argument mining - modeling implicit argumentation using stance*. In *Conference on Natural Language Processing*.

A. Best Prompts for LLMs in Zero-shot and Fine-tuning Settings

Prompt 1

""Your task is to analyze each sentence of the text and determine whether it is *Implicit* or *Explicit*.

Explicit refers to transparent and clearly understandable content.

Implicit refers to hidden meanings or assumptions that are unclear from the given text alone.

The text surrounding the sentence is your context.

Instructions:

- For each Implicit or Explicit sentence found, return a separate JSON object with exactly two fields:

- "sentence": the exact span from the input sentence expressing the argument.

- "type": "Explicit" or "Implicit".

- If none of them is found, return one JSON object with both fields set to "".

- Do **not** wrap the JSON objects in a list (no square brackets).

- Separate multiple JSON objects with commas and spaces only, e.g.:

"sentence": "...", "type": "Implicit", "sentence": "...", "type": "Explicit"

- The output must be strictly valid JSON:

- Use double quotes only

- Close all braces correctly

- Do not include trailing commas

- Do not include any explanation, notes, or extra text. Output only the JSON objects.

Sentence:

{sentence}

""

Prompt 2

""Your task is to classify each sentence in the following text as Implicit or Explicit.

Definitions:

- Explicit refers to transparent and clearly understandable content.

- Implicit refers to hidden meanings or assumptions that are unclear from the given text alone.

Instructions:

- Wrap each sentence in either <Implicit> sentence </Implicit> or <Explicit> sentence </Explicit> tags based on the classification.

- Output only the tagged text, with no explanations or extra formatting.

Sentence:

{sentence}

""

B. Some low-quality outputs of LLMs

Truncated Output

{"input": "title CMV: People who cheat on their spouses can and should, in some cases, be forgiven. id= "cn8tuyn" author= "Li54" There seems to be a commonly held belief that people who cheat on their SO / partner should be dropped immediately with no further contact. I understand that physical / sexual conduct with another person can seem like a betrayal, but in some cases, there are many worse things that can happen in a relationship, which couples seem to be able to work through. There are a few assumptions underlying the belief that " sex with another person = relationship over " that I don't think hold up. * Sex is the most important part of a relationship, and therefore, betraying it and that agreement mean the relationship should end. My belief: sex is a very important part of a relationship, but not *the* most important part. Immediately breaking up with someone because of something like this seems excessive. * Nobody makes mistakes ever, so when they do, the relationship should end. My belief: obviously this isn't true, and a cornerstone of strong relationships is the ability to work through mistakes / issues together. * Cheaters will always cheat again, so better to end it now. my view: That may be true in some cases, but other than anecdotal evidence, I haven't seen anything to support this claim. [But would be really interested in reading anything scientific about this, actually!] * Cheaters cheat because of underlying issues with the relationship. My belief: if that 's true, there are a lot of possibilities - not just breaking up - that could fix the problem, such as ... talking about the problem. Maybe fixing that issue will resolve the motivation to cheat. Caveat: sometimes, cheating happens because there are parts of the relationship that are fundamentally broken and / or the relationship is too far gone / unsalvageable and / or this is the straw the broke the camel 's back etc. Cheating shouldn't always signify an immediate end to a relationship. In some cases, couples can and should choose to work through it. CMV!

reply id= "cn8tm9p" author= "looklistencreate" I don't think this depends as much on the cheater as the person he / she cheated on. Some people are unable to stay with an unfaithful partner and should not be pressured into continuing a relationship after that point. Forgiveness depends on the forger,

not the perpetrator. Nobody "should" be forgiven for cheating, and I'm comfortable saying that anyone who knowingly betrays the trust of a relationship would deserve losing that relationship. I wouldn't go so far as to say that it's expected to break up with the cheater, because some people are willing to be forgiving. But I can't envision a situation where cheating "should" be forgiven. Forgiveness is a gift, not an expectation.

reply id="cn8tuyn" author="Li54" This makes the most sense to me - very coherent argument. Thanks!",

"**mistral_output**": {"sentence": "In some cases, couples can and should choose to work through it.", "type": "Explicit"}, {"sentence": "My belief: sex is a very important part of a relationship, but not the most important part.", "type": "Implicit"}, {"sentence": "Nobody 'should' be forgiven for cheating.", "type": "Explicit"}, {"sentence": "Forgiveness depends on the forgiver, not the perpetrator.", "type": "Explicit"} }

Label Hallucinations

"pred": "<CLAIM> There have been 55 million abortions in the US since 1973 </CLAIM> <CLAIM> Something in the ballpark of 45 % of women who have abortions have more than one abortion </CLAIM> <CLAIM> These people would have been raised by their incompetent parents to drain down society, increase crime rates, suck up resources, and generally screw things up </CLAIM> <CLAIM> One possibility is that there would be wars waged to try to kill them all, perhaps even with other countries with similar problems </CLAIM> <CLAIM> Waging war to purposefully lower population, even with a country facing similar issues, would cost not only resources, but it would also cause political, cultural, and global issues </CLAIM> <CLAIM> America's innovation and education rankings would be lower </CLAIM> <CLAIM> Every middle-class child in the country would receive a lower quality education if there were 55 million more around </CLAIM> <CLAIM> Without the extra population from abortions, has there been proportionately more immigration to fill labor needs, or are the millions of illegals from Mexico as likely to be so numerous even if there was a higher non-abortion population </CLAIM> <CLAIM> One possibility is that there would be wars waged to try to kill them all, perhaps even with other countries with similar problems </CLAIM> <CLAIM> America's innovation and education rankings would be lower </CLAIM> <CLAIM> Every middle-class child in the country would receive a lower quality education if there were 55 million more around </CLAIM> <CLAIM> It's debatable how many women are having abortions due to financial reasons </CLAIM>

<CLAIM> surely its a significant number </CLAIM> <CLAIM> how many of these women wouldn't be having financial troubles if there wasn't competition from cheap illegal labor to keep wages lower </CLAIM> <CLAIM> This holds true even if illegal immigration is an overall plus to the countries value </CLAIM> <CLAIM> Mexicans are the supreme race, native Americans will go extinct through abortion and then out-breeding. </CLAIM> <CLAIM> this discussion is abstract enough to go into some weird places and that it is necessary to have a stopping point </CLAIM> <CLAIM> Is a more highly populated and economically productive America today going to result in an America tomorrow that can handle over population problems </CLAIM> <CLAIM> having a higher population from non-abortions just add on to the problems of a world going to shit </CLAIM> <CLAIM> An aborted fetus is likely to have been a less productive person than a never-considered-for-abortion fetus in an industrial society </CLAIM> <CLAIM> its likely that these abortions would still be more productive than not </CLAIM> <CLAIM> This is to be considered if one is convinced that having a higher population coming from non-abortions results in a strain for resources in the country </CLAIM> <CLAIM> Modern technology like the internet must be considered </CLAIM> <CLAIM> Do you really think abortions are only had by incompetent people who couldn't be good parents? </CLAIM> <CLAIM> This appears to be a belief that has led you to conclude that America must be better off </CLAIM> <CLAIM> you are really misinformed about the type of women who obtain abortions </CLAIM> <CLAIM> That may be true. </CLAIM> <CLAIM> My thinking was that abortions were mainly required by unfortunate people, who in turn from their difficult circumstances, would've raised the child to be the same. </CLAIM> <CLAIM> Can you counter that? </CLAIM> <CLAIM> I have a [source] that shows that abortions are typically (and in some years, mostly) had by people without enough wealth to care for there kids. </CLAIM> <CLAIM> Some stats for example include 31 % of women obtaining abortions are in school; 68 % of women who have abortions are employed. </CLAIM> <CLAIM> 33 % of women who obtain abortions have family incomes under \$ 11,000 annually, 11 % of abortions are obtained by women whose household incomes are \$ 50,000 or more. </CLAIM> <CLAIM> that means the majority of those having abortions have income between \$ 11,000 and \$ 55,000, which is above the Federal Poverty line. </CLAIM> <CLAIM> I can't confirm but I do recall reading once that a common reason for women to abort is becoming pregnant too soon after having a child </CLAIM> <CLAIM> It could be about finances but could just be too soon for a woman to

go through again. </CLAIM> <CLAIM> Pregnancy is hard on the body and far riskier than an abortion performed within a few weeks. </CLAIM>",

"ref": "<premise> There have been 55 million abortions in the US since 1973 </premise> <premise> Something in the ballpark of 45 % of women who have abortions have more than one abortion </premise> <claim> These people would have been raised by their incompetent parents to drain down society, increase crime rates, suck up resources, and generally screw things up </claim> <claim> Various ways to lower a resource negative population would have to be explored, if not because of this 55 million, then because of the next 55 million </claim> <claim> One possibility is that there would be wars waged to try to kill them all, perhaps even with other countries with similar problems </claim> <premise> Waging war to purposefully lower population, even with a country facing similar issues, would cost not only resources, but it would also cause political, cultural, and global issues </premise> <premise> America's innovation and education rankings would be lower </premise> <premise> Every middle-class child in the country would receive a lower quality education if there were 55 million more around </premise> <claim> Its a complicated issue </claim> <premise> Without the extra population from abortions, has there been proportionately more immigration to fill labor needs, or are the millions of illegals from Mexico as likely to be so numerous even if there was a higher none-abortion population </premise> <premise> Doesn't immigration of a working class citizen on such a massive scale cause dissonance in a country more so than it does cultural exchange </premise> <claim> it would have been preferable to have our abortions alive </claim> <premise> Or are we talking about a population so huge that race and national identity are insignificant, that there is always going to be a huge amount of hostility between different demographics </premise> <claim> Perhaps the immigration of illegals is lowering the value of blue collar work and makes it harder for the borderline impoverished citizens to provide for their children </claim> <claim> Its debatable how many women are having abortions due to financial reasons </claim> <premise> surely its a significant number </premise> <premise> how many of these women wouldn't be having financial troubles if there wasn't competition from cheap illegal labor to keep wages lower </premise> <claim> This holds true even if illegal immigration is an overall plus to the countries value </claim> <claim> Mexicans are the supreme race, native Americans will go extinct through abortion and then out-breeding. </claim> <claim> this discussion is abstract enough to go into some weird places and that it is necessary to have a stopping point </claim> <claim> Is

a more highly populated and economically productive America today going to result in an America tomorrow that can handle over population problems </claim> <claim> would having a higher population from none-abortion just add on to the problems of a world going to shit </claim> <premise> An aborted fetus is likely to have been a less productive person than a never-considered-for-abortion fetus </premise> <claim> in an industrial society, its likely that these abortions would still be more productive than not </claim> <claim> Then there was the debate as to whether or not a larger population with less resources per person is more innovative than a smaller one </claim> <claim> This is to be considered if one is convinced that having a higher population coming from none-abortions results in a strain for resources in the country </claim> <claim> Modern technology like the internet must be considered </claim> <claim> Do you really think abortions are only had by incompetent people who couldn't be good parents? </claim> <claim> This appears to be a belief that has led you to conclude that America must be better off </claim> <claim> you are really misinformed about the type of women who obtain abortions. </claim> <claim> That may be true. </claim> <claim> My thinking was that abortions were mainly required by unfortunate people, who in turn from their difficult circumstances, would've raised the child to be the same. </claim> <premise> Can you counter that? </premise> <premise> I have a [source] that shows that abortions are typically (and in some years, mostly) had by people without enough wealth to care for there kids. </premise> <premise> Some stats for example include 31 % of women obtaining abortions are in school; 68 % of women who have abortions are employed. </premise> <premise> 33 % of women who obtain abortions have family incomes under \$ 11,000 annually, 11 % of abortions are obtained by women whose household incomes are \$ 50,000 or more. </premise> <premise> that means the majority of those having abortions have income between \$ 11,000 and \$ 55,000, which is above the Federal Poverty line. </premise> <claim> I can't confirm but I do recall reading once that a common reason for women to abort is becoming pregnant too soon after having a child </claim> <premise> It could be about finances but could just be too soon for a woman to go through again. </premise> <premise> Pregnancy is hard on the body and far riskier than an abortion performed within a few weeks. </premise>"