“It’s a further exercise in futility”: implicit content detection and classification in Italian political discourse. A pilot study.

Walter Paci (Università degli Studi di Firenze)

walter.paci(at)unifi.it

Abstract

Implicit content, such as implicatures and presuppositions, is a key feature of political discourse, allowing speakers to convey meaning indirectly and influence audience interpretation. While Large Language Models (LLMs) have demonstrated impressive capabilities in natural language understanding, their ability to process implicit meaning in real-world contexts remains an open question. This study investigates whether state-of-the-art LLMs can detect and classify implicit content in Italian political speech. Using a subset of the IMPAQTS corpus we assess nine multilingual models, both open-weight and proprietary. The study comprises two tasks: a binary detection task, where models determine whether a given sentence contains implicit content, and a binary classification task, in which models identify whether the implicit content is conveyed through implicature or presupposition. To enhance model performance, we employ six different prompting techniques. Results reveal that while some proprietary models exhibit moderate success in detecting implicit content, none surpass chance-level performance in classification. Open-weight models consistently underperform, with accuracy scores hovering near random guessing. Among prompting strategies, more structured techniques achieve marginal improvements in detection but fail to enhance classification accuracy. These findings highlight the persistent challenges LLMs face in pragmatic reasoning, defining implicit content detection and classification as unresolved tasks in NLP.

Keywords

Implicit content, political discourse, Large Language Models, implicatures, presuppositions, NLP, prompting techniques

1 Introduction

Implicit communication is a fundamental characteristic of political discourse, where speakers frequently rely on pragmatic mechanisms such as implicatures and presuppositions to shape audience perceptions and convey messages indirectly. By employing these implicit strategies, politicians can subtly influence interpretation, leveraging shared knowledge, cultural references, and ideological predispositions to persuade without making explicit claims (Morency, Oswald and De Saussure 2008; Lombardi Vallauri 2019). See the example below:

(1) Questa è l’ulteriore, ennesima discussione sterile!
‘This discussion is a further, umpteenth, exercise in futility!’[1]

[1] Here and in the rest of the paper, example sentences have been manually translated.

DOI: 10.62408/ai-ling.v2i2.24
AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses ISSN: 2943-0070

The excerpt presupposes that there were already many other pointless discussions, otherwise that could not be the *further*, *umpteenth* futile discussion. This phenomenon exemplifies presuppositions (Strawson 1964; Garner 1971; Ducrot 1972), a central topic in pragmatic theory, alongside implicatures (Grice 1975; Sperber and Wilson 1987).

Since both are widespread in political discourse, this domain serves as an optimal setting for examining implicit communication strategies (Van Dijk 1992; Lombardi Vallauri and Masia 2020).

The rapid advancements in Large Language Models (LLMs) have positioned them as powerful tools for processing natural language, with demonstrated capabilities in semantic comprehension (Wang et al. 2018; Williams, Nangia and Bowman 2018) and emerging competence in pragmatic reasoning (Zheng et al. 2021; Hu et al. 2023; Kim, Taylor and Kang 2023). However, the extent to which these models can accurately interpret implicit content in real-world political discourse remains largely unexamined. Unlike artificially constructed pragmatic stimuli, political language is characterized by its rhetorical complexity, ideological framing, and strategic ambiguity (Van Dijk 1992; Lombardi Vallauri and Masia 2020), making it an ideal candidate for evaluating LLMs' ability to process implicit meaning beyond surface-level text.

In this pilot study, we address this gap by assessing the ability of pre-trained LLMs to detect and classify implicit content in Italian political speeches. Leveraging a small subset of the IMPAQTS corpus (Cominetti, Gregori, Lombardi Vallauri and Panunzi 2024), a large-scale dataset of transcribed political speeches with expert-annotated pragmatic phenomena, we systematically evaluate LLMs on two tasks:

1. A binary identification task, where LLMs must provide a Yes/No answer to a question asking if the input sentence has some implicit meaning.
2. A binary classification task where the LLMs are asked to recognize if implicit contents are conveyed through implicatures or presuppositions.

Results show that no model achieves results significantly better than chance, indicating that the task is highly challenging for current LLMs. The application of different prompting techniques (PTs) does not appear to have a meaningful impact on the models' overall ability to understand the task, suggesting that structural limitations in their reasoning persist regardless of prompting adjustments. More compact LLMs, which contain fewer parameters yet approach state-of-the-art capabilities in other domains, prove entirely inadequate for both detecting and classifying implicit content.

2 Related Work

Several studies have examined whether LLMs can interpret implicit meaning by testing their performance on specific pragmatic tasks. Jeretic, Warstadt, Bhooshan and Williams (2020) explored the extent to which models could infer scalar implicatures and recognize presuppositional triggers. Their findings revealed that while LLMs can learn certain systematic patterns, their understanding remains inconsistent, especially when dealing with more complex or less frequent cases. Similarly, Zheng et al. (2021) assessed LLMs' ability to resolve conversational implicatures based on Grice's maxims, showing that while models performed reasonably well in structured multiple-choice tasks, they struggled significantly when required to infer meaning in more open-ended or naturalistic contexts.

The impact of different prompting strategies on pragmatic reasoning has also been investigated: Hu et al. (2023) analysed whether LLMs could infer pragmatic meaning without task-specific fine-tuning, finding that although models displayed some implicit understanding, their performance suggested a reliance on linguistic surface patterns rather than true inferential reasoning. Kim, Taylor and Kang (2023) extended this research by using binary-choice tasks to assess conversational implicatures, showing that models improved notably when guided through structured reasoning using chain-of-thought (CoT) prompting. More recently, Ruis et al. (2023) introduced a framework for evaluating implicature resolution in LLMs, demonstrating that even state-of-the-art models exhibit substantial gaps when compared to human reasoning. Collectively, these studies indicate that while LLMs have made progress in handling pragmatic phenomena, their ability to detect and classify implicit meaning remains limited.

Beyond general pragmatic language understanding, NLP research has increasingly focused on computational approaches to political discourse. Researchers have leveraged large-scale datasets of political speech and social media discourse to analyse ideological framing, sentiment, and rhetorical strategies. Katre (2019) applied NLP techniques to investigate linguistic patterns in political speech, using computational methods to visualize lexical trends and thematic structures. Huguet Cabot et al. (2020) examined how metaphor and emotional rhetoric contribute to political persuasion, demonstrating that combining metaphor and sentiment detection enhances the ability to classify political perspectives and ideological framing. More broadly, Németh (2023) reviewed NLP applications in the study of political polarization, highlighting how computational methods have been used to track shifts in discourse and identify ideological divisions over time.
Recent advancements in LLMs have further expanded the scope of political discourse analysis. Marino and Giglietto (2024) explored the use of LLMs for political annotation in social media, showing that these models improve upon traditional topic modelling techniques by providing richer contextual insights. Similarly, Li et al. (2024) introduced Political-LLM, a framework for incorporating LLMs into political science research, emphasizing their potential for large-scale text analysis while also raising concerns regarding biases and limitations in interpretability. However, despite these advances, little research has focused on the role of implicit meaning in political communication.

3 The IMPAQTS Corpus

To investigate the ability of Large Language Models (LLMs) to interpret implicit meaning in political discourse, we leverage the IMPAQTS corpus (Cominetti, Gregori, Lombardi Vallauri and Panunzi 2024), a large-scale dataset of Italian political speeches annotated for implicit content. This corpus provides a rich and diverse collection of transcribed monologues spanning multiple decades, offering an ideal resource for assessing how LLMs can process naturally occurring implicatures and presuppositions. Unlike many prior studies that rely on artificial and toy examples, IMPAQTS presents real-world political speech, thereby ensuring an ecologically valid evaluation of LLMs in complex communicative contexts.

IMPAQTS comprises 1,500 speeches delivered by 150 prominent Italian politicians between 1946 and 2023, encompassing a total of approximately 2.65 million tokens. The dataset includes both video and audio recordings, with manually verified transcriptions serving as the basis for our study. A key feature of this corpus is its expert annotations, which systematically identify passages embedding non-bona fide true implicit content— that is, implicitly conveyed statements that are interpreted as true within a given context by the listener even if they stem from questionable assumptions or presuppositions. Such implicit contents enable speakers to advance potentially controversial positions through seemingly straightforward or presupposed statements, reducing the cognitive effort required to process them and eliciting uncritical acceptance (Morency, Oswald and De Saussure 2008; Lombardi Vallauri 2019).

IMPAQTS contains tags for four types of implicit content: *implicature*, *presupposition*, *topicalization*, and *vagueness*. For clarity, we briefly illustrate each category with a simple example.[2]

(2) Alla cultura torna Franceschini. Ma come si fa a stare in un Paese che non ha Franceschini al Governo? [MSAL19-N1][3]
At [the ministry of] culture comes back Franceschini. After all, how is it possible to have a country without Franceschini in its government?’

(3) Volevate riavvicinare i cittadini alla cosa pubblica. [MREN16-F1]
‘You wanted to bring back citizens to the State.’

(4) Questo grande spazio di libertà noi continueremo a coltivarlo [AALF11-P1]
‘This great space of freedom, we will continue to cultivate it.’

[2] For an extensive presentation of these phenomena see Cominetti, Gregori, Lombardi Vallauri and Panunzi (2024).

[3] Each speech in IMPAQTS is uniquely identified by a 9-character alphanumeric code, such as MSAL19-N1. The first four letters indicate the speaker’s first initial followed by the first three initials of their last name. In the case of compound surnames, the second letter of the code corresponds to the first letter of the first part of the surname, while the third and fourth letters correspond to the first and second letters of the second part (e.g., MSAL = Matteo Salvini; EDNI = Enrico De Nicola; RRIE = Rosa Russo Iervolino). This is followed by two digits indicating the last two digits of the year in which the speech was delivered (e.g., 19 for 2019, 46 for 1946). A hyphen (‘-’) follows. The next character is a letter indicating the type of speech: Parliamentary speeches (A), rally speeches (C), party assemble speeches (P), statement in presence (F), broadcast statements (T), face-to-face statement (F), new media statements (N), interview (I) and operative assembly (O). The last number in the code is used to disambiguate more speeches uttered in the same year by the same speaker.

(5) Hanno fatto passare il principio che, per governare, era necessario non rispettare i programmi e pazienza per i cittadini. [LDMA18-T1]
‘They’ve pushed the idea that, in order to govern, it was necessary to disregard the political agenda and that citizens should simply put up with it.’

In (1) we have an implicature where the speaker uses irony to flout the Maxim of Quality: a false assertion is made to convey its opposite, i.e., that Franceschini’s returning as minister is not due to his merits.

In (2) a presupposition is exemplified. The speaker gives for granted two things: (i) citizens are not interested in politics (*avvicinare*, ‘bring close’) and (ii) citizen were closer to politics before (*riavvicinare*, ‘bring back again’).

Topicalizations, as the one presented in (3), mark cases where a speaker presents a referent or proposition as given or already salient in the discourse, even when it has not been explicitly introduced in the preceding co-text or context. It relies on the cognitive distinction between topic, i.e. information assumed to be shared or activated in the listener’s short-term memory, and focus, i.e. what introduces new information. Linguistically, topicalization is typically encoded through syntactic structures that foreground the topic (such as clitic left-dislocation in Italian), or through prosodic prominence in spoken discourse. In topicalizations the hearer is encouraged to treat the topicalized element as part of the common ground. In our example, the *grande spazio di libertà* (‘great space of freedom’) is topicalized in respect to the actions the speaker is presenting.

Vagueness is presented in (4), with the speaker that leaves unclear the understood subject that hold the verb phrase *hanno fatto passare l’idea* (‘pushed the idea’), presupposing the existence of some inscrutable agents who worked behind the curtains to disregard political agendas and citizen’s needs.

For the detection task, we focus on all the categories presented above while, for the classification task, we chose to focus on the two most pervasive and functionally significant forms of implicit communication in political language: implicatures and presuppositions (Grice 1975; Sperber and Wilson 1987; Lombardi Vallauri 2016).

Each annotated passage within IMPAQTS is accompanied by an expert-generated explanation that explicitly reconstructs the intended meaning of the implicit content. These explanations follow a standardized structure, ensuring consistency across annotations. In particular, implicatures are introduced with the phrase *implica che…* (‘it implies that...’), presuppositions with *presuppone che…* (‘it presupposes that...’), topicalizations with *dà per attivo nel discorso che…* (‘it gives as active in the discourse that…’), vagueness with *lascia vago che…* (‘it leaves vague that…’).

To construct a balanced and representative evaluation dataset, we randomly extracted 100 samples for each type of implicit content—implicatures, presuppositions, topicalizations and sentences containing vagueness— from the IMPAQTS corpus, resulting in a total of 400 instances of implicit meaning. Additionally, we selected 400 samples containing no implicit content to serve as a control set, ensuring a meaningful comparison between sentences that require pragmatic inference and those that do not. The final dataset comprises 800 utterances, allowing us to systematically analyse LLM performance across different types of implicit meaning while controlling for potential confounding factors such as lexical complexity and discourse structure.

4 The Experiment

Using the dataset presented above, we challenge models to detect and classify a sentence's implicit content through different prompting techniques. Once each technique has been tested and evaluated, we select the best performing model and the two best performing techniques for that model; then, we engineer a prompt that comprises both to check if they elicit better results.

In this section, we describe which models we tested, how experiments were conducted and the experimental details common to both tasks.

4.1 Models

We experiment with nine models pre-trained with multilingual data, needed to process Italian language; six of these models are open-weight while three are proprietary. For the former class, we test LLAMA2 7B (Touvron et al. 2023), Falcon 7B (Almazrouei et al. 2023), Multi-Verse 7B,[4] Mistral 7B (Mistral AI 2023a), Mixtral 7B (Mistral AI 2023b), Laser-Dolphin Mixtral 2x7B;[5] for the latter we test ChatGPT3.5-turbo (OpenAI 2023a), ChatGPT4-turbo (OpenAI 2023b). We also test GPT-4o (OpenAI 2024) but only on the best performing prompting techniques to contain project costs.

We choose models with a strong difference of parameters' scale in order to investigate if bigger models could perform better than smaller ones. The open-weight models have less parameters (7B to 9B) with respect to proprietary ones (the exact number of which is not known).

4.2 Prompting Techniques

For both tasks, we compute each model with six different prompting techniques: zero-shot (ZS), few-shot (FS), Chain-of-Thought (CoT), Zero-Shot chain-of-thought (CoT-ZS), Sociodemographic (SD) prompting (Beck, Schuff, Lauscher and Gurevych 2024) and Generated Knowledge (GN) Prompting. We briefly describe them below:[6]

Zero-shot: No prior examples or detailed guidance are given to the models; they solely rely on their pre-training knowledge.

[4] Multi-Verse is a Mistral-based model tuned with an innovative technique. For details see https://huggingface.co/MTSAIR/multi_verse_model

[5] Models introduced by the label "laser-dolphin" are models fine-tuned with a prompting technique that tries to avoid the so-called *refusal mechanism*. For details see https://huggingface.co/TheBloke/laser-dolphin-mixtral-2x7b-dpo-GGUF

[6] Here we use simplify and shortened versions of each prompt setting of the detection task for space reasons. Prompting template and details can be found in Appendix.

(6) `Dimmi se l'enunciato seguente veicola un contenuto implicito non bona fide vero. Rispondi solo sì o no.`
`Enunciato:` [frase da analizzare]
`Presenza di contenuto implicito non bona fide vero:`
`‘Tell me if this sentence conveys a non bona fide true implicit content. Give a yes or no answer only.`
`Sentence:` [sentence to be analysed]
`Presence of non bona fide true implicit content:’`

Few-shot: Instead of relying on a fixed set of manually selected demonstrations, we adopt a *query-adaptive* few-shot strategy. For each test instance, the model receives four demonstrations randomly sampled from a constrained subset of 20 items (5 implicatures, 5 presuppositions, 10 negative instances). Sampling is performed independently for each query and each model, so every input is paired with a newly generated set of balanced exemplars drawn from the same distribution.

This design choice reflects growing evidence that static few-shot prompts have limited flexibility and often fail when test instances diverge from the fixed exemplars (Liu et al. 2022; Min et al. 2022). Manually selecting a static example set also risks introducing unintended biases, since performance may depend strongly on the specific demonstrations chosen (Zhao et al. 2021). By contrast, dynamic, retrieval-inspired few-shot prompting, where demonstrations are resampled or retrieved per query, has been shown to increase robustness and reduce sensitivity to prompt choices (Rubin, Herzig and Berant 2022; Liu et al. 2022).[7]

(7) `Enunciato:` [Primo esempio di enunciato con contenuto implicito]
`Presenza di contenuto implicito non bona fide vero: Sì`
`Enunciato:` [Primo esempio di enunciato senza contenuto implicito]
`Presenza di contenuto implicito non bona fide vero: No.`
[…]
`Enunciato:` [enunciato che si vuole analizzare]
`Presenza di contenuto implicito non bona fide vero:`

`‘Sentence:` [First sample with implicit content]
`Presence of non bona fide true implicit content: Yes.`
`Sentence:` [First sample without implicit content]
`Presence of non bona fide true implicit content: No.`
[…]
`Sentence:` [sentence we want to analyse]
`Presence of non bona fide true implicit content:’`

Chain-of-Thought: We give an extensive step-by-step explanation to the model to elicit the reasoning process needed to decode the implicit content. The explanation has been manually written by an expert linguist and contains instructions to tackle each of the four types of implicit content present in our experimental dataset.

(8) `L'implicatura è un'inferenza che un ascoltatore può fare basandosi sul contesto e su come viene formulato un enunciato. […]`
`Le presupposizioni non sono parte del contenuto esplicito dell'enunciato ma sono implicite e devono essere vere affinché l'enunciato abbia senso. […]`

[7] We acknowledge that this design choice has some important limitations. We present them in Section 7.

```
La topicalizzazione è un fenomeno linguistico che riguarda la
strutturazione dell'informazione in un enunciato, mettendo in
evidenza l'argomento (o "topic") della frase. […]
La vaghezza è una strategia pragmatica utilizzata per
esprimere contenuti in modo non specifico o indeterminato.
[…]
Alla luce di quanto detto, dimmi se l'enunciato seguente
veicola un contenuto implicito non bona fide vero. Procedi
passo dopo passo e produci una risposta "sì/no".
Enunciato: [enunciato che si vuole analizzare]
Presenza di contenuto implicito non bona fide vero:

'An implicature is an inference that a listener can make based
on the context and the way a statement is formulated. […]
Presupposition is information that is considered given,
accepted as true, and shared among the participants in the
conversation. […]
Topicalization is a linguistic phenomenon that concerns the
structuring of information in a statement, highlighting the
subject (or 'topic') of the sentence. […]
Vagueness is a pragmatic strategy used to express content in
a non-specific or indeterminate way. […]
According to what has been said, tell me if the following
statement conveys non-bona fide true implicit content.
Proceed step by step and produce a 'yes/no' answer.
Sentence: [sentence to be analysed]
Presence of non-bona fide true implicit content:'
```

Zero-shot Chain-of-Thought: This prompting technique is an enhanced version of the ZS prompt; in fact, no prior examples or detailed guidance are given to the models but a "procedi passo dopo passo" (*proceed step-by-step*) clause is added before the output formatting instructions.

(9)

```
Dimmi se l'enunciato seguente veicola un contenuto implicito
non bona fide vero. Procedi passo dopo passo e produci una
risposta sì/no.
Enunciato: [frase da analizzare]
Presenza di contenuto implicito non bona fide vero:

'Tell me if this sentence conveys a non bona fide true implicit
content. Give a yes or no answer only. Proceed step-by-step
and give me a yes/no answer.
Sentence: [sentence to be analysed]
Presence of non bona fide true implicit content:'
```

Sociodemographic prompting: The Sociodemographic prompting technique aims to assess whether providing additional contextual information about the speaker influences the model's ability to interpret implicit content. Given that political discourse is deeply shaped by the speaker's background, ideological stance, and historical context, integrating sociodemographic cues could help LLMs better approximate the pragmatic inferences that human listeners naturally make when interpreting political statements.

To implement this technique, we design a two-step approach: in the first step, we prompt the model to generate a concise biographical description[8] of the politician who uttered the sentence being analysed. This description should provide

[8] This information is extracted from the metadata of the IMPAQTS corpus.

contextual insights into the speaker's political affiliation, historical period, and public persona, all of which may influence how implicit content can be interpreted. Even if this information is generated by the model itself rather than extracted from a verified database, we do not check for hallucinations in this process, as the goal is to observe whether a model can retrieve the contextual and world knowledge needed to understand an implicit meaning from its pretraining data.

In the second step, we adapted the short zero-shot prompt presented above, instructing the model to perform the target task.

The two steps were implemented as separate API calls: the reasoning process for the main task did not depend on hidden conversational memory but only on the explicit textual description passed to the model.

(10) [Prompt per la generazione della descrizione del parlante]
`Scrivi una breve presentazione di` [politico]`, parlando del periodo di attività della sua carriera politica, del partito di appartenenza e delle sue principali idee.`

[Prompt per il task di detection]
`Di seguito troverai la descrizione di una persona e un enunciato. L'enunciato è stato pronunciato dalla persona descritta. Dimmi se l'enunciato ha del contenuto implicito non bona fide vero, anche considerando la persona che lo ha pronunciato e il suo pensiero e orientamento politico. Rispondi solo sì o no.`

`Descrizione:` [testo generato col prompt precedente]
`Enunciato:` [enunciato che si vuole analizzare]
`Presenza di contenuto implicito non bona fide vero:`

'[Prompt for generating the speaker's description]
`Write a brief introduction of` [politician]`, discussing the period of their political career, their party affiliation, and their main ideas.`

[Prompt for the detection task]
`Below, you will find a description of a person and a statement. The statement was made by the person described. Tell me if the statement contains non-bona fide implicit content, considering also the person who made it and their political views and orientation. Answer only yes or no.`

`Description:` [previously generated description]
`Sentence:` [sentence to be analysed]
`Presence of non-bona fide true implicit content:`'

Generate Knowledge: In implementing the Generate Knowledge prompting technique, we designed a structured two-step process to enhance the model's ability to interpret implicit content by first eliciting relevant background knowledge before task execution.

In the first step, we prompted the model to generate a concise yet comprehensive definition of five core pragmatics concepts directly relevant to the interpretation of implicit content in political discourse. These concepts included implicature, presupposition, entailment, speech acts, and pragmatic inferences, selected for their fundamental role in shaping non-explicit meaning. The goal was to ensure that the model established a foundational understanding of these concepts before attempting to process implicit language. We systematically verified that each generation consistently produced exactly five distinct definitions, maintaining clarity and coherence while avoiding irrelevant or incomplete responses. Any deviations, such as missing definitions or off-topic explanations, were discarded, and the process was repeated until the desired output structure was achieved.

In the second step, we appended the short zero-shot prompt previously presented to the generated knowledge, explicitly instructing the model to apply its newly articulated understanding to the target task.

(11) [Prompt per la generazione delle conoscenze]
`Genera cinque diverse conoscenze relative alla pragmatica linguistica. Utilizza il template seguente per la generazione.`

`Input: {Concetto di Linguistica}`
`Conoscenza: {Breve ed esaustiva spiegazione del concetto}`

[Prompt per il task di detection:]
`Input:` [Conoscenza 1 generata precedentemente]
`Conoscenza:` [Descrizione di conoscenza 1]
[…]
`Conoscenza:` [Descrizione di conoscenza 5]
`Basandoti anche sulle conoscenze presentate sopra, identifica se nell'enunciato seguente sono presenti contenuti impliciti non bona fide veri. Rispondi solo sì o no.`
`Enunciato:` [enunciato da analizzare]
`Presenza di contenuto implicito non bona fide vero:`

'[Prompt for generating knowledge]
`Generate five different pieces of knowledge related to pragmatics. Use the following template for generation.`

`Input:` {Linguistic Concept}
`Knowledge:` {Brief and comprehensive explanation of the concept}

[Prompt for the detection task]
`Input:` [Knowledge 1 generated previously]
`Knowledge:` [Description of Knowledge 1]
[…]
`Knowledge:` [Description of Knowledge 5]
`Based also on the knowledge presented above, identify if the following sentence contains non-bona fide true implicit content. Answer only yes or no.`
`Sentence:` [sentence to be analysed]
`Presence of non-bona fide implicit content:`'

4.3 Evaluation

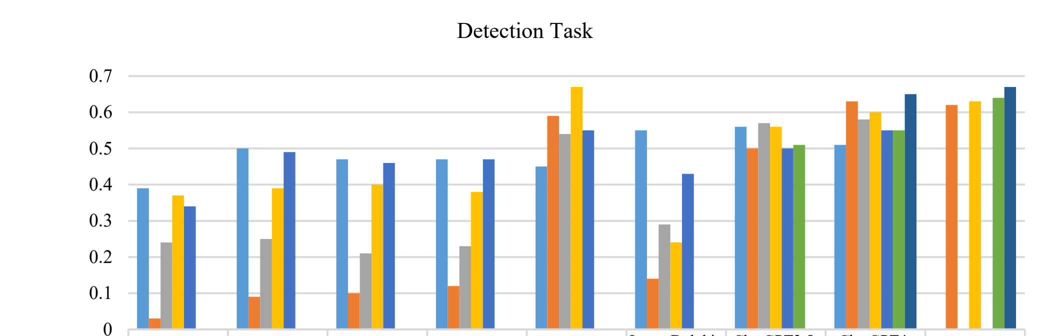
Given the fact that both tasks are binary and that we have a ground-truth available through the corpus annotation, we evaluate model performance through plain accuracy, defining chance guessing at 50%.

5 Results
5.1 Detection task

Results for the detection task are available in Figure 1. The implicit content detection task reveals substantial differences in performance across models and prompting techniques. Open-weight models such as LLAMA2, Falcon, and Mistral 7B generally struggle with the task, showing low scores across all prompting techniques. Notably, Few-Shot (FS) prompting yields particularly poor results for these models, with scores close to chance. ZS-CoT and CoT improve performance slightly, but the gap between open-weight and closed-weight models remains evident. The GN prompt consistently produced hallucinations for all open-weight models, preventing reliable evaluation.

Among the open-weight models, Mixtral 7B performs best, particularly under CoT, reaching 67% accuracy. Laser-Dolphin Mixtral 2x7B is the best performing open-weight model in the ZS setting, achieving 55% accuracy.

Conversely, ChatGPT models consistently outperform open-weight LLMs, with ChatGPT-4-turbo and GPT-4o achieving the highest accuracy scores: GPT-4o was evaluated only on the best-performing prompting techniques and achieved results comparable to, or slightly surpassing, the other proprietary models. While ZS-CoT and CoT provide significant improvements across models, the combination of FS and CoT yields the highest performance overall, reaching 65% in ChatGPT-4-turbo and 67% in GPT-4o, confirming that more complex prompting techniques can elicit better performance.



	LLAMA2	Falcon	Multiverse	Mistral 7B	Mixtral 7B	Laser-Dolphin Mixtral 2x7B	ChatGPT3.5-turbo	ChatGPT4-turbo	GPT-4o
ZS	0.39	0.5	0.47	0.47	0.45	0.55	0.56	0.51	
FS	0.03	0.09	0.1	0.12	0.59	0.14	0.5	0.63	0.62
ZS-CoT	0.24	0.25	0.21	0.23	0.54	0.29	0.57	0.58	
CoT	0.37	0.39	0.4	0.38	0.67	0.24	0.56	0.6	0.63
SD	0.34	0.49	0.46	0.47	0.55	0.43	0.5	0.55	
GN							0.51	0.55	0.64
FS + CoT								0.65	0.67

Figure 1: Detection Task. Models Accuracy.

A qualitative analysis of model outputs shows why improvements lead by prompting techniques remain limited. Below we report two sentences that all models correctly identified as containing an implicit content in the ZS setting.

(12) Il talk show è una sorta di Uomini e Donne della politica in cui l'unica cosa che cercano è lo scontro. Ma noi non ci andiamo, perché nei talk-show non si parla di TTIP.
'Talk shows are a kind of political Uomini e Donne, where the only thing they look for is confrontation. But we don't go on them, because in talk shows nobody talks about TTIP.'

(13) Ecco perché io dico a voce alta che loro sono ladri di futuro.
'This is why I say out loud that they are thieves of the future.'

These sentences contain salient rhetorical cues like metaphorical framings (*Il talk show è una sorta di Uomini e Donne della politica*) and causal connectors (*Ecco perché*) that make implicit meaning easily identifiable.

However, the same reliance on surface cues also generates systematic false positives: all models tagged the following sentences as containing an implicit content in the ZS-CoT setting.

(14) Noi riteniamo che questo strumento sia illegittimo, oltre che essere anticostituzionale.
'We believe that this instrument is illegitimate, as well as unconstitutional.'

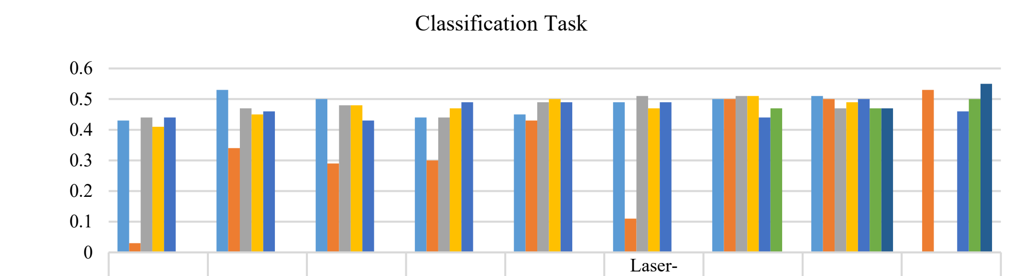
(15) I vertici militari hanno asserito che dal punto di vista tecnico bastano poche decine di giorni.
'The military leadership stated that, from a technical standpoint, just a few dozen days are enough.'

Utterances that contain explicit, non-inferential evaluations (*We believe that this instrument is illegitimate*) or neutral reported speech (*The military leadership stated that…*) are repeatedly misclassified as implicit. In these cases, models appear to overgeneralize from distributional regularities in political discourse, treating evaluative adjectives and epistemic framing as marker of implicitness even when no inference is required.

5.2 Classification task

Results for this task show without a doubt the challenging nature of classifying implicit content, with no model significantly surpassing chance guessing level performance (50%) across all prompting techniques. Open-weight models such as LLAMA2, Falcon, and Mistral 7B exhibit limited improvements across different prompting methods, with scores fluctuating around random guessing. FS prompting proves particularly ineffective, with extreme cases of hallucinations—for instance, LLAMA2 drops from 43% of accuracy in the ZS setting to 3% of FS; Laser-Dolphin Mixtral 2×7B falls from 49% of ZS to 11% of FS. SD prompting does not yield substantial improvements across models, producing results mostly in line with ZS and CoT techniques.

For closed-weight models, ChatGPT4-turbo and GPT-4o maintain the highest accuracy, though their performance remains comparable to open-weight models in most cases. GPT-4o, tested only on the best-performing techniques, reaches 53% accuracy in the FS setting and 55% accuracy in the FS + SD setting. Both results are still only slightly above chance; nonetheless, FS + SD confirms that, as in the detection task, more complex prompting techniques can elicit (slightly) better reasoning.



	LLAMA2	Falcon	Multiverse	Mistral 7B	Mixtral 7B	Laser-Dolphin Mixtral 2x7B	ChatGPT3.5 -turbo	ChatGPT4-turbo	GPT-4o
ZS	0.43	0.53	0.5	0.44	0.45	0.49	0.5	0.51	
FS	0.03	0.34	0.29	0.3	0.43	0.11	0.5	0.5	0.53
ZS-CoT	0.44	0.47	0.48	0.44	0.49	0.51	0.51	0.47	
CoT	0.41	0.45	0.48	0.47	0.5	0.47	0.51	0.49	
SD	0.44	0.46	0.43	0.49	0.49	0.49	0.44	0.5	0.46
GN							0.47	0.47	0.5
FS + SD								0.47	0.55

Figure 2: Classification Task. Models Accuracy.

As we have already seen for the detection task, a qualitative inspection of the classification outputs clarifies why performance on this task remains poor across prompting techniques. When models correctly classified implicatures, they did so in sentences exhibiting highly conventionalized rhetorical frames. Below we report two examples identified as implicatures in the ZS setting. Similar patterns emerge also in other prompting techniques.

(16) […] coloro che stanno chiusi nei palazzi […]
‘[…] those who remain shut inside their palaces […]’

(17) Finalmente il Governo ha fatto marcia indietro.
‘Finally the Government backtracked.’

Both utterances rely on salient inferential cues: in (16) the metaphor *stanno chiusi nei palazzi* indirectly conveys political detachment, while, in (17) *ha fatto marcia indietro* expresses evaluative judgment through an idiom. These forms of conventionalized figurative language can mark implicatures without requiring substantial pragmatic reconstruction.

However, the same reliance on surface-level rhetorical cues also explains why presuppositions are correctly classified primarily when they contain explicit lexical triggers. Consider the following sentences, which models systematically identified as presuppositions with ZS-CoT and CoT prompting:

(18) […] stanziando risorse per equilibrare le differenze territoriali.
'[…] allocating resources to balance territorial differences.'

(19) Oggi si fa un altro passo in avanti […].
'Today we take another step forward […].'

Expressions like *un altro passo* or *il risveglio di attenzione* carry clear change-of-state or iterative presuppositions: models' correct classifications reflect not deep pragmatic competence but their ability to detect canonical presupposition triggers.

Error analysis reveals also two systematic patterns of misclassification: First, many implicatures were misidentified as presuppositions whenever they contained aspectual or temporal markers, or change-of-state verbs, even when the interpretive process required inference rather than encoded presupposition. For instance, under ZS-CoT, models classified the following implicatures as presuppositions:

(20) Finora questo diritto partiva prima dalla condizione dello straniero, ora parte dai cittadini italiani.
'Until now this right started firstly from the foreigners' condition, now it starts from Italian citizens...'

(21) C'è un vento che si è alzato […].
'There's a wind that has risen […].'

Here the idiomatic or metaphorical meaning is inferential, but the surface form contains bona fide true presuppositional structures: in fact, *finora* presupposes that now things are different while *un vento che si è alzato* takes as granted that no wind was blowing before. Models appear to overgeneralize from these syntactic patterns, prioritising formal resemblance over the pragmatic mechanism at stake, i.e. manipulative implicit content.

Second, several presuppositions were misclassified as implicatures, particularly when embedded in metaphorical or evaluative rhetoric. Examples include:

(22) Lei non ha esitato a fare a pezzi interi ministeri.
'She did not hesitate to tear entire ministries to pieces.'

(23) Le istituzioni dovrebbero anche fare pace con il cervello.
'Institutions should also find peace within themselves.[9]'

Although these expressions contain non-bona fide true presuppositions – in fact both in (22) and (23) the presuppositions rely on a moral judgment of the action

[9] Literally "make peace with their brains".

described in the sentences, i.e. the fact that to tear to pieces ministries and not to be in peace with oneself are bad things –, their figurative and emotive framing prompted models to interpret them as implicatures. In these cases, strong metaphors and moralizing language overshadow structural presuppositional cues, leading to systematic category confusions.

From these examples appears evident that models do not disambiguate implicatures from presuppositions on the basis of their underlying pragmatic properties but instead, they classify according to lexical-syntactic correlates, with metaphor and evaluative polarity playing a crucial role in implicature and aspectual or iterative morphology in presupposition, irrespective of whether these cues encode actual non-bona fide implicit content or merely resemble canonical forms.

6 Conclusions

In this paper, we investigated the ability of LLMs to detect and classify implicit content in Italian political discourse, focusing on various prompting techniques to assess whether structured guidance improves performance. We conducted a twofold evaluation: a detection task, which aimed to identify the presence of implicit content, and a classification task, which required models to accurately determine the nature of the implicit meaning conveyed.

Our results indicate that implicit content detection is feasible for advanced LLMs, particularly State-of-the-Art closed-weight models such as ChatGPT4-turbo and GPT-4o, which significantly benefit from structured prompting techniques like CoT and engineered, task-specific ones. Conversely, open-weight models struggle considerably, with performance often near chance, and some prompting strategies—such as FS—even harming rather than enhancing detection accuracy. This suggests that, while state-of-the-art closed models manage to capture some implicit content signals, open-weight alternatives still face substantial limitations in handling these linguistic phenomena.

Our qualitative analysis of detection outputs showed how models perform best when overt rhetorical and lexical signatures, such as metaphorical framings or idioms, are present. Conversely, they systematically produce false positives in explicit evaluative assertions or indirect discourses, revealing an overgeneralisation from distributional regularities rather than true pragmatic inference.

Moreover, the classification task proved substantially more challenging across all models. Even the most advanced LLMs, including GPT-4o, failed to significantly outperform random guessing, revealing that implicit meaning understanding remains a largely unsolved problem in computational linguistics. While prompting techniques helped detection performance, they did not lead to comparable gains in classification accuracy. Qualitative inspection of the outputs showed that implicatures were correctly classified by all models only when expressed through highly conventionalized figurative or adversative constructions, whereas presuppositions were classified correctly only when marked by explicit lexical and syntactical triggers. Misclassifications arose because models relied on surface-level morphological cues rather than on the pragmatic mechanism at work: aspectual or temporal markers led models to misclassify implicatures as presuppositions while strong metaphorical or moralizing language overshadowed presuppositional triggers, causing the reverse misclassification. These patterns seem to suggest that models are detecting surface-level signals rather than engaging in deep pragmatic reasoning. While this reliance on surface cues is clearly insufficient for capturing manipulative implicatures, it is worth noting that presuppositions are indeed activated by lexical or structural triggers: in this sense, the models' classification of some presuppositions demonstrates that they recognize canonical markers. However, the broader pattern shows that this ability does not generalize to cases where presuppositional and inferential elements coexist, i.e. where the truth value of what is presupposed is needed to disambiguate between bona fide and non-bona fide implicit content.

Additionally, we explored the impact of Sociodemographic prompting, which introduced contextual speaker-related information to assess whether models exhibit biases in processing implicit content. This approach aimed to approximate the contextualized pragmatic reasoning that human listeners naturally perform in political discourse. Our findings suggest that explicitly integrating speaker background does not significantly enhance model performance. Nonetheless, more research must be done on this subject to give a final evaluation of this approach: better world and contextual knowledge creation and retrieval methods should be tested and evaluated.

Overall, our study highlights the limitations of even state-of-the-art LLMs in handling implicit meaning, reinforcing the need for more advanced methods to bridge the gap between computational and human pragmatic understanding.

Future research should extend the present study in several directions: first, more models with different number of parameters should be tested to investigate how model scale and performance correlates. Second, more reliable methods for generating, retrieving, or constraining speaker- and world-knowledge may shed light on whether contextual enrichment can meaningfully support pragmatic interpretation. Third, expanding the analysis incorporating prosodic or multimodal cues could offer a more comprehensive picture of how LLMs process political communication. Finally, integrating qualitative error analysis with explainability techniques may help identify the specific linguistic triggers that lead models to conflate implicatures and presuppositions, paving the way for architectures or training procedures better suited to pragmatic reasoning.

7 Limitations

This study presents several limitations that should be acknowledged. First, the Few-Shot prompting condition revealed substantial instability, particularly in open-weight models; this suggests that FS prompting may introduce noise or spurious cues when the model lacks robust prior knowledge of Italian political discourse. Second, although our tasks were designed to isolate specific pragmatic phenomena, the evaluation was necessarily limited to textual input: prosodic, multimodal, and interactional cues were not considered. Third, the study relies on Italian data, a language for which many LLMs have less extensive pretraining exposure compared to English, potentially affecting generalizability. Finally, as with all prompting-based evaluations, the results may depend on particular wording and formatting choices, and further work is needed to assess the stability of these findings across alternative prompt designs and annotation schemas.

8 Conflict of Interest

The author declares no conflicts of interest regarding the publication of this contribution.

References

Almazrouei, Ebtesam & Alobeidli, Hamza & Alshamsi, Abdulaziz & Cappelli, Alessandro & Cojocaru, Ruxandra & Debbah, Mérouane & Goffinet, Étienne & Hesslow, Daniel & Launay, Julien & Malartic, Quentin et al. 2023. The Falcon series of open language models. *arXiv*. https://doi.org/10.48550/arXiv.2311.16867

Beck, Tilman & Schuff, Hendrik & Lauscher, Anne & Gurevych, Iryna. 2024. Sensitivity, performance, robustness: deconstructing the effect of sociodemographic prompting. In Graham, Yvette & Purver, Matthew (eds), *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics* 1 (St. Julian's, March 17-22, 2024), 2589–2615. Stroudsburg: Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.eacl-long.0

Cominetti, Federica & Gregori, Lorenzo & Lombardi Vallauri, Edoardo & Panunzi, Alessandro. 2024. IMPAQTS: A multimodal corpus of parliamentary and other political speeches in Italy (1946–2023), annotated with implicit strategies. In Fiser, Darja & Eskevich, Maria & Bordon, David (eds), *Proceedings of the IV Workshop on Creating, Analysing, and Increasing Accessibility of Parliamentary Corpora ParlaCLARIN @ LREC-COLING 2024* (Torino, May 20, 2024), 101–109. ELRA/ICCL. https://aclanthology.org/2024.parlaclarin-1.15/

Ducrot, Oswald. 1972. *Dire et ne pas dire: Principes de sémantique linguistique*. Paris: Hermann.

Garner, Richard. 1971. Presupposition in philosophy and linguistics. In Fillmore, Charles J. & Langendoen, D. Terence (eds), *Studies in linguistic semantics*, 23–45. New York: Irvington.

Grice, Herbert Paul. 1975. Logic and conversation. In Cole, Peter & Morgan, Jerry L. (eds), *Speech acts*, 41–58. New York: Academic Press.

Hu, Jennifer & Floyd, Samuel & Jouravlev, Olessia & Fedorenko, Evelina & Gibson, Edward. 2023. A fine-grained comparison of pragmatic language understanding in humans and language models. In Rogers, Anna & Boyd-Graber, Jordan & Okazaki, Naoaki (eds), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* 1 (Toronto, July 9-14, 2023), 4194–4213. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.230

Huguet Cabot, Pere-Lluís & Dankers, Verna & Abadi, David & Fischer, Agneta & Shutova, Ekaterina. 2020. The pragmatics behind politics: modelling metaphor, framing and emotion in political discourse. In Cohn, Trevor & He, Yulan & Liu, Yang (eds), *Findings of the Association for Computational Linguistics EMNLP 2020* (Online, November 16-20, 2020), 4479–4488. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.findings-emnlp.402

Jeretic, Paloma & Warstadt, Alex & Bhooshan, Suvrat & Williams, Adina. 2020. Are natural language inference models IMPPRESsive? Learning implicature and presupposition. In Jurafsky, Dan & Chai, Joyce & Schluter, Natalie & Tetreault, Joel (eds), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (Online, July 5-10, 2020), 8690–8705. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.768

Katre, Prakash D. 2019. NLP-based text analytics and visualization of political speeches. *International Journal of Recent Technology and Engineering* 8(3). 8574–8579. Blue Eyes Intelligence Engineering and Sciences Publication. https://doi.org/10.35940/ijrte.C6503.098319

Kim, Zae Myung & Taylor, Dylan E. & Kang, Dongyeop. 2023. "Is the Pope Catholic?" Applying chain-of-thought reasoning to understanding conversational implicatures. *arXiv*. https://doi.org/10.48550/arXiv.2305.13826

Li, Lincan & Li, Jiaqi & Chen, Catherine & Gui, Fred & Yang, Hongjia & Yu, Chenxiao & Wang, Zhengguang & Cai, Jianing & Aaron Zhou, Junlong & Shen, Bolin et al. 2024. Political-LLM: Large language models in political science. *arXiv*. https://doi.org/10.48550/arXiv.2412.06864

Liu, Jiachang & Shen, Dinghan & Zhang, Yizhe & Dolan, Bill & Carin, Lawrence & Chen, Weizhu. 2022. What makes good in-context examples for GPT-3? In Agirre, Eneko & Apidianaki, Marianna & Vulić, Ivan (eds), *Proceedings of Deep Learning Inside Out DeeLIO 2022* (Dublin, May 27, 2022), 100–114. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.deelio-1.10

Lombardi Vallauri, Edoardo. 2016. The "exaptation" of linguistic implicit strategies. *SpringerPlus* 5(1), 1106. 1–24. https://doi.org/10.1186/s40064-016-2788-y

Lombardi Vallauri, Edoardo. 2019. *La lingua disonesta: Contenuti impliciti e strategie di persuasione*. Bologna: Il Mulino.

Lombardi Vallauri, Edoardo & Masia, Viviana. 2020. La comunicazione implicita come dimensione di variazione tra tipi testuali. In Visconti, Jacqueline & Manfredini, Manuela & Coveri, Lorenzo (eds), *Linguaggi settoriali e specialistici: Sincronia, diacronia, traduzione, variazione. Atti del XV Convegno SILFI* (Genova, 28-30 maggio 2018), 113–120. Florence: Franco Cesati Editore.

Marino, Giada & Giglietto, Fabio. 2024. Integrating large language models in political discourse studies on social media: Challenges of validating an LLMs-in-the-loop pipeline. *Sociologica* 18(2). 87–107. https://doi.org/10.6092/issn.1971-8853/19524

Min, Sewon & Lyu, Xinxi & Holtzman, Ari & Artetxe, Mikel & Lewis, Mike & Hajishirzi, Hannaneh & Zettlemoyer, Luke. 2022. Rethinking the role of demonstrations: What makes in-context learning work? In Goldberg, Yoav & Kozareva, Zornitsa & Zhang, Yue (eds), *Proceedings of the 2022 Conference* *on Empirical Methods in Natural Language Processing* (Abu Dhabi, December 7-11, 2022), 11048–11064. Association for Computational Linguistics. https://aclanthology.org/2022.emnlp-main.0/

Mistral AI. 2023a. *Mistral 7B Model Card*. https://www.mistral.ai/models/mistral-7b (last accessed on 18/12/2025)

Mistral AI. 2023b. *Mixtral 7B Model Card*. https://www.mistral.ai/models/mixtral-7b (last accessed on 18/12/2025)

Morency, Patrick & Oswald, Steve & De Saussure, Louis. 2008. Explicitness, implicitness and commitment attribution: A cognitive pragmatic approach. In De Brabanter, Philippe & Dendale, Patrick (eds), *Commitment. Belgian Journal of Linguistics* 22(1). 197–219. Amsterdam: John Benjamins Publishing.

Németh, Renáta. 2023. A scoping review on the use of natural language processing in research on political polarization: Trends and research prospects. *Journal of Computational Social Science* 6(1). 289–313. London: Springer Publishing.

OpenAI. 2023a. *GPT-3.5-Turbo*. https://platform.openai.com/docs/models/gpt-3.5-turbo (last accessed on 18/12/2025).

OpenAI. 2023b. *GPT-4-Turbo*. https://platform.openai.com/docs/models/gpt-4-turbo (last accessed on 18/12/2025).

OpenAI. 2024. *GPT-4o: Multimodal AI model integrating text, images, and audio*. https://openai.com/index/hello-gpt-4o/ (last accessed on 18/12/2025).

Rubin, Ohad & Herzig, Jonathan & Berant, Jonathan. 2022. Learning to retrieve prompts for in-context learning. In Carpuat, Marine & de Marneffe, Marie-Catherine & Meza Ruiz, Ivan Vladimir (eds), *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Seattle, July 10-15, 2022), 2655–2671. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.naacl-main.191

Ruis, Laura & Khan, Akbir & Biderman, Stella & Hooker, Sara & Rocktäschel, Tim & Grefenstette, Edward. 2023. The Goldilocks of pragmatic understanding: Fine-tuning strategy matters for implicature resolution by LLMs. *Advances in Neural Information Processing Systems* 36. 20827–20905. https://doi.org/10.48550/arXiv.2210.14986

Sperber, Dan & Wilson, Deirdre. 1987. Précis of Relevance: Communication and cognition. *Behavioral and Brain Sciences* 10(4). 697–710. Cambridge: Cambridge University Press. https://doi.org/10.1017/S0140525X00055515

Strawson, Peter F. 1964. Identifying reference and truth-values. *Theoria* 30(2). 96–118. https://doi.org/10.1111/j.1755-2567.1964.tb00404.x

Touvron, Hugo & Martin, Louis & Stone, Kevin & Albert, Peter & Almahairi, Amjad & Babaei, Yasmine & Bashlykov, Nikolay & Batra, Soumya & Bhargava, Prajjwal & Bhosale, Shruti et al. 2023. *LLaMA 2: Open foundation and fine-tuned chat models*. https://doi.org/10.48550/arXiv.2307.09288

Van Dijk, Teun A. 1992. Discourse and the denial of racism. *Discourse & Society* 3(1). 87–118. London: SAGE Publications.

Van Dijk, Teun A. 1997. What is political discourse analysis. In Blommaert, Jan & Bulcaen, Chris (eds), *Political Linguistics. Belgian Journal of Linguistics* 11(1). 11–52. Amsterdam: John Benjamins Publishing.

Wang, Alex & Singh, Amanpreet & Michael, Julian & Hill, Felix & Levy, Omer & Bowman, Samuel. 2018. GLUE: A Multi-Task benchmark and analysis platform for Natural Language Understanding. In Linzen, Tal & Chrupala, Grzegorz & Alishahi, Afra (eds), *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP* (Brussels, November 1, 2018), 353–355. Association for Computational Linguistics. https://doi.org/10.18653/v1/W18-5446

Williams, Adina & Nangia, Nikita & Bowman, Samuel R. 2018. A broad-coverage challenge corpus for sentence understanding through inference. In Walker, Marilyn & Ji, Heng & Stent, Amanda (eds), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* 1 (New Orleans, June 1-6, 2018), 1112–1122. Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1101

Zhao, Tony Z. & Wallace, Eric & Feng, Shi & Klein, Dan & Singh, Sameer. 2021. Calibrate before use: improving few-shot performance of Language Models. In Meila, Marina & Zhang, Tong (eds.), *Proceedings of the 38th International Conference on Machine Learning* (Online, July 18-24, 2021), 12697–12706. Proceedings of Machine Learning Research. https://doi.org/10.48550/arXiv.2102.09690

Zheng, Zilong & Qiu, Shuwen & Fan, Lifeng & Zhu, Yixin & Zhu, Song-Chun. 2021. GRICE: A grammar-based dataset for recovering implicature and conversational reasoning. In Zong, Chengqing & Xia, Fei & Li, Wenjie, Li & Navigli, Roberto (eds), *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (Online, August 1-6, 2021), 2074–2085. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-acl.182

Appendix

Details

In both tasks, model responses are generated using greedy decoding, operationalized by setting the temperature parameter to 0, i.e. fully deterministic. For the detection task, both open-weight and proprietary models are allowed to generate up to 5 new tokens in each prompt setting. For the classification task, models could generate up to 25 new tokens with each prompting technique.

To compute open-weight models, we leverage a Nvidia Titan (15 GB) GPU for a total of roughly 250 compute hours. The cost of running experiments using OpenAI APIs were approximately €150 for all three models.

Prompt templates

Detection

Zero-Shot

```
Dimmi se l’enunciato seguente veicola un contenuto implicito non bona fide vero. Rispondi solo sì o no.
Enunciato: [frase da analizzare]
Presenza di contenuto implicito non bona fide vero:

‘Tell me if this sentence conveys a non bona fide true implicit content. Give a yes or no answer only.
Sentence: [sentence to be analysed]
Presence of non bona fide true implicit content:’
```

Few-Shot

```
Enunciato: [esempio di enunciato con contenuto implicito]
Presenza di contenuto implicito non bona fide vero: Sì
Enunciato: [esempio di enunciato senza contenuto implicito]
Presenza di contenuto implicito non bona fide vero: No.
Enunciato: [esempio di enunciato con contenuto implicito]
Presenza di contenuto implicito non bona fide vero: Sì
Enunciato: [esempio di enunciato senza contenuto implicito]
Presenza di contenuto implicito non bona fide vero: No.
Enunciato: [enunciato che si vuole analizzare]
Presenza di contenuto implicito non bona fide vero:

‘Sentence: [sample with implicit content]
Presence of non bona fide true implicit content: Yes.
Sentence: [sample without implicit content]
Presence of non bona fide true implicit content: No.
Sentence: [sample with implicit content]
Presence of non bona fide true implicit content: Yes.
Sentence: [sample without implicit content]
Presence of non bona fide true implicit content: No.
Sentence: [sentence we want to analyse]
Presence of non bona fide true implicit content:’
```

CoT

L'implicatura è un'inferenza che un ascoltatore può fare basandosi sul contesto e su come viene formulato un enunciato. Le implicature non sono esplicitamente enunciate, ma vengono dedotte dal contesto e dalle norme conversazionali. Le implicature possono essere di due tipi principali: conversazionali e convenzionali. Le implicature conversazionali si basano sui principi della comunicazione cooperativa, descritti da H.P. Grice nel suo saggio "Logic and Conversation" del 1975. Grice ha proposto il "Principio di Cooperazione", secondo il quale i partecipanti a una conversazione si aspettano che gli altri siano collaborativi e seguano certe massime: Massima di quantità, ovvero fornire informazioni sufficienti per la comprensibilità del discorso; massima di qualità, ovvero dire la verità, o almeno non dire ciò che si crede sia falso o di cui non si hanno prove sufficienti; massima di relazione, cioè essere pertinente con l'argomento del discorso; massima di modo, ovvero essere chiari, evitare ambiguità e essere brevi e ordinati.
Le implicature convenzionali non dipendono dalle massime conversazionali ma sono legate alla convenzione linguistica e al significato delle parole stesse. Ad esempio, l'uso di parole come "ma", "però" o "infatti" introduce un tipo specifico di implicatura.
La presupposizione è un'informazione che è considerata data, accettata come vera e condivisa tra i partecipanti alla conversazione. Le presupposizioni non sono parte del contenuto esplicito dell'enunciato ma sono implicite e devono essere vere affinché l'enunciato abbia senso. Le presupposizioni possono essere testate tramite il test della negazione: se la presupposizione sopravvive alla negazione dell'enunciato, è una presupposizione.
La topicalizzazione è un fenomeno linguistico che riguarda la strutturazione dell'informazione in un enunciato, mettendo in evidenza l'argomento (o "topic") della frase. In molte lingue, inclusa l'italiano, la topicalizzazione può alterare l'ordine normale delle parole per sottolineare ciò di cui si sta parlando.La vaghezza è una strategia pragmatica utilizzata per esprimere contenuti in modo non specifico o indeterminato. Questo può essere fatto per vari motivi, come evitare conflitti, mantenere la cortesia, o per lasciare spazio all'interpretazione.Le implicature, le topicalizzazioni, la vaghezza e le presupposizioni sono tutte strategie linguistiche che possono essere utilizzate in contesti comunicativi manipolatori per influenzare, persuadere o ingannare gli interlocutori. In questo caso si dice che un enunciato contiene dei contenuti impliciti non "bona fide" veri. Alla luce di quanto detto, dimmi se l'enunciato seguente veicola un contenuto implicito non bona fide vero. Procedi passo dopo passo e produci una risposta "sì/no".
Enunciato: [enunciato che si vuole analizzare]
Presenza di contenuto implicito non bona fide vero:

'An implicature is an inference that a listener can make based on the context and the way a statement is formulated. Implicatures are not explicitly stated but are deduced from the context and conversational norms. Implicatures can be of two main types: conversational and conventional. Conversational implicatures are based on the principles of cooperative communication, described by H.P. Grice in his 1975 essay named "Logic and Conversation". Grice proposed the 'Cooperative Principle,' according to which participants in a conversation expect others to be collaborative and to follow certain maxims: Maxim of Quantity, meaning to provide sufficient information for the comprehensibility of the discourse; Maxim of Quality, meaning to tell the truth, or at least not to say what one believes to be false or for which one has insufficient evidence; Maxim of Relation, meaning to be relevant to the topic

of discourse; Maxim of Manner, meaning to be clear, avoid ambiguity, and be brief and orderly.
Conventional implicatures do not depend on conversational maxims but are tied to linguistic convention and the meaning of words themselves. For example, the use of words like 'but,' 'however,' or 'indeed' introduces a specific type of implicature.
Presupposition is information that is considered given, accepted as true, and shared among the participants in the conversation. Presuppositions are not part of the explicit content of the statement but are implicit and must be true for the statement to make sense. Presuppositions can be tested using the negation test: if the presupposition survives the negation of the statement, it is a presupposition.
Topicalization is a linguistic phenomenon that concerns the structuring of information in a statement, highlighting the subject (or 'topic') of the sentence. In many languages, including Italian, topicalization can alter the normal word order to emphasize what is being talked about. Vagueness is a pragmatic strategy used to express content in a non-specific or indeterminate way. This can be done for various reasons, such as avoiding conflict, maintaining politeness, or leaving room for interpretation.
Implicatures, topicalizations, vagueness, and presuppositions are all linguistic strategies that can be used in manipulative communicative contexts to influence, persuade, or deceive interlocutors. In this case, it is said that a statement contains non-bona fide true implicit content. According to what has been said, tell me if the following statement conveys non-bona fide true implicit content. Proceed step by step and produce a 'yes/no' answer.

Sentence: [sentence to be analysed]
Presence of non-bona fide true implicit content:'

CoT-ZS

Dimmi se l'enunciato seguente veicola un contenuto implicito non bona fide vero. Procedi passo dopo passo e produci una risposta sì/no.
Enunciato: [frase da analizzare]
Presenza di contenuto implicito non bona fide vero:

'Tell me if this sentence conveys a non bona fide true implicit content. Give a yes or no answer only. Proceed step by step and give me a yes/no answer.
Sentence: [sentence to be analysed]
Presence of non bona fide true implicit content:'

Sociodemographic Prompting

Prompt per la generazione della descrizione del parlante:

Scrivi una breve presentazione di [politico], parlando del periodo di attività della sua carriera politica, del partito di appartenenza e delle sue principali idee.

Prompt per il task di detection:

Di seguito troverai la descrizione di una persona e un enunciato. L'enunciato è stato pronunciato dalla persona descritta. Dimmi se l'enunciato ha del contenuto implicito non bona fide vero, anche considerando la persona che lo ha pronunciato e il suo pensiero e orientamento politico. Rispondi solo sì o no.

`Descrizione:` [testo generato col prompt precedente]
`Enunciato:` [enunciato che si vuole analizzare]
`Presenza di contenuto implicito non bona fide vero:`

‘Prompt for generating the speaker's description:
`Write a brief introduction of` [politician]`, discussing the period of their political career, their party affiliation, and their main ideas.`

Prompt for the detection task:
`Below, you will find a description of a person and a statement. The statement was made by the person described. Tell me if the statement contains non-bona fide implicit content, considering also the person who made it and their political views and orientation. Answer only yes or no.`
`Description:` [previously generated description]
`Sentence:` [sentence to be analysed]
`Presence of non-bona fide true implicit content:`’

Generate Knowledge prompting

Prompt per la generazione delle conoscenze:
`Genera cinque diverse conoscenze relative alla pragmatica linguistica. Utilizza il template seguente per la generazione.`

```
Input: {Concetto di Linguistica}
Conoscenza: {Breve ed esaustiva spiegazione del concetto}
```

Prompt per il task di detection:
`Input:` [Conoscenza 1 generata precedentemente]
`Conoscenza:` [Descrizione di conoscenza 1]
`Input:` [Conoscenza 2 generata precedentemente]
`Conoscenza:` [Descrizione di conoscenza 2]
`Input:` [Conoscenza 3 generata precedentemente]
`Conoscenza:` [Descrizione di conoscenza 3]
`Input:` [Conoscenza 4 generata precedentemente]
`Conoscenza:` [Descrizione di conoscenza 4]
`Input:` [Conoscenza 5 generata precedentemente]
`Conoscenza:` [Descrizione di conoscenza 5]
`Basandoti anche sulle conoscenze presentate sopra, identifica se nell'enunciato seguente sono presenti contenuti impliciti non bona fide veri. Rispondi solo sì o no.`

`Enunciato:` [enunciato da analizzare]
`Presenza di contenuto implicito non bona fide vero:`

‘Prompt for generating knowledge:
`Generate five different pieces of knowledge related to pragmatics. Use the following template for generation.`

```
Input: {Linguistic Concept}
Knowledge: {Brief and comprehensive explanation of the concept}
```

Prompt for the detection task:
`Input:` [Knowledge 1 generated previously]
`Knowledge:` [Description of Knowledge 1]

Input: [Knowledge 2 generated previously]
Knowledge: [Description of Knowledge 2]
Input: [Knowledge 3 generated previously]
Knowledge: [Description of Knowledge 3]
Input: [Knowledge 4 generated previously]
Knowledge: [Description of Knowledge 4]
Input: [Knowledge 5 generated previously]
Knowledge: [Description of Knowledge 5]
Based also on the knowledge presented above, identify if the following sentence contains non-bona fide true implicit content. Answer only yes or no.

Sentence: [sentence to be analysed]
Presence of non-bona fide implicit content:'

Classification

Zero-Shot

L'enunciato seguente veicola un contenuto implicito non bona fide vero. Esso può essere un'implicatura o una presupposizione. Dimmi di quale tipo di contenuto implicito non bona fide vero si tratta. Rispondi solo "implicatura" o "presupposizione".
Enunciato: [frase da analizzare]
Tipo di contenuto implicito non bona fide vero presente:

'The following sentence conveys a non bona fide true implicit content. It can be an implicature or a presupposition. Tell me which kind of non bona fide true implicit content it is. Answer with "implicature" or "presupposition" only.
Sentence: [sentence to be analysed]
Type of non bona fide true implicit content:'

Few-Shot

Enunciato: [esempio di enunciato con implicatura]
Tipo di contenuto implicito non bona fide vero: implicatura
Enunciato: [esempio di enunciato con presupposizione]
Tipo di contenuto implicito non bona fide vero: presupposizione.
Enunciato: [esempio di enunciato con implicatura]
Tipo di contenuto implicito non bona fide vero: implicatura
Enunciato: [esempio di enunciato con presupposizione]
Tipo di contenuto implicito non bona fide vero: presupposizione.

Enunciato: [enunciato che si vuole analizzare]
Tipo di contenuto implicito non bona fide vero:

'Sentence: [sample with an implicature]
Type of non bona fide true implicit content: implicature.
Sentence: [sample with a presupposition]
Type of non bona fide true implicit content: presupposition.
Sentence: [sample with an implicature]
Type of non bona fide true implicit content: implicature.
Sentence: [sample with a presupposition]
Type of non bona fide true implicit content: presupposition.
Sentence: [sentence we want to analyse]
Presence of non bona fide true implicit content:'

CoT

L'implicatura è un'inferenza che un ascoltatore può fare basandosi sul contesto e su come viene formulato un enunciato. Le implicature non sono esplicitamente enunciate, ma vengono dedotte dal contesto e dalle norme conversazionali. Le implicature possono essere di due tipi principali: conversazionali e convenzionali. Le implicature conversazionali si basano sui principi della comunicazione cooperativa, descritti da H.P. Grice nel suo saggio "Logic and Conversation" del 1975. Grice ha proposto il "Principio di Cooperazione", secondo il quale i partecipanti a una conversazione si aspettano che gli altri siano collaborativi e seguano certe massime: Massima di quantità, ovvero fornire informazioni sufficienti per la comprensibilità del discorso; massima di qualità, ovvero dire la verità, o almeno non dire ciò che si crede sia falso o di cui non si hanno prove sufficienti; massima di relazione, cioè essere pertinente con l'argomento del discorso; massima di modo, ovvero essere chiari, evitare ambiguità e essere brevi e ordinati.
Le implicature convenzionali non dipendono dalle massime conversazionali ma sono legate alla convenzione linguistica e al significato delle parole stesse. Ad esempio, l'uso di parole come "ma", "però" o "infatti" introduce un tipo specifico di implicatura.
La presupposizione è un'informazione che è considerata data, accettata come vera e condivisa tra i partecipanti alla conversazione. Le presupposizioni non sono parte del contenuto esplicito dell'enunciato ma sono implicite e devono essere vere affinché l'enunciato abbia senso. Le presupposizioni possono essere testate tramite il test della negazione: se la presupposizione sopravvive alla negazione dell'enunciato, è una presupposizione.
La topicalizzazione è un fenomeno linguistico che riguarda la strutturazione dell'informazione in un enunciato, mettendo in evidenza l'argomento (o "topic") della frase. In molte lingue, inclusa l'italiano, la topicalizzazione può alterare l'ordine normale delle parole per sottolineare ciò di cui si sta parlando.La vaghezza è una strategia pragmatica utilizzata per esprimere contenuti in modo non specifico o indeterminato. Questo può essere fatto per vari motivi, come evitare conflitti, mantenere la cortesia, o per lasciare spazio all'interpretazione.Le implicature, le topicalizzazioni, la vaghezza e le presupposizioni sono tutte strategie linguistiche che possono essere utilizzate in contesti comunicativi manipolatori per influenzare, persuadere o ingannare gli interlocutori. In questo caso si dice che un enunciato contiene dei contenuti impliciti non "bona fide" veri. Basandoti su quanto detto sopra, analizza l'enunciato seguente. Esso contiene un contenuto implicito non bona fide vero che può essere una presupposizione o una implicatura. Procedi passo dopo passo e dimmi se si tratta di una implicatura o una presupposizione. Rispondi solo con "implicatura" o "Presupposizione".

Enunciato: [enunciato che si vuole analizzare]
Presenza di contenuto implicito non bona fide vero:

'An implicature is an inference that a listener can make based on the context and the way a statement is formulated. Implicatures are not explicitly stated but are deduced from the context and conversational norms. Implicatures can be of two main types: conversational and conventional. Conversational implicatures are based on the principles of cooperative communication, described by H.P. Grice in his 1975 essay named "Logic and Conversation". Grice proposed the 'Cooperative Principle,' according to which participants in a conversation expect others to be collaborative and to follow certain maxims: Maxim of Quantity, meaning to provide sufficient information for the comprehensibility of the discourse;

Maxim of Quality, meaning to tell the truth, or at least not to say what one believes to be false or for which one has insufficient evidence; Maxim of Relation, meaning to be relevant to the topic of discourse; Maxim of Manner, meaning to be clear, avoid ambiguity, and be brief and orderly.
Conventional implicatures do not depend on conversational maxims but are tied to linguistic convention and the meaning of words themselves. For example, the use of words like 'but,' 'however,' or 'indeed' introduces a specific type of implicature.
Presupposition is information that is considered given, accepted as true, and shared among the participants in the conversation. Presuppositions are not part of the explicit content of the statement but are implicit and must be true for the statement to make sense. Presuppositions can be tested using the negation test: if the presupposition survives the negation of the statement, it is a presupposition.
Topicalization is a linguistic phenomenon that concerns the structuring of information in a statement, highlighting the subject (or 'topic') of the sentence. In many languages, including Italian, topicalization can alter the normal word order to emphasize what is being talked about. Vagueness is a pragmatic strategy used to express content in a non-specific or indeterminate way. This can be done for various reasons, such as avoiding conflict, maintaining politeness, or leaving room for interpretation.
Implicatures, topicalizations, vagueness, and presuppositions are all linguistic strategies that can be used in manipulative communicative contexts to influence, persuade, or deceive interlocutors. In this case, it is said that a statement contains non-bona fide true implicit content. According to what has been said, analyze the following sentence. It contains a non bona fide true implicit content. It can be an implicature or a presupposition. Proceed step by step and tell me if it is an implicature or a presupposition. Answer with only "implicature" or "presupposition".

Sentence: [sentence to be analyzed]
Type of non-bona fide true implicit content:'

CoT-ZS

L'enunciato seguente veicola un contenuto implicito non bona fide vero. Esso può essere un'implicatura o una presupposizione. Dimmi di quale tipo di contenuto implicito non bona fide vero si tratta. Procedi passo dopo passo e produci una risposta del tipo "implicatura" o "presupposizione".
Enunciato: [frase da analizzare]
Tipo di contenuto implicito non bona fide vero:

'Tell me if this sentence conveys a non bona fide true implicit content. Give a yes or no answer only. Proceed step by step and give me a "implicature" or "presupposition" answer.
Sentence: [sentence to be analysed]
Type of non bona fide true implicit content:'

Sociodemographic Prompting

Prompt per la generazione della descrizione del parlante:
Scrivi una breve presentazione di [politico], parlando del periodo di attività della sua carriera politica, del partito di appartenenza e delle sue principali idee.

Prompt per il task di detection:
Di seguito troverai la descrizione di una persona e un enunciato. L'enunciato è stato pronunciato dalla persona descritta. L'enunciato contiene un contenuto implicito non bona fide vero che può essere una implicatura o una presupposizione. Dimmi se si tratta di una implicatura o di una presupposizione. Rispondi solo con "implicatura" o "Presupposizione".

Descrizione: [testo generato col prompt precedente]
Enunciato: [enunciato che si vuole analizzare]
Tipo di contenuto implicito non bona fide vero:

'Prompt for generating the speaker's description:
Write a brief introduction of [politician], discussing the period of their political career, their party affiliation, and their main ideas.

Prompt for the detection task:
Below, you will find a description of a person and a statement. The statement was made by the person described and contains a non bona fide true implicit content. It can be an implicature or a presupposition. Tell me if it is an implicature or a presupposition, considering the person who said the sentence and their political views and orientation. Answer with only "implicature" or "presupposition".
Description: [previously generated description]
Sentence: [sentence to be analysed]
Type of non-bona fide true implicit content:'

Generate Knowledge prompting

Prompt per la generazione delle conoscenze:

```
Genera cinque diverse conoscenze relative alla pragmatica linguistica. Utilizza il template seguente per la generazione.

Input: {Concetto di Linguistica}
Conoscenza: {Breve ed esaustiva spiegazione del concetto}
```

Prompt per il task di detection:

`Input:` [Conoscenza 1 generata precedentemente]
`Conoscenza:` [Descrizione di conoscenza 1]
`Input:` [Conoscenza 2 generata precedentemente]
`Conoscenza:` [Descrizione di conoscenza 2]
`Input:` [Conoscenza 3 generata precedentemente]
`Conoscenza:` [Descrizione di conoscenza 3]
`Input:` [Conoscenza 4 generata precedentemente]
`Conoscenza:` [Descrizione di conoscenza 4]
`Input:` [Conoscenza 5 generata precedentemente]
`Conoscenza:` [Descrizione di conoscenza 5]
`Basandoti anche sulle conoscenze presentate sopra, analizza la frase seguente. Essa contiene un contenuto implicito non bona fide vero che può essere una implicatura o una presupposizione. Dimmi se si tratta di una implicatura o una presupposizione. Rispondi solo “Implicatura” o “presupposizione”.`
`Enunciato:` [enunciato da analizzare]
`Tipo di contenuto implicito non bona fide vero:`

‘**Prompt for generating knowledge:**

```
Generate five different pieces of knowledge related to pragmatics. Use the following template for generation.

Input: {Linguistic Concept}
Knowledge: {Brief and comprehensive explanation of the concept}
```

Prompt for the detection task:

`Input:` [Knowledge 1 generated previously]
`Knowledge:` [Description of Knowledge 1]
`Input:` [Knowledge 2 generated previously]
`Knowledge:` [Description of Knowledge 2]
`Input:` [Knowledge 3 generated previously]
`Knowledge:` [Description of Knowledge 3]
`Input:` [Knowledge 4 generated previously]
`Knowledge:` [Description of Knowledge 4]
`Input:` [Knowledge 5 generated previously]
`Knowledge:` [Description of Knowledge 5]
`Based also on the knowledge presented above, analyse the following sentence. It contains a non-bona fide true implicit content that can be an implicature or a presupposition. Tell me if it is an implicature or a presupposition. Answer with “Implicature” or “Presupposition” only.`

`Sentence:` [sentence to be analysed]
`Presence of non-bona fide implicit content:`’