LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

EXPLORING THE PRAGMATIC COMPETENCE OF AI CHATBOTS IN UNDERSTANDING IMPLICATURE AND PRESUPPOSITION

Bushra Abeer[*1], **Sidra Bibi**[2]

[*1]*BS English Scholar, Department of English, Bahauddin Zakariya University*
[2]*M. Phil Scholar, Department of Humanities, Comsat University Islamabad, Wah Campus*

[*1]bushraabeer0@gmail.com, [2]melthesis512@gmail.com

DOI: https://doi.org/10.5281/zenodo.23724484

Keywords
pragmatic competence, AI chatbots, conversational implicature, presupposition, contextual interpretation, Gricean pragmatics, computational linguistics

Article History
Received: *25 June 2026*
Accepted: *16 September 2026*
Published: *28 October 2026*

Copyright @Author
Corresponding Author:
Bushra Abeer

Abstract

Artificial intelligence (AI) chatbots have transformed human–computer communication, but their ability to interpret implied meanings and presupposed information remains an important question in pragmatics. This study examines the pragmatic competence of AI chatbots in understanding conversational implicature and presupposition, focusing on contextual interpretation, speaker intention, and inferential reasoning. The research employs a qualitative, analysis-based design that evaluates representative chatbot responses to selected conversational scenarios. The theoretical framework draws on three influential works: Grice's (1975) account of conversational implicature, Levinson's (1983) foundational treatment of pragmatics, and Stalnaker's (1974) theory of presupposition and common ground. The analysis considers indirect requests, scalar implicatures, presupposition triggers, ambiguous utterances, irony, and context-dependent expressions. It suggests that AI chatbots can often identify conventional linguistic patterns and generate plausible interpretations when contextual information is explicit. However, interpretation becomes less reliable when meaning depends on cultural knowledge, unstated assumptions, irony, or competing conversational intentions. These observations highlight the distinction between producing contextually appropriate language and demonstrating robust pragmatic competence. The study contributes to discussions of computational pragmatics by identifying the importance of contextual modeling, inferential consistency, and sensitivity to speaker assumptions. Its implications extend to natural language processing, chatbot evaluation, and language education. Because the study is conceptual and analysis-based rather than an empirical benchmark experiment, its observations should be treated as theoretically informed analytical conclusions requiring systematic validation through controlled testing.

INTRODUCTION

The emergence of AI has revolutionized the way humans interact with digital systems. Modern chatbots can respond to inquiries, clarify intricate concepts, summarize documents, and engage in dialogues that closely

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
LJLLR
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

mimic human interaction. Their fluency has inspired researchers to look into the possibility of these systems understanding the meaning of language beyond the literal sense of the words. Semantic coherence and grammatical correctness are crucial components of language processing, but successful communication also involves understanding the speaker's implication, assumption, and intention in certain contexts.

Pragmatics is one of the major areas of linguistics which explores how meaning is communicated and interpreted in the context. Pragmatics deals with the relationship between expressions, speaker, listener and communicative situation; while Semantics focuses on the conventional meaning of expressions. A sentence can contain information not mentioned, but an utterance can contain information that is assumed as present. These phenomena are especially relevant in day-to-day communication, in which speakers often need not report all information explicitly, but may instead depend on shared background information.

Conversational implicature and presupposition are two important concepts in pragmatic analysis. Conversational implicature is an effect whereby the speaker implies something more than the literal meaning of what he or she actually says. As a result, for instance, if a student asks, “Did you like the lecture?” and gets a response such as, “The slides were nice,” the student might deduce that the lecture wasn't that much fun. This interpretation is not verbalized; it comes from the conversation context and assumptions about the relevance and intentions of the speaker.

Presupposition, on the other hand, is about what an utterance presupposes as already accepted or available in the background of the conversation. Think about this sentence: “Sara has discontinued the use of the chatbot.” The word ‘stopped' conveys the typical style of a verb, implying that Sara has used the chatbot before. The sentence conveys information about a change that is currently happening, and earlier use is regarded as background. Understanding this distinction is crucial to understanding speakers' structuring of information and common ground management.

The tricky part with AI chatbots is that some meaning can't be conveyed just by the grammar or word by word. An utterance can be understood in multiple ways and the right interpretation can be a function of the social relationship, cultural expectations, turn of the previous speaker, or the speaker's communicative intention. A pattern-recognizing chatbot can be confused by the intent of the pattern.

The theory of conversational implicature, by Grice (1975), underlies many explanations and studies of how listeners draw meaning from the assumptions made by speakers in cooperative communication. Levinson (1983) provides a thorough discussion of the concepts of pragmatics, which sets the stage for the theoretical framework of understanding context, inference and pragmatic meaning. How communication relies on assumptions that are shared, presumed, or assumed together by interlocutors is explained by Stalnaker (1974), who discussed presupposition and common ground. In sum, these works offer a theoretical framework for the pragmatic performance of AI chatbots.

The re-emergence of the issue is due to recent developments in large language models. Such models produce output based upon learned linguistic patterns and context, but without fluency, it is not sufficient to demonstrate that they are consistent in their interpretation of implied meaning in various contexts. Bender et al. (2021) note that they are not the same thing as convincing linguistic performance, which does not necessarily imply human-like language understanding. Likewise, computational research on pragmatic reasoning remains an exploration of the interaction between the linguistic form and communicative context (Frank & Goodman, 2012; Hu et al., 2023).

This study looks at the understanding of implicature and presupposition in AI chatbots using a qualitative and theory-driven analysis of selected conversational scenarios. It is about the ability of a chatbot to get the drift, to pick up on assumptions, and to appease to unclear or context-dependent utterances. The study does not attempt to validate the performance of any specific chatbot through experiments but rather aims

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

create an analytical framework for grasping strengths and weaknesses which will be tested in future experiments.

Problem Statement

While AI chatbots are becoming more adept in their interactions, evaluating their actual ability is challenging with standard conversational skills.As the AI chatbots become more fluent with their interactions, it becomes harder to measure their competency in a pragmatic sense with regular conversation. The response might seem like it is in the right style, but not address the speaker's true meaning or assumptions in the speech. This distinction is particularly relevant in cases of implicatures or presuppositions that often rely on context for their interpretation, rather than being explicitly communicated through linguistic structure.

While there are existing theoretical models of how humans make inferences about implied meaning and how humans tolerate background assumptions, there are ongoing challenges in implementing these models in computational systems. AI chatbots can correctly recognize certain familiar presupposition triggers, or common indirect requests, but fail to recognize irony, cues for cultural references, and competing conversational implications inconsistently. The central question, then, concerns the possibility of assessing pragmatic competence without mixing out the linguistic competence with valid interpretation of the context.

Research Objectives

The study pursues five objectives:

1. To explain the theoretical relationship between pragmatic competence, conversational implicature, and presupposition.
2. To examine how AI chatbots may interpret indirect meanings in representative conversational contexts.
3. To analyze the identification of presupposed information through linguistic triggers and contextual assumptions.
4. To explore potential limitations in chatbot interpretation of ambiguity, irony, cultural references, and speaker intentions.
5. To develop a theory-informed framework for evaluating pragmatic competence in AI-mediated communication.

Research Questions

The study addresses the following questions:

1. How can pragmatic competence in AI chatbots be understood through established pragmatic theories?
2. How do chatbot responses reflect the interpretation of conversational implicatures?
3. How effectively can AI chatbots identify presupposed information in context-sensitive utterances?
4. What challenges arise when implicature and presupposition depend on ambiguity, irony, or shared cultural knowledge?
5. What implications do these challenges have for computational pragmatics and chatbot evaluation?

Significance of the Study

This study contributes to computational pragmatics by connecting established linguistic theories with the evaluation of AI-generated communication. Its framework can help researchers distinguish literal interpretation from contextual inference and identify specific conditions under which chatbot responses require closer examination.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

The study also has implications for language education, particularly in teaching learners to recognize implied meanings, presupposed assumptions, and context-dependent interpretations. For developers, it highlights the value of evaluating conversational systems across diverse pragmatic situations rather than relying exclusively on fluency or grammatical correctness.

Literature Review

Pragmatic competence refers to the ability to interpret and use language appropriately in particular communicative situations. It involves understanding how speakers communicate intentions, how listeners infer information beyond literal expressions, and how contextual assumptions influence interpretation. In human communication, pragmatic competence supports indirect requests, polite refusals, conversational implicatures, presuppositions, humor, irony, and the negotiation of shared knowledge.

For AI chatbots, pragmatic competence presents a distinctive challenge. A chatbot may produce a grammatically correct response without identifying the communicative purpose of the preceding utterance. For example, if a user says, "It is rather cold in here," the statement may simply describe the temperature. In a different context, however, it may function as an indirect request to close a window. Interpreting the statement requires sensitivity to the situation and the relationship between the participants.

Modern language models can use preceding conversational turns to generate contextually relevant responses. Nevertheless, the presence of contextual information does not guarantee that a model will consistently select the intended interpretation. Hu et al. (2023) compared pragmatic language understanding in humans and language models and found that model performance varies across pragmatic phenomena, particularly when social expectations are involved. Their findings support the need to evaluate different forms of pragmatic reasoning separately rather than treating language understanding as a single, uniform ability.

Bender et al. (2021) similarly question the assumption that fluent linguistic output necessarily demonstrates an understanding comparable to human language comprehension. Their critique is relevant to pragmatic analysis because a chatbot may generate a convincing explanation of an implied meaning without reliably identifying that meaning in unfamiliar contexts.

Pragmatic competence in AI should therefore be examined through observable interpretive performance. Relevant questions include whether a chatbot identifies plausible implications, recognizes presupposed information, distinguishes literal from intended meaning, and revises its interpretation when contextual evidence changes.

Grice's (1975) account of conversational implicature is one of the most influential theoretical foundations for understanding implied meaning. Grice proposes that conversational participants generally interpret one another against an assumption of cooperative communication. His Cooperative Principle states that contributions are expected to be appropriate to the accepted purpose and direction of the conversation.

Grice organizes conversational expectations into four categories: Quantity, Quality, Relation, and Manner.

- Quantity: Provide as much information as the situation requires, but not more than necessary.
- Quality: Avoid statements believed to be false or unsupported.
- Relation: Make contributions relevant to the conversation.
- Manner: Communicate clearly and avoid unnecessary obscurity or ambiguity.

These maxims are not absolute rules that speakers must always follow. Rather, they help explain how listeners interpret utterances when a speaker appears to communicate indirectly or departs from an expected conversational pattern.

Consider the following exchange:

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

Question: "Did all the students pass the examination?"
Response: "Some students passed."
The response may lead the listener to infer that not all students passed. However, the word some does not logically exclude all. The inference arises because the speaker could have used the stronger expression all but chose a weaker alternative. Depending on the context, the inference may be reasonable, defeasible, or inappropriate.
For an AI chatbot, the challenge is not simply recognizing the word some. It must determine whether the stronger alternative is relevant, whether the context supports the inference, and whether the speaker intends to communicate a limitation.
Cho and Kim (2024) investigated scalar implicature in language models and reported differences in how models handled implicatures when contextual information and questions under discussion were introduced. Their findings illustrate why pragmatic interpretation must be assessed in relation to both linguistic expressions and contextual conditions.
Another example involves an indirect request:
Speaker A: "Could you open the window?"
Speaker B: "Yes, I could."
The literal response addresses the speaker's ability to open the window. Yet in ordinary conversation, Speaker A is generally requesting an action rather than asking about physical ability. A chatbot that treats the exchange literally may misunderstand the intended speech act.
The same utterance can also receive different interpretations depending on context. In a discussion about physical ability, "Could you open the window?" might genuinely ask whether a person is capable of doing so. In a classroom or office, it may function as a polite request.
Grice's framework helps explain these distinctions because it links interpretation to assumptions about communicative purpose. Applied to AI chatbots, it provides a basis for evaluating whether a system recognizes implied meanings, identifies relevant contextual evidence, and avoids treating every conversational inference as certain.
However, Grice's theory also has limitations as a complete model of computational communication. Human interpretations may depend on social hierarchies, emotional relationships, cultural conventions, and shared experiences. The Cooperative Principle does not independently specify how an AI system should represent every such factor. It is therefore best used as a theoretical framework rather than as a complete computational solution.
Levinson's (1983) Pragmatics offers a broad theoretical account of the relationship between linguistic meaning, context, inference, and communication. The work is particularly valuable because it situates implicature and presupposition within a larger account of how language users interpret utterances.
A central insight in pragmatic theory is that speakers do not need to express every relevant piece of information explicitly. Listeners use linguistic conventions, conversational expectations, and contextual knowledge to recover meanings that extend beyond sentence structure.
Levinson's discussion of generalized and particularized conversational implicatures is relevant to AI systems. Generalized implicatures tend to arise in relatively ordinary contexts, while particularized implicatures depend more heavily on specific circumstances.
For example, "Some of the guests arrived" may ordinarily suggest that not all guests arrived. This is a generalized inference that can be cancelled if the speaker adds, "In fact, all of them arrived." By contrast, interpreting "The lights are still on" as a request to switch them off depends on the surrounding situation and is therefore more context-sensitive.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

A

chatbot may successfully identify common generalized implicatures because they occur frequently in ordinary language. More specific inferences, however, may require information that is not explicitly included in the prompt. If the model assumes that a statement is a request when it is actually a neutral observation, it may produce an inappropriate response.

Levinson's framework also helps distinguish pragmatic inferences from logical entailments. An entailment follows from the truth conditions of a statement, whereas an implicature is an inferred meaning that may be withdrawn or cancelled without contradiction.

Consider:

"Some applicants met the requirements, and perhaps all of them did."

The second clause is compatible with the first because some does not logically mean some but not all. If a chatbot treats the exclusion of all as an unavoidable semantic fact, it fails to distinguish a defeasible pragmatic inference from literal meaning.

This distinction matters in educational, legal, and professional settings. A chatbot summarizing an institutional report, for example, should not convert an uncertain implication into a definite claim. A reliable system must preserve the difference between what a source explicitly states and what readers may reasonably infer.

Levinson's work thus provides an analytical foundation for examining the range of inferences that AI chatbots make and the conditions under which these inferences are justified.

Stalnaker's (1974) account of presupposition emphasizes the role of shared assumptions in communication. Conversations proceed against a background of propositions that participants accept, provisionally assume, or treat as mutually established. This background is commonly described as the common ground.

Presuppositions are important because utterances frequently introduce new information while treating other information as already available. For example:

"Maria has returned to university."

The verb returned suggests that Maria attended the university previously or had some earlier association with it. The speaker presents the return as the main development while leaving the earlier association in the background.

A chatbot interpreting this sentence should distinguish between the explicitly asserted proposition—that Maria has returned—and the presupposed information concerning her previous association with the university. It should also avoid assuming that every presupposition is necessarily true in the real world. The speaker may be mistaken, or the presupposition may be challenged by later information.

Stalnaker's account is especially relevant to AI because chatbots must track information introduced during earlier conversational turns. Consider the following exchange:

User: "My sister recently started a new job."

Chatbot: "Where does she work?"

User: "She works at a hospital."

Chatbot: "Has she been working there for many years?"

The final question conflicts with the information that she recently started the job, unless the chatbot has a reason to distinguish her current employment from an earlier period of work at the same hospital. This illustrates the importance of maintaining a coherent conversational record.

The challenge becomes more complex when presuppositions are embedded in questions or descriptions. For example, "When did you stop using the service?" presupposes that the person previously used it and has

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

now stopped. If the person never used the service, the question contains a questionable background assumption.

A pragmatically competent chatbot should identify this assumption and, when necessary, seek clarification rather than accepting it automatically.

Stalnaker's approach also explains why common ground should not be equated with universally shared knowledge. A proposition may be common ground between two participants but unknown to an outsider. Similarly, a chatbot should not infer a user's personal history merely because a sentence grammatically presupposes it.

For AI evaluation, the relevant questions are whether a model can recognize the background assumptions triggered by an expression, track those assumptions across turns, and revise its interpretation when a participant rejects them.

Recent research increasingly treats pragmatics as a distinct area of language-model evaluation. This development is important because traditional language benchmarks may not reveal whether a system distinguishes literal meaning from conversational inference.

Sravanthi et al. (2024) introduced the Pragmatics Understanding Benchmark (PUB), covering implicature, presupposition, reference, and deixis through multiple-choice tasks. Their evaluation identified substantial variation across pragmatic phenomena and a gap between model and human performance. This supports the view that a chatbot's overall linguistic competence cannot be inferred from success in one pragmatic category.

Cong (2024) examined manner implicatures in large language models and reported limitations in evidence that pretraining alone produces robust competence in these inferences. The study reinforces the importance of distinguishing the ability to generate plausible language from the ability to interpret subtle differences in communicative meaning.

More recent work has also examined pragmatic awareness in actual conversational interactions. Andersson and McIntyre (2025), for example, investigated ChatGPT's recognition of impoliteness. They reported that the model could recognize some implied impoliteness but was less precise in non-impolite cases such as banter and encountered difficulties with register and contextual appropriateness. Although impoliteness is not identical to implicature or presupposition, the findings demonstrate how pragmatic interpretation depends on distinctions that are difficult to capture through surface linguistic patterns alone.

Ma et al. (2025) reviewed datasets and evaluation approaches for pragmatics in large language models. Their survey identifies continuing challenges in the design of comprehensive benchmarks and the assessment of nuanced language use. It also provides a foundation for evaluating pragmatic abilities across multiple phenomena rather than relying on isolated examples.

Taken together, these studies suggest that pragmatic performance should be assessed through controlled tasks that vary context, linguistic form, and communicative intention. They also show why conclusions about AI pragmatic competence should be based on documented model responses rather than general impressions of conversational fluency.

Theoretical Framework

The study integrates three foundational theoretical texts: Grice (1975), Levinson (1983), and Stalnaker (1974). Each contributes a distinct analytical perspective.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

Grice: Inference and Cooperative Communication

Grice's framework is used to examine how chatbots interpret meanings that are not directly expressed. The analysis considers whether a response is relevant to the preceding question, whether the wording supports a plausible inference, and whether alternative interpretations are available.

The framework is applied to indirect requests, scalar expressions, and conversational responses that appear incomplete or indirect.

Levinson: Contextual Meaning and Pragmatic Inference

Levinson's account provides the basis for distinguishing generalized implicatures from context-dependent inferences. It also supports the distinction between entailment and implicature.

This perspective helps evaluate whether a chatbot treats an inference as defeasible or incorrectly presents it as a necessary consequence of the utterance.

Stalnaker: Presupposition and Common Ground

Stalnaker's theory is used to identify background assumptions and assess how these assumptions are maintained or challenged during conversation.

The analysis examines presupposition triggers, consistency across conversational turns, and the handling of information that a speaker may not accept.

Integrated Analytical Model

Together, the three theories establish four analytical dimensions:

1. Literal and implied meaning: Does the chatbot distinguish the explicit proposition from a plausible conversational inference?
2. Contextual sensitivity: Does the interpretation change appropriately when the conversational circumstances change?
3. Presupposition recognition: Does the chatbot identify background assumptions and distinguish them from asserted information?
4. Interpretive flexibility: Does the chatbot acknowledge uncertainty and revise an interpretation when contrary evidence appears?

These dimensions provide a coherent framework for examining pragmatic performance without assuming that a single response establishes general competence.

Methodology

Research Design

The study adopts a qualitative, analysis-based research design. It uses established pragmatic theories and representative conversational scenarios to develop a structured examination of the interpretive demands placed on AI chatbots.

The design is analytical rather than experimental. It does not claim to measure the accuracy of a particular chatbot through a newly collected dataset, nor does it present simulated responses as observed empirical results. Instead, it identifies the interpretive requirements that can be used to evaluate chatbot performance and illustrates how different responses may be assessed against the theoretical framework.

This approach is appropriate for developing a conceptual evaluation model. However, its conclusions concern theoretically anticipated strengths, limitations, and evaluation criteria rather than experimentally established performance rates.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
LJLLR
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

Data and Illustrative Materials

The analytical materials consist of representative conversational examples designed to illustrate implicature and presupposition. They include:

- Scalar expressions, such as some, many, and all.
- Indirect requests and indirect responses.
- Presupposition triggers, including stop, return, again, and continue.
- Ambiguous utterances whose interpretation depends on contextual information.
- Examples involving irony, social expectations, and conflicting background assumptions.

The examples are illustrative analytical materials rather than a collected corpus of chatbot outputs. They are intended to clarify the theoretical distinctions and demonstrate how a systematic evaluation could be conducted.

Analytical Procedure

The analysis follows four stages.

First, each example is classified according to its principal pragmatic phenomenon. This establishes whether the central interpretive issue concerns implicature, presupposition, or the interaction between them.

Second, the expected interpretation is identified using the three theoretical frameworks. Grice's Cooperative Principle supports the analysis of conversational inferences; Levinson's account informs the distinction between generalized and context-dependent implicatures; and Stalnaker's theory guides the identification of presupposed information and common ground.

Third, alternative interpretations are considered. This stage is particularly important because pragmatic inferences are often defeasible. A response is not necessarily incorrect merely because it selects an interpretation different from the most common one. Its appropriateness depends on whether that interpretation is supported by the available context.

Finally, the examples are examined according to the four analytical dimensions: distinction between literal and implied meaning, contextual sensitivity, presupposition recognition, and interpretive flexibility.

Analytical Criteria

The framework uses the following criteria to organize the evaluation.

Criterion	Central question	Theoretical basis
Implicature recognition	Is an implied meaning identified where context supports it?	Grice (1975)
Contextual interpretation	Does the interpretation reflect the conversational situation?	Levinson (1983)
Presupposition identification	Is background information distinguished from asserted content?	Stalnaker (1974)
Consistency	Does the interpretation remain coherent across conversational turns?	Stalnaker (1974)
Flexibility	Can the interpretation be revised when contextual evidence changes?	Integrated framework

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
LJLLR
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

Criterion	Central question	Theoretical basis

Scope and Limitations of the Method

The analysis is limited to English-language examples and selected pragmatic phenomena. It does not establish whether a particular commercial chatbot outperforms another, and it does not estimate an accuracy percentage.

A future empirical study could apply the framework to a balanced dataset of chatbot responses, include multiple models and human annotators, and report accuracy, inter-annotator agreement, and error patterns. Such a design would provide evidence about the extent to which the theoretical expectations developed here correspond to actual system performance.

Analysis and Discussion

Interpretation of Conversational Implicature

Conversational implicature presents a major test of pragmatic competence because the intended meaning is not always explicitly encoded in the words. The listener must connect the utterance to the conversational purpose and decide whether an additional inference is warranted.

Consider the following illustrative exchange:

Question: "Did you like the research presentation?"

Response: "The slides were well organized."

The response provides positive information about one aspect of the presentation but does not directly evaluate the presentation as a whole. In ordinary conversation, the listener may infer that the speaker has reservations about the presentation. However, this inference is not inevitable. The speaker may genuinely have liked the slides and the presentation, or may be avoiding a direct answer for reasons unrelated to dissatisfaction.

A chatbot that automatically interprets the response as negative would overstate the evidence. A more careful interpretation would identify the possible implicature while acknowledging that the context does not establish the speaker's overall evaluation.

This example demonstrates the difference between recognizing an inference and treating it as certain. Grice's theory helps explain why the response may invite an indirect interpretation, while Levinson's framework highlights the importance of contextual conditions in determining whether the inference is justified.

A related problem arises with scalar implicatures. Consider:

"The lecturer answered some of the questions."

A common inference is that the lecturer did not answer all the questions. Yet the statement remains logically compatible with the possibility that every question was answered. If the context later establishes that the lecturer answered all questions, the initial inference must be revised.

Research on scalar implicature suggests that language models can exhibit some sensitivity to these inferences, but performance varies with task design and contextual information (Cho & Kim, 2024). This finding supports a cautious evaluation strategy: models should be tested not only on whether they select a familiar inference but also on whether they recognize when that inference is cancellable.

The same principle applies to indirect requests. When a user says, "I cannot see the screen clearly," the utterance may function as a request to adjust the display, increase the font size, or change the seating arrangement. A chatbot should consider the context before deciding which action is intended. If no context

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

is

available, asking a clarifying question may be more appropriate than selecting an unsupported interpretation.

These examples suggest that pragmatic evaluation should reward contextual justification rather than merely matching a single predetermined interpretation. A model demonstrates stronger interpretive performance when it identifies plausible meanings, distinguishes them from certainty, and revises them in response to additional information.

Understanding Presupposition

Presupposition introduces a different interpretive challenge because background assumptions can remain implicit even when they are central to the meaning of an utterance.

Consider the statement:

"Ahmed has started using the application again."

The expression again indicates a repeated action and suggests that Ahmed used the application previously. The main assertion concerns his current renewed use, whereas the earlier use is treated as background information.

A chatbot may identify the presupposition through the lexical trigger. However, recognizing the trigger is not sufficient to establish that the system can manage the assumption appropriately. It must also distinguish the speaker's assumption from verified information.

For instance, if a user states, "Ahmed has started using the application again, although he never used it before," the utterance contains a conflict between the presupposition associated with again and the subsequent denial of earlier use. A careful chatbot should recognize the inconsistency and ask for clarification or explain the competing interpretations.

This is where Stalnaker's concept of common ground becomes particularly useful. A presupposition may be treated as accepted background information in a conversation, but conversational participants can challenge it. A chatbot should not automatically transform every presupposition into an established fact about the world.

A similar issue arises in questions containing definite descriptions:

"When did the principal announce the revised policy?"

The question normally assumes that the principal announced such a policy. If the user has not established that an announcement occurred, a chatbot should avoid inventing a date or treating the event as confirmed. It may instead state that the question presupposes an announcement and request supporting context.

This distinction is particularly important in legal, educational, and administrative settings. A question may contain an inaccurate assumption, and accepting that assumption uncritically can lead to misleading answers. Pragmatic competence therefore includes the ability to identify presupposed information while maintaining a distinction between assumed and verified facts.

Contextual Sensitivity and Ambiguity

Contextual sensitivity involves recognizing that an utterance can support different interpretations in different circumstances. The same words may function as an observation, a complaint, a request, or an ironic comment.

Consider:

"You have arrived early."

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

In a formal meeting where punctuality is valued, the statement may express approval. In a situation where the meeting has been delayed and the speaker is being sarcastic, it may communicate criticism. Without contextual information, neither interpretation is guaranteed.

A chatbot that consistently selects one interpretation may appear confident but fail to account for the variability of pragmatic meaning. A more reliable approach would identify the literal content first, then consider the contextual evidence supporting alternative interpretations.

Ambiguity also arises when speakers use indirect criticism. "That was an interesting decision" could express genuine curiosity, cautious approval, or disapproval. Its interpretation depends on intonation in spoken communication, the relationship between speakers, and the wider conversational context. Text-based chatbots often lack access to these signals unless they are explicitly described.

Hu et al. (2023) found that pragmatic performance varies across tasks, with particular challenges where social expectations influence interpretation. This supports the need to assess contextual ambiguity as a distinct dimension of chatbot evaluation rather than treating it as a minor variation in language comprehension.

Ambiguity should not automatically be classified as model failure. In some situations, the available evidence genuinely supports multiple interpretations. The more appropriate criterion is whether the chatbot recognizes the uncertainty and responds proportionately to the evidence.

Irony, Sarcasm, and Social Meaning

Irony and sarcasm are especially challenging because the intended meaning may diverge from the literal meaning. Consider the statement:

"Excellent work—you deleted the entire document."

The phrase excellent work is ordinarily positive, but the second clause creates a context in which the praise is likely ironic. The intended message is probably criticism rather than approval.

A chatbot may identify this contrast when the context is explicit. Nevertheless, the same phrase may be sincere in a different situation. Consequently, pragmatic evaluation should include both clear examples and borderline cases in which the intended stance is less certain.

Cong's (2024) findings concerning manner implicatures show that subtle pragmatic distinctions remain difficult for language models to handle consistently. Similarly, Andersson and McIntyre (2025) report that ChatGPT can recognize some implied impoliteness while struggling with banter and register-sensitive interpretation. These findings suggest that pragmatic competence involves more than identifying negative words or familiar linguistic patterns.

The implications extend to online communication, where users frequently rely on humor, indirect criticism, understatement, and shared cultural references. Misinterpreting such expressions may cause a chatbot to provide an unnecessarily serious response, misidentify a friendly exchange as hostile, or fail to recognize genuine dissatisfaction.

A stronger evaluation framework would therefore distinguish between recognition of clear sarcasm, interpretation of context-dependent irony, and identification of ambiguous social meaning. It would also examine whether a chatbot can explain the evidence behind its interpretation rather than merely assigning a label.

The Interaction Between Implicature and Presupposition

Implicature and presupposition are analytically distinct, but they can interact within the same utterance. Consider:

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
LJLLR
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

"Even Nadia managed to complete the assignment."

The word even commonly triggers the presupposition that Nadia's completion was unexpected or that she was a relatively unlikely person to complete the task. Depending on the context, the sentence may also implicate that the assignment was difficult or that other people were expected to complete it more easily.

A chatbot must therefore distinguish between the background assumption associated with even and the additional inference that the assignment was difficult. The latter depends more heavily on contextual expectations and is not necessarily established by the sentence alone.

Another example is:

"Why did the manager finally approve the proposal?"

The word finally suggests that approval occurred after a delay or period of waiting. The question presupposes approval and may also invite an inference that the decision was overdue. However, the exact reason for the delay cannot be inferred from the sentence alone.

These examples demonstrate why a chatbot may identify one pragmatic phenomenon while missing another. Recognizing a presupposition trigger does not guarantee that the model will correctly identify a related implicature. Likewise, generating a plausible implicature does not establish that the model has correctly tracked the utterance's background assumptions.

The PUB benchmark developed by Sravanthi et al. (2024) is relevant here because it treats implicature and presupposition as separate evaluation categories while situating them within a broader set of pragmatic tasks. Its findings indicate that model performance can vary substantially across these categories.

Future evaluations should therefore include examples in which multiple pragmatic phenomena interact. Such tests would provide a more realistic picture of the interpretive demands of ordinary communication.

Consistency Across Conversational Turns

Pragmatic competence is not limited to interpreting isolated sentences. In extended conversations, participants build on earlier information, revise assumptions, and correct misunderstandings.

For example:

User: "I have never visited Istanbul."

Chatbot: "What did you enjoy most about your last visit?"

The chatbot's question conflicts with the user's statement. The error is not simply a failure to recognize a presupposition trigger; it also reflects inadequate use of information established in the conversation.

A more appropriate response would ask whether the user plans to visit Istanbul or discuss what they would like to see there. The system should not presuppose a previous visit when the available conversational record explicitly denies one.

A related difficulty arises when users correct earlier statements. Suppose a user first says that a colleague works at a university and later clarifies that the colleague has recently left. A chatbot should incorporate the correction rather than continue treating the earlier information as current.

Stalnaker's common-ground framework helps explain this requirement. Conversational information is not static; participants can introduce, reject, and revise assumptions. The chatbot must therefore track the status of propositions instead of simply accumulating all previously mentioned statements.

In practical evaluation, researchers can test this ability by introducing explicit corrections, contradictory assumptions, and new contextual details. They can then examine whether the chatbot updates its interpretation appropriately.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

Distinguishing Linguistic Fluency From Pragmatic Competence

A central issue in AI research is whether fluent output reliably demonstrates successful interpretation. A chatbot may explain a pragmatic concept accurately when directly asked to define it but still misinterpret that concept in an unfamiliar conversation.

For example, a model may provide a correct definition of presupposition and identify stop as a presupposition trigger. Yet when asked, "When did the applicant stop submitting documents?" it might provide an invented date rather than identify the assumption that the applicant previously submitted documents and subsequently stopped.

This distinction highlights the difference between explaining a linguistic rule and applying it consistently in communication.

Bender et al. (2021) caution against interpreting the production of convincing language as sufficient evidence of human-like understanding. Contemporary pragmatic evaluation research similarly emphasizes task-specific testing and careful examination of error patterns (Ma et al., 2025).

The appropriate conclusion is not that chatbots lack every form of pragmatic ability. Rather, pragmatic performance must be examined empirically, with attention to the conditions under which a system succeeds, fails, or requires clarification.

A robust evaluation should consider whether the model can:

- Identify an inference supported by the context.
- Avoid presenting an uncertain inference as a fact.
- Recognize background assumptions embedded in a sentence.
- Distinguish a speaker's assumption from independently verified information.
- Maintain consistency across conversational turns.
- Revise its interpretation when new evidence becomes available.

These criteria provide a more informative basis for assessment than fluency alone.

Implications of the Study

Implications for Computational Linguistics

The analysis highlights the importance of treating pragmatics as a distinct dimension of natural language processing. Conventional language-processing systems often evaluate performance through grammatical accuracy, semantic similarity, or task completion. Although these measures are useful, they may not adequately capture the ability to interpret implied meanings and presupposed information.

Ma et al. (2025) emphasize the importance of comprehensive datasets and evaluation methods for assessing pragmatic abilities in large language models. Their survey supports the development of evaluation tasks that cover diverse pragmatic phenomena rather than relying on a small number of familiar examples.

A practical implication is that AI developers can incorporate contextual variations into model evaluation. For example, the same utterance can be presented with different preceding conversational turns to determine whether the model adjusts its interpretation appropriately. Researchers can also introduce contradictory information to assess whether a system recognizes and revises unsupported presuppositions.

Such evaluation would help identify the difference between a model that reproduces common linguistic associations and one that consistently responds to contextual evidence.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
LJLLR
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

Implications for Chatbot Design

The analysis suggests several directions for improving chatbot reliability.

First, systems should distinguish explicitly stated information from inferred meanings. If a user says, "Some students submitted their assignments," a chatbot should not automatically report that the remaining students failed to submit theirs. It should recognize that the inference may be plausible without treating it as logically established.

Second, conversational systems should maintain a coherent representation of previously established information. This would help them avoid accepting presuppositions that conflict with earlier statements.

Third, systems should be able to request clarification when the available context does not support a confident interpretation. For example, a chatbot encountering "That was brilliant" after a serious mistake should consider the possibility of sarcasm rather than automatically interpreting the statement as praise.

Finally, model evaluation should include culturally diverse and socially sensitive scenarios. Interpretations that are common in one linguistic community may not be appropriate in another. A reliable chatbot should account for such variation rather than treating a single cultural convention as universal.

These suggestions represent design implications derived from the theoretical analysis. Their effectiveness would need to be assessed through controlled empirical studies.

Implications for English Language Teaching

The study also has educational implications, particularly for teaching English pragmatics. Language learners may understand the grammatical structure of a sentence while misunderstanding its communicative function. Consequently, explicit instruction in implicature and presupposition can help learners develop more effective communication skills.

Teachers can use AI chatbots to develop exercises involving indirect requests, implied criticism, scalar implicatures, and presupposition triggers. For example, learners can compare the following statements:

1. "Some students completed the task."
2. "All students completed the task."
3. "The students completed the task again."

Students can identify the differences between explicit meaning, possible conversational inference, and presupposed information.

Chatbots can also generate alternative conversational contexts to demonstrate how interpretations change. However, teachers should verify the examples because a chatbot may offer a plausible but contextually inappropriate interpretation.

For textbook writers and language educators, the analysis supports a learning-outcome-based approach in which students demonstrate their ability to distinguish literal meaning from implied meaning, identify presupposition triggers, and justify interpretations using contextual evidence.

Implications for Human–AI Communication

Misinterpretation of pragmatic meaning can affect user trust. A chatbot that repeatedly misunderstands indirect requests, sarcasm, or background assumptions may appear insensitive or unreliable, even when its grammar is accurate.

In educational settings, such errors may lead learners to accept incorrect interpretations. In professional settings, they may cause misunderstandings about instructions, responsibilities, or decisions. In customer service, failure to recognize indirect dissatisfaction may result in responses that do not address the user's actual concern.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
LJLLR
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

These risks suggest that AI systems should be evaluated not only on whether their responses sound natural but also on whether their interpretations are appropriately supported by the conversation. When uncertainty is substantial, acknowledging ambiguity may be preferable to producing an overconfident answer.

Recommendations for Future Research

The present analysis identifies several directions for future empirical investigation.

First, researchers should develop controlled pragmatic datasets. These datasets could contain examples of scalar implicatures, indirect requests, presupposition triggers, irony, and context-dependent meanings. Each item should include sufficient context to establish the intended interpretation while allowing for legitimate alternatives where appropriate.

Second, researchers should compare multiple AI chatbots. Different systems may perform differently depending on their architecture, training, instruction-following abilities, and conversational context. Comparative testing would help determine whether observed patterns are general or specific to particular models.

Third, human judgments should be incorporated into evaluation. Multiple human annotators could assess chatbot interpretations and identify acceptable alternatives. Inter-annotator agreement would help establish whether an example has a sufficiently clear intended interpretation for reliable scoring.

Fourth, evaluation should include context manipulation. Researchers could present the same utterance in different contexts to determine whether a chatbot changes its interpretation when the evidence changes. This would be especially valuable for sarcasm, indirect requests, and cancellable implicatures.

Fifth, presupposition tracking should be investigated across longer conversations. Researchers could test whether chatbots preserve previously established information, recognize corrections, and avoid treating unsupported assumptions as facts.

Sixth, cross-cultural and multilingual studies should be conducted. Pragmatic conventions differ across languages and communities. Evaluations involving English and other languages could clarify whether a model's apparent competence depends on culturally familiar patterns.

Finally, future studies should report transparent methods, reproducible prompts, model versions, scoring procedures, and limitations. Such reporting would make findings easier to compare and reduce the risk of generalizing from isolated examples.

Limitations of the Study

Several limitations must be acknowledged.

First, the study uses a qualitative, analysis-based design rather than a controlled experiment. The illustrative examples demonstrate theoretical distinctions but do not establish the actual success rate of any particular AI chatbot.

Second, the analysis focuses primarily on English-language communication. Its conclusions may not apply equally to languages with different grammatical structures, politeness conventions, or discourse practices.

Third, the three theoretical frameworks provide a strong foundation for examining implicature and presupposition, but they do not encompass every aspect of pragmatic competence. Speech acts, deixis, politeness, turn-taking, and multimodal communication may also influence interpretation.

Fourth, pragmatic interpretations can vary among speakers. An utterance that one listener interprets as sarcastic may be understood by another as sincere or playful. Evaluation criteria must therefore distinguish clear cases from genuinely ambiguous ones.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
LJLLR
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

Fifth, AI systems are continually updated. Findings obtained from one model version or prompting strategy may not generalize to other systems or later versions.

These limitations do not undermine the theoretical value of the study. Instead, they establish the need for controlled empirical research to determine how the proposed analytical framework corresponds to actual chatbot behavior.

Conclusion

This study aimed to examine the pragmatic competence in understanding conversational implicature and presupposition in AI chatbots, which is informed by three influential theoretical frameworks: conversational implicature by Grice (1975), a comprehensive theory of pragmatics by Levinson (1983), and a theory of presupposition and common ground by Stalnaker (1974).

The analysis shows that interpreting pragmatically is not just picking out words and grammatical structures. It involves distinguishing literal content from implied meaning, identifying background assumptions, taking contextual factors into account and revising interpretations when new evidence emerges.

Grice's theory is used to describe how conversational expectations facilitate understanding of indirect meanings. The relationship between linguistic expressions and pragmatic inference is illuminated by the account offered by Levinson and the role of shared assumptions in communication is emphasized by the account offered by Stalnaker. These theories offer a helpful lens to consider AI-generated interpretations.

Modern studies corroborate the importance of a more comprehensive appraisal of pragmatic competence. Research on implicature, pragmatic understanding benchmarks, and conversational awareness shows that there is some variation between tasks, and ongoing issues in context-sensitive interpretation (Andersson & McIntyre, 2025; Cong, 2024; Sravanthi et al., 2024). These results show that a fluent answer does not necessarily indicate full pragmatic competence.

This study's major contribution is a theory-informed four-dimensional framework to study the interpretation of chatrobots regarding the recognition of implied meaning, contextual sensitivity, identifying presupposition, and interpretive flexibility. Its illustrative analysis illuminates what is unexplored and needs to be explored further without making any claims of experimental findings that are not measured.

In the end, there is a need for evaluation methods that can evaluate the extent to which a chatbot's interpretations are accurate and well-supported by linguistic and contextual evidence. Consistency and responsibility in the understanding of implied meanings and presupposed information from AI chatbots can be evaluated through controlled testing, different datasets, human evaluation and open reporting.

REFERENCES

Andersson, M., & McIntyre, D. (2025). Can ChatGPT recognize impoliteness? An exploratory study of the pragmatic awareness of a large language model. Journal of Pragmatics, 239, 16–36. https://doi.org/10.1016/j.pragma.2025.02.001

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Cho, Y., & Kim, S. (2024). Pragmatic inference of scalar implicature by LLMs. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
LJLLR
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 | ISSN (Print): 3006-3167
Volume 5, Issue 8 October, 2026

Workshop) (pp. 10–20). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-srw.2

Cong, Y. (2024). Manner implicatures in large language models. Scientific Reports, 14, Article 29113. https://doi.org/10.1038/s41598-024-80571-3

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. Science, 336(6084), 998. https://doi.org/10.1126/science.1218633

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and semantics: Vol. 3. Speech acts (pp. 41–58). Academic Press.

Hu, J., Floyd, S., Jouravlev, O., Fedorenko, E., & Gibson, E. (2023). A fine-grained comparison of pragmatic language understanding in humans and language models. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 4194–4213). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.230

Levinson, S. C. (1983). Pragmatics. Cambridge University Press.

Ma, B., Li, Y., Zhou, W., Gong, Z., Liu, Y. J., Jasinskaja, K., Friedrich, A., Hirschberg, J., Kreuter, F., & Plank, B. (2025). Pragmatics in the era of large language models: A survey on datasets, evaluation, opportunities and challenges. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 8679–8696). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.acl-long.425

Sravanthi, S., Doshi, M., Tankala, P., Murthy, R., Dabre, R., & Bhattacharyya, P. (2024). PUB: A pragmatics understanding benchmark for assessing LLMs' pragmatics capabilities. In Findings of the Association for Computational Linguistics: ACL 2024. Association for Computational Linguistics. https://aclanthology.org/2024.findings-acl.719/

Stalnaker, R. C. (1974). Pragmatic presuppositions. In M. K. Munitz & P. K. Unger (Eds.), Semantics and philosophy (pp. 197–213). New York University Press.

Austin, J. L. (1962). How to do things with words. Oxford University Press.

Beaver, D. I., & Geurts, B. (2014). Presupposition. In E. N. Zalta (Ed.), The Stanford encyclopedia of philosophy. https://plato.stanford.edu/entries/presupposition/

Carston, R. (2002). Thoughts and utterances: The pragmatics of explicit communication. Blackwell Publishing.

Huang, Y. (2014). Pragmatics (2nd ed.). Oxford University Press.

Searle, J. R. (1969). Speech acts: An essay in the philosophy of language. Cambridge University Press.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. Trends in Cognitive Sciences, 20(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Kabbara, J., & Cheung, J. C. K. (2022). The pragmatics of presupposition in language models.

Potts, C. (2015). Presupposition and implicature. In S. Lappin and C. Fox (Eds.), The handbook of contemporary semantic theory (2nd ed.). Wiley-Blackwell.

Clark, H. H. (1996). Using language. Cambridge University Press.

Jurafsky, D., & Martin, J. H. (2025). Speech and language processing (3rd ed. draft). Online manuscript. https://web.stanford.edu/~jurafsky/slp3/