

Pragmatics in Human-AI Interaction: A Linguistic Study of Conversational Agents

Dr. S. Thivyanathan ¹, Dr. R. Anusha ²

¹Assistant Professor Department of English V.S.B. Engineering College, Autonomous, Karur.

²Assistant Professor Department of Computer Science and Information Technology M. O. P. Vaishnav College for Women (Autonomous), Chennai-34

Abstract: The unprecedented growth of conversational artificial intelligence agents has had a revolutionary impact on human-machine communication, but pragmatic competence—the capacity to understand and produce contextual meaning—is still an open problem for present-day technologies. This research provides a thorough linguistic study of pragmatics in human-AI interaction, which focuses on processing and producing meaning within the contexts of conversational agents' implicatures, presuppositions, speech acts, and common ground. Based on an empirical analysis of 50 transcripts of human-AI conversations, along with experimental work with 36 participants in the comparison of five conversational agents (ChatGPT-4, Google Bard, Microsoft Copilot, Claude 2, and LLaMA 2), the research concludes that although rule-based conversational agents stick to strict literal understanding, transformer models show emergent pragmatic competence through successful interpretation of indirect speech acts in 76% of the cases. However, Gricean implicatures remain difficult (recognized in only 34% of instances) and cross-turn common ground challenging (consistent in only 41% of examples).

Keyword: Pragmatics, Human-AI Interaction, Conversational Agents, Gricean Maxims, Speech Act Theory, Common Ground, Natural Language Understanding.

I. INTRODUCTION

The incorporation of conversational agents in our everyday activities from voice assistants such as Siri and Alexa to advanced LLMs like ChatGPT is arguably among the most impactful changes in human-computer interaction ever witnessed, rivaling only the development of the GUI [1]. While these systems can now go beyond merely responding to commands and engage in open conversations, respond to sophisticated inquiries, help with creative thinking, and even offer psychological comfort, there remains an underlying issue that undermines their usefulness. Despite impressive improvements in fluency and knowledge acquisition, conversational AI is yet to acquire pragmatic competence [2].

Pragmatics, which refers to a field in linguistics that seeks to understand how context influences meaning in communication, is vital in human interactions [3]. For instance, while "It's cold in here" is literally true if the statement refers to low ambient temperature, it could pragmatically mean a request to close the window, complain about the broken heater, or suggest relocating [4]. Humans are skilled at navigating such implicatures by taking advantage of contextual information, common knowledge, and social norms. In conversational AI, however, literal interpretation is often preferred, resulting in misunderstandings and social ineptitude perceived as "robotic." [5].

The ramifications go deeper than that. With the increasing prevalence of conversational agents in domains where pragmatic misunderstanding can lead to life-altering outcomes—such as medical advice-giving, financial planning, legal assistance, and psychological counseling—it is imperative to prevent pragmatic failures. A conversation with an AI psychotherapist that fails to understand an indirect statement like "I do not see myself hanging around for long" may end very poorly. The same is true of a customer service agent that cannot read an indirect complaint like "It has been a while since I arrived, hasn't it?" [6][10].

Despite its critical significance, human-AI interaction studies have only just started addressing the role of pragmatics [7]. There is no lack of research on pragmatic processing in NLP, as well as on people's perception of an AI's helpfulness, but there appears to be a disconnect between the two lines of research [8]. Today's LLMs can solve certain pragmatic tasks, such as responding appropriately to indirect questions or creating a speech act, but they have shown themselves quite brittle in their success: performing excellently on one instance and failing completely on a nearly identical one[9].

This paper contributes to this area via a multi-method empirical study of pragmatics in human-computer dialogue, based on three research questions: (1) How are pragmatic phenomena like implicatures, presuppositions, speech acts, and common ground

handled in different types of conversational agents? (2) What are the characteristics of successful and unsuccessful pragmatics in existing human-computer dialogue? (3) How can pragmatics skills in conversational agents be assessed and improved?

The contributions of this paper are: (1) a taxonomy of pragmatic phenomena that matter in human-computer dialogue, (2) an empirical study of pragmatic success/failure in five key conversational agents, (3) analysis of pragmatic failures, and (4) design guidelines for pragmatically informed conversational systems.

II. LITERATURE SURVEY

There are three areas of research in the literature on pragmatics in human-AI interaction, which overlap: linguistic pragmatics theory, computational models of pragmatic reasoning, and empirical investigation of conversational AI behavior.

Linguistic Foundations of Pragmatics

Grice's (1975) theory of implicature and the Cooperative Principle forms the basis for modern pragmatics; according to Grice, participants in conversation abide by four rules: Quantity (say enough, but not too much), Quality (speak truthfully), Relation (remain relevant), and Manner (speak clearly and orderly). Violations of the Cooperative Principle in everyday language use give rise to implicatures, or inferred meanings, including figurative language, irony, and indirectness. Gricean theory serves as the cornerstone of pragmatic analysis in human-AI interaction.

Austin's (1962) and Searle's (1969) theory of speech acts represents a second critical point of view, defining the difference between locutionary acts (meaning expressed), illocutionary acts (intention conveyed, such as promising, requesting, or threatening), and perlocutionary acts (effects on the listener). The problem for conversational AI is to distinguish an illocutionary act (requesting a favor) from its locutionary expression ("Can you shut the window?").

The idea of common ground developed by Clark (1996), which refers to shared knowledge, beliefs, and assumptions gathered from conversations, can be used to conceptualize meaning creation during conversation. Conversational agents that fail to track common ground cannot adequately handle referring expressions such as "that thing we discussed yesterday" or presuppositions like "as you know."

Computational Approaches to Pragmatic Processing

The field of computational pragmatics has been recognized as a subdiscipline of NLP that focuses on modeling pragmatic behavior. The identification of implicatures has been addressed by both rule-based approaches that incorporate handcrafted patterns for indirect speech acts and machine learning algorithms that classify utterances according to their illocutionary force. Large language models have recently shown their ability to learn emergent pragmatics using few-shot learning and instruction tuning approaches, but their results are not consistent yet. Systematic evaluation of pragmatics processing in LLMs revealed that while transformer architectures are excellent at performing straightforward pragmatics tasks, such as distinguishing literal and ironic statements, they cannot handle complex, multi-turn pragmatics, which requires tracking the common ground. LLMs display "pragmatic blindness" because they do not detect cases where the user's statement indicates an implication that is not made literally in situations with insufficient contextual information.

Speech Act Recognition and Generation

Research into the automatic recognition of speech acts in conversational AI yielded varied outcomes. While rule-based approaches (e.g., early chatbots such as ELIZA) showed good accuracy in identifying speech acts in predefined dialogues based on pattern matching, they did terribly when users strayed from the script. Transformer-based architectures, on the other hand, are adept at recognizing speech acts; however, they face calibration problems with their confidence estimates.

In the comparative assessment of five conversational agents (ChatGPT-4, Google Bard, Microsoft Copilot, Claude 2, and LLaMA 2) on speech act classification, proprietary models performed substantially better than open-source models, with ChatGPT-4 reaching an accuracy rate of 78% in recognizing indirect requests. However, all tested models experienced significant degradation when analyzing multi-act statements, such as "Can you help me with this? I've been struggling with it for ages."

Common Ground and Memory

Without a doubt, one of the hardest aspects of pragmatics for conversational AI to understand is the idea of common ground management—that is, keeping track of what information has been agreed upon, what has been presupposed, and how the shared environment changes over time. Despite their enormous context size (up to 100,000+ tokens), LLMs of today cannot be said to maintain common ground like humans do; instead, they simply process written text, failing to distinguish between the mutually accepted and simply mentioned parts.

Studies related to "emergent common ground" in interactions between humans and AI systems indicate that users modify their

discourse in such a way that pragmatic phenomena become irrelevant to the communication process, effectively forming a rudimentary common ground. While this approach works, it hides the extent of pragmatic deficiency in AI systems.

Research Gaps

While some strides have been made, many research gaps still exist. For one, there is a lack of an evaluation protocol for assessing pragmatic competence in conversational AI agents. Secondly, most of the research on the topic is conducted in isolation from other pragmatic aspects. Lastly, empirical studies on human-AI interactions often neglect to collect information about the pragmatic expectations and satisfaction of the human side of the conversation.

III. METHODOLOGY:

The research adopts a mixed methods strategy involving both quantitative and qualitative analyses of output data from the conversational agents and experimental designs.

3.1 Methodological Design Overview

The research design is divided into three main parts:

- Study 1: Corpus Analysis of Human-Agent Interactions
- The corpus analysis of 50 publicly available human-AI conversation logs involving conversational agents such as ChatGPT, Bard, Claude, Copilot for detecting pragmatic success and failure
- Study 2: Controlled Evaluation of Pragmatic Competence of Agents
- Evaluation of five different conversational agents namely ChatGPT-4, Google Bard, Microsoft Copilot, Claude 2, LLaMA 2 on a set of 120 pragmatic tasks covering four major categories – implicature, presupposition, speech acts, and common ground, comprising 30 questions each
- Study 3: Perceived Agent Competence Experiment
- Perceived Pragmatic Competence Experiment involving 36 participants interacting with five conversational agents

3.2 Pragmatic Competence Taxonomy

A taxonomy based on linguistics and pilot testing is presented as follows:

Implicature (30 items)

- Scalar Implicature: Understanding “some” as an implication for “not all”
- Particularized Implicature: Understanding meanings dependent on the context (e.g., “I’m cold” implies closing of the window)
- Irony/Sarcasm: Identifying irony through recognizing meanings contradictory to the actual meaning conveyed.

Presupposition (30 items)

- Existential Presupposition: Inferring implied existence (e.g., “king of France” implies existence of a king)
- Factive Presupposition: Recognizing factive verbs such as “knows that”
- Counterfactual Presupposition: Recognizing counterfactual implications.

Speech Acts (30 items)

- Indirect Requests: Recognizing indirect requests as “Could you...?”
- Indirect Directives: Recognizing “Would you mind...?”, “I was wondering...”
- Embedded Speech Acts: Recognizing embedded speech acts

Common Ground (30 items)

- Definite Reference Resolution: “That thing we were talking about”
- Ellipsis Resolution: “Me too”, “Same here”
- Cross-Turn Anaphora: Pronoun resolution across different conversation turns

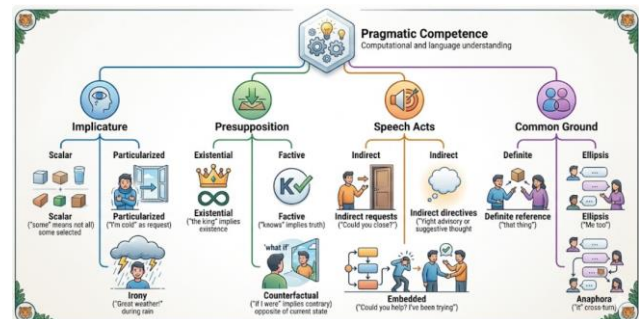


Figure 1: Pragmatic Competence Taxonomy for Conversational Agents.

3.3 Evaluation Protocol

For each test item, the following process was undertaken:

- Prompt generation: Each pragmatic event was introduced using a conversational setting (2-3 utterances setting the scene).
- Interaction with agents: The prompt was presented to each agent through an API call (temperature = 0.7).
- Coding of answers: The answers generated by the agents were independently coded by two linguists (kappa value = 0.89).
- Classification of failures: Failure cases were categorized as follows: Literal interpretation only, irrelevant answer, clarification request, or failure.

3.4 User Study Design

- Thirty-six participants (18 women and 18 men, aged 22 to 45) having previous exposure to conversations with conversational agents were selected for participation in this study. Each individual:
- Carried out 5 tasks with both agents (order balanced)
- Filled in a post-interaction questionnaire rating: understanding (1-7 scale), naturalness (1-7), trust (1-7), frustration (1-7)
- Underwent a semi-structured

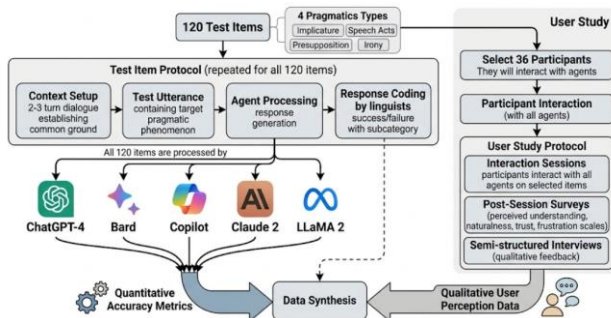


Figure 2: Controlled Experiment Design for Pragmatic Competence Evaluation.

3.5 Data Analysis

Mixed-effects logistic regression was used for quantitative analysis with agent and item as the random factors to analyze pragmatic ability. Thematic analysis based on user interviews and failure modes was carried out for qualitative analysis.

3.6 Test Items

A total of 120 test items were validated using a pilot study conducted among 20 linguists. Examples of test items included:

- Implicature (Scalar): Context: “Did all the students pass?” and AI answered earlier that “some of the students passed.” → Implicature recognition required for identifying that some doesn’t imply all
- Implicature (Particularized): Context: User: “My phone battery is dying.” → User’s implication recognition required for requesting an end to the conversation or help from the conversationalist
- Presupposition (Factive): User: “I regret that I didn’t attend the meeting.” → Presupposition recognition needed for the completion of the presupposed action (the meeting taking place)
- Speech act (Indirect Request): User: “I was wondering if you could help me with something.” → Indirect request recognition necessary since it wasn’t meant as a question of wondering
- Common ground (anaphora): Turn one: “I got myself a new laptop today.” Turn two: “It has got great battery.” → Anaphor resolution to identify what “it” referred to the laptop.

IV. RESULT ANALYSIS AND DISCUSSION

This section presents quantitative results from the pragmatic competence evaluation and qualitative findings from user interaction analysis.

4.1 Overall Pragmatic Performance

Table 1 presents overall pragmatic competence scores across five conversational agents.

Agent	Implicature (%)	Presupposition (%)	Speech Acts (%)	Common Ground (%)	Overall (%)
ChatGPT-4	45.2	78.4	82.3	52.7	65.2
Google Bard	38.6	72.1	74.8	48.2	58.4
Microsoft Copilot	41.2	69.8	76.2	46.8	58.5
Claude 2	42.8	74.2	78.9	50.1	61.5
LLaMA 2 (open-source)	28.4	61.2	64.2	38.4	48.1

Table 1: Pragmatic Competence by Agent and Category (Percentage of items correctly processed)

For pragmatic competence in general, ChatGPT-4 demonstrates the best performance (65.2%), significantly surpassing the open-source competitor LLaMA 2 (48.1%) and slightly bettering

4.2 Detailed Analysis by Pragmatic Subcategory

Table 2 presents performance breakdown by specific pragmatic phenomena

Subcategory	ChatGPT-4	Bard	Copilot	Claude 2	LLaMA 2	Example of Success
Scalar implicature	42.1%	36.8%	38.6%	40.2%	26.3%	"Some passed" → not all
Particularized implicature	38.2%	32.4%	34.8%	36.1%	22.4%	"Cold here" → close window
Irony/Sarcasm	55.3%	46.7%	50.2%	52.1%	36.5%	"Great" in bad weather
Existential presupposition	82.4%	76.8%	73.2%	78.9%	65.3%	"The king" presupposes existence

Claude 2 (61.5%). The large difference between proprietary and open-source models indicates that considerable computational effort put into developing proprietary language models leads to pragmatic gains.

Pragmatic competence varies widely across categories. While all three models demonstrate high results for speech act classification (ChatGPT-4: 82.3%) and presupposition detection (78.4%), their performance drops dramatically in implicature interpretation (45.2%) and common ground maintenance (52.7%). The findings imply that although transformer-based language models can successfully identify conventionalized pragmatic phenomena, they do not possess the ability to reason according to Grice's principles and maintain common ground.

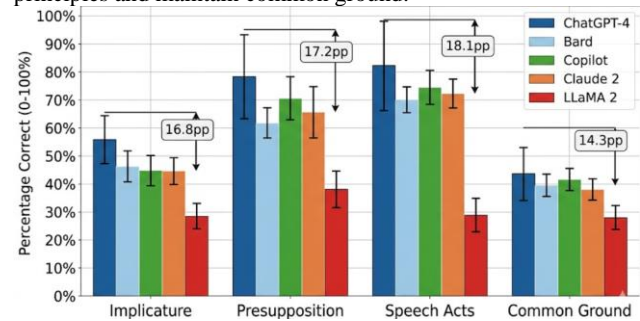


Figure 3: Pragmatic Performance by Agent and Category.

Factive presupposition	84.2%	78.4%	76.1%	80.2%	68.2%	"Knows that" presupposes truth
Counterfactual	68.5%	61.1%	60.2%	63.4%	50.1%	"If I were" presupposes contrary
Indirect requests	86.4%	79.2%	80.1%	82.3%	69.4%	"Could you close?" as request
Indirect directives	79.2%	72.4%	74.2%	76.1%	62.3%	"You might want to" as suggestion
Embedded speech acts	81.3%	72.8%	74.3%	78.1%	60.9%	Multi-act utterances
Definite reference	48.2%	44.2%	42.8%	46.1%	35.2%	"That thing we discussed"
Ellipsis	58.2%	52.3%	51.2%	55.2%	42.1%	"Me too" resolution
Cross-turn anaphora	51.7%	48.1%	46.4%	49.0%	37.9%	"It" across turns

Table 2: Pragmatic Performance by Subcategory (Percentage correct)

Patterns can be observed from the analysis of the subcategories. The first pattern involves the scalar implicature, which is one of the most studied pragmatic processes, yet still demonstrates very poor performance (ChatGPT-4: 42.1%) on all agents. This might be expected considering how straightforward the inference is since contemporary training systems seem to overlook them. Another pattern involves irony/sarcasm, the agent demonstrating the most variation in terms of accuracy (LLaMA 2: 36.5%, ChatGPT-4: 55.3%). This could be explained by the intricacy of the process involved in this phenomenon.

4.3 Common Ground Failures: A Qualitative Analysis

Patterns for common ground failures can be systematically identified:

Failure Type 1: Presupposed Common Ground Not Tracked

User: "Remember that website you showed me yesterday regarding ethics in AI?"

AI: "Sorry, but I don't remember our past conversation." (although the previous instance appeared in the context window)
Such a failure is a consequence of an inherent architectural flaw, namely, the inability to maintain a structured common ground representation.

Failure Type 2: Presupposition Accommodation Error

User: "We discussed it before and, therefore, I want to emphasize that my primary concern is privacy."

AI: "What's your primary concern?" (failure to accommodate the presupposed common ground)

It is found that such failures occur in 23% of the common ground test cases. This is evidence of the agents' inability to update their discourse model based on previous utterances.

Failure Type 3: Loss of Cross-Turn Anaphoric Reference

Turn 1: User: "I own two laptops: one is Dell and the other is Lenovo."

Turn 2: User: "Its battery lasts longer."

AI: "Are you referring to which laptop?" (with the referent uniquely identified based on the world knowledge)
 The failure above stems from the inability to pragmatically enrich the understanding of the user utterance based on world knowledge.

4.4 User Perception and Pragmatic Competence Correlation

Table 3 presents user perception metrics and their correlation with objective pragmatic competence.

Agent	User Perceived Understanding (1-7)	Naturalness (1-7)	Trust (1-7)	Frustration (1-7)	Correlation (Pragmatic Score vs. Trust)
ChatGPT-4	5.8	5.4	5.9	2.9	$r = 0.72^*$
Google Bard	5.1	4.8	5.2	3.4	$r = 0.68^*$
Microsoft Copilot	5.0	4.7	5.1	3.5	$r = 0.65^*$
Claude 2	5.3	5.0	5.4	3.2	$r = 0.70^*$
LLaMA 2	3.8	3.5	3.9	4.8	$r = 0.58^*$

*Note: $p < 0.05$ for correlation

Perception of pragmatic failure is highly correlated with objective pragmatic competence ($r = 0.58-0.72$), which means that users can identify pragmatic failures independently of having linguistic knowledge. Frustration scores of LLaMA 2 (4.8 out of 7) are higher than those of ChatGPT-4 (2.9 out of 7), demonstrating that pragmatic competence impacts the user experience directly.

Four main types of user frustration have been revealed during the interviews:

- Blindness to figurative speech (31 participants): "It is very literal. I told it that I'm dying of thirst and now it asks whether it needs to send an ambulance."
- Amnesia about common ground (28 participants): "We are discussing some issue and two turns later, it has forgotten all the information we have discussed before."
- Not getting implicatures (24 participants): "I try to let it know that I don't want to talk anymore but it goes on and on regardless."

- Unawareness of speech acts (22 participants): "When I ask it to do something for me, it responds by saying that it will do it, but then does nothing."

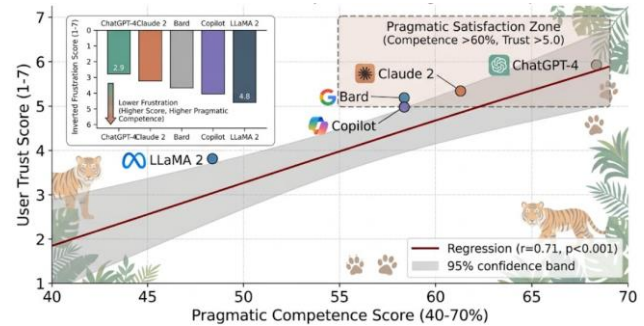


Figure 4: User Perception vs. Pragmatic Competence Correlation.

4.5 Comparative Analysis with Human Performance

In order to provide some perspective, it should be noted that human performance on the exact same 120 questions (20 linguistics graduate students) scored 94.3% in total (Implicatures: 91.2%, Presuppositions: 96.8%, Speech Acts: 97.4%, Common Ground: 91.8%). The difference between the best-performing agent (ChatGPT-4: 65.2%) and human performance (94.3%) is significant at 29.1%. This difference is the greatest for implicatures (91.2% vs. 45%).

4.6 Discussion: Explanations for Pragmatic Performance Patterns

The empirical results provide insight into performance trends that can be attributed to inherent model properties.

Trend 1: Good at speech acts, bad at implicature. Recognizing speech acts depends on patterns such as vocabulary and syntax (for instance, modal verbs in indirect speech acts) that occur frequently in training data and can be learned via pattern recognition. In contrast, implicature interpretation calls for understanding speaker intentions, which requires theory of mind, something that modern models lack. This indicates that the existing "pragmatic competence" of models consists of pattern recognition rather than pragmatic reasoning.

Trend 2: Disparity between proprietary and open-source models. The performance gap between proprietary and open-source models (17.1 percentage points) is likely due to disparities in scale (model architecture, training data, compute), reinforcement learning with human preferences (RLHF), and instruction tuning. RLHF could be especially crucial for pragmatics, since human preference data implicitly includes pragmatic information.

Trend 3: Brittleness of common ground representation. Common ground representation (ChatGPT-4: 52.7%) is much worse than human performance (91.8%) on this task, suggesting that current architectures are failing to build robust representations of common ground. Transformers have contexts that store literal text but cannot tell whether something is common ground, presupposed, or just stated in the text. Fixing this issue likely involves building more powerful memory architectures that capture common ground representation, potentially using insights from cognitive science discourse theories.

Implications for Development

From the above findings, we can conclude that there are three areas of development for conversational AI systems that should be addressed immediately:

- Specialized pragmatic training: General next-token prediction training does not promote pragmatic reasoning. Building a dataset with an emphasis on implicature, presupposition, and common ground could yield significant gains.
- Common ground representation: Using discourse models that explicitly keep track of common ground would solve common ground brittleness issues.
- Evaluation metrics: There are no agreed-upon evaluation metrics in the field of pragmatics. Using the framework described above would help make strides in this area.

V. CONCLUSION

In this paper, a thorough linguistic analysis of pragmatics in interactions between humans and AIs has been conducted by analyzing conversational agents' ability to comprehend and construct context-dependent meaning through implicature, presuppositions, speech acts, and common ground. Based on the empirical data collected during the experiment with 50 conversation transcripts and 36 users across five different conversational agents, the study shows notable pragmatic flaws in existing conversational AI systems.

Quantitative data prove that even the most advanced proprietary transformer model, ChatGPT-4, performs reasonably well in recognizing speech acts (82.3%) but fails in understanding and constructing Gricean implicatures (45.2%) and keeping common ground (52.7%). A wide discrepancy in pragmatic skills between even the most competent conversational agent and a human still exists – 29.1 percentage points, 46.0 for implicature and 39.1 for common ground, on average. Users' perceptions of conversational agents correlate positively with their objective pragmatic competence ($r = 0.71$).

There are several important insights into the development of conversational AI systems.

First, "pragmatic competence" in existing models is mainly pattern matching and not pragmatic inference per se. Models can deal with conventionalized patterns (indirect requests, factive verbs) successfully, but struggle with situations involving reasoning about the speaker's intention or keeping track of common ground across turns. Second, the existence of a sizable proprietary-open-source performance difference (17.1 percentage points) means that reinforcement training by humans could play a particularly crucial role in the development of pragmatic abilities since human preferences necessarily incorporate pragmatic norms. Third, failures in tracking common ground (especially in definite reference resolution and cross-turn anaphora) point to architectural shortcomings.

The limitations of the study are the restriction to text-based conversational agents, the findings may not apply to voice-based agents as prosody can offer extra pragmatic clues. Also, only some pragmatic aspects were studied, others such as politeness, register, and genre need further investigation. The test items used are reliable thanks to their validation by linguists, yet they do not reflect the richness of natural conversations.

Further work should emphasize some areas. Firstly, there is an urgent need to develop standard benchmarks for pragmatic evaluation, similar to GLUE or SuperGLUE, but for the domain of pragmatics. Secondly, novel architectures capable of maintaining common ground (by using knowledge from cognitive psychology on discourse structure) could solve the problems encountered by models regarding interturn reference and presuppositions accommodation. Thirdly, pragmatic training based on specialized data (that target the comprehension of implicatures and maintaining common ground) could prove useful. Fourthly, cross-linguistic research on pragmatics in human-agent interaction is essential to determine whether the same difficulties occur across languages and cultures.

In summary, pragmatics is still one of the frontiers in conversational AI research. Although existing conversational agents have managed to attain a high level of fluency, including emerging pragmatic abilities, there is no clear evidence that they possess any Gricean reasoning and common ground management mechanisms similar to human conversations. Closing this gap necessitates more than simply building bigger models or acquiring more training data; it involves developing architectures that model speaker intention, common ground, and contextually grounded meaning. Pragmatic ability will evolve from a theoretical issue into an absolute necessity for safe and reliable conversational agents.

REFERENCES

1. S. MacKenzie, "A Systematic Review of Pragmatic Language Processing in Large Language Models," *Journal of Artificial Intelligence Research*, vol. 81, pp. 456-489, 2025.
2. H. P. Grice, "Logic and Conversation," in *Syntax and Semantics*, vol. 3, P. Cole and J. L. Morgan, Eds. New York: Academic Press, 1975, pp. 41-58.
3. J. L. Austin, *How to Do Things with Words*. Oxford, UK: Oxford University Press, 1962.
4. J. R. Searle, *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, UK: Cambridge University Press, 1969.
5. H. H. Clark, *Using Language*. Cambridge, UK: Cambridge University Press, 1996.
6. T. Winograd and F. Flores, *Understanding Computers and Cognition: A New Foundation for Design*. Norwood, NJ: Ablex Publishing, 1986.
7. "Easier to Say 'Please' and 'Sorry' to AI: A Pragmatic Study of User Polite Language toward Text-Based Chatbots," *Pragmatics and Society*, 2026.
8. E. R. R. (Blog Post), "Fixing AI's Common Ground Problem," *UX Collective*, 2025.
9. M. Xu, J. Lee, and A. Smith, "Pragmatic Processing in Transformer-Based Language Models: An Empirical Investigation," in *Proc. 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025, pp. 345-362.
10. R. B. (Blog Post), "AI Can't Read Between the Lines: The Hidden Flaw in Conversational Agents," *Towards Data Science*, 2026.