African Journal of Advanced Studies in Humanities and Social Sciences (AJASHSS)
المجلة الإفريقية للدراسات المتقدمة في العلوم الإنسانية والاجتماعية
Online-ISSN: 2957-5907
Volume 5, Issue 2, 2026
Page No: 1-9
Website: https://aaasjournals.com/index.php/ajashss/index
Arab Impact factor 2025: 0.83 **SJIFactor 2024: 6.856** **ISI 2025: 1.191**

Reducing Misunderstanding in Human-AI Communication: A Speech Act Model for Better Language Interaction

Zeena Al-Asi *
Department of English, Faculty of Education, Zuwara, University of Zawia, Libya

تقليل سوء الفهم في التواصل بين الإنسان والذكاء الاصطناعي: نموذج 'أفعال الكلام' لتحسين التفاعل اللغوي

زينة العاصي *
قسم اللغة الإنجليزية، كلية التربية - زوارة، جامعة الزاوية، ليبيا

*Corresponding author: zynhalasy7@gmail.com

Received: January 04, 2026	Accepted: March 20, 2026	Published: April 05, 2026

Abstract
Human-AI talk often fails when a system reads sentence form but misses user action. A user may ask, hint, soften, refuse, or seek repair. Many systems answer the words on the screen, not the act behind them. This paper studies that gap and offers a speech act model for better language interaction. The study uses secondary analysis of public resources and published experiments. The paper reviews classic speech act theory and recent benchmarks of pragmatic ability. The public materials include DailyDialog, MultiWOZ, the Stanford Politeness Corpus, the Switchboard Dialogue Act Corpus, PUB, DialogBench, and INDIR-IT. The review shows four repeated failure points. Systems struggle with indirect requests, politeness, dialogue act balance, and repair after confusion. Based on these findings, the paper proposes a four-stage model: act detection, context reading, face-risk check, and repair-based response planning. The model treats misunderstanding as a mismatch between user intention and system reply. It uses clarification when force is unclear. It also adjusts tone to the social setting. It links classic pragmatics with open benchmark practice. The paper argues that better human-AI talk needs pragmatic design, not only fluent text. The model offers a clear path for future testing, system tuning, and linguistics-based evaluation of intelligent language systems.

Keywords: human-AI communication; speech acts; pragmatics; dialogue acts; politeness; indirect speech; large language models; repair strategies.

الملخص
غالبًا ما يخفق الحوار بين الإنسان والذكاء الاصطناعي عندما يحلل النظام صيغة الجملة دون إدراك الغرض من فعل المستخدم؛ فقد يسأل المستخدم، أو يلمح، أو يتلطف في الطلب، أو يرفض، أو يسعى لتصحيح فهم معين، إلا أن الكثير من الأنظمة تكتفي بالرد على الكلمات الظاهرة على الشاشة عوضًا عن الفعل الكامن خلفها. تتقصى هذه الورقة البحثية تلك الفجوة وتطرح نموذجًا لأفعال الكلام لتعزيز التفاعل اللغوي. تعتمد الدراسة على تحليل ثانوي لمصادر عامة وتجارب منشورة، وتراجع نظرية أفعال الكلام الكلاسيكية ومعايير القدرة التداولية (البرجماتية) الحديثة. شملت المواد العامة قواعد بيانات متنوعة منها: (DailyDialog، MultiWOZ، Stanford Politeness Corpus، Switchboard Dialogue Act Corpus، PUB، DialogBench، وINDIR-IT).
وقد كشفت المراجعة عن أربع نقاط فشل متكررة، حيث تعاني الأنظمة في التعامل مع: الطلبات غير المباشرة، وقواعد اللباقة، وتوازن أفعال الحوار، وعملية "الإصلاح" أو التصحيح بعد حدوث خلط في الفهم. وبناءً على هذه النتائج، تقترح الورقة نموذجًا من أربع مراحل: كشف الفعل، وقراءة السياق، وفحص مخاطر "الوجه" (حفظ المقامات)، وتخطيط الاستجابة القائم على التصحيح. يعالج النموذج سوء الفهم باعتباره تعارضًا بين نية المستخدم ورد النظام، ويستخدم أسلوب الاستيضاح عندما يكون القصد غير واضح، كما يضبط نبرة الحديث لتناسب السياق الاجتماعي. تربط الورقة بين اللسانيات التداولية

الكلاسيكية وممارسات التقييم الحديثة، وتخلص إلى أن تحسين الحوار مع الذكاء الاصطناعي يتطلب تصميمًا تداوليًا لا يقتصر على طلاقة النص فحسب. يقدم النموذج مسارًا واضحًا للاختبارات المستقبلية، وضبط الأنظمة، والتقييم اللغوي للأنظمة الذكية.

الكلمات المفتاحية :التواصل بين الإنسان والذكاء الاصطناعي؛ أفعال الكلام؛ التداولية؛ أفعال الحوار؛ اللباقة؛ الكلام غير المباشر؛ النماذج اللغوية الكبيرة؛ استراتيجيات التصحيح.

1. Introduction

Large language systems can write fluent text. Yet fluent text does not always show real understanding. Many failures begin when a system reads words but misses the user's action. A user may ask, refuse, hint, warn, or apologize. Each act asks for a different reply. When the system answers the sentence form only, misunderstanding grows.

Speech act theory offers a strong lens for this problem. Austin (1962) showed that utterances do things. Searle (1969) later described key act types. Grice (1975) explained how speakers mean more than they say. Brown and Levinson (1987) linked indirectness and politeness to social risk. These ideas remain useful for human-AI talk.

Recent public work shows why the issue matters. Sravanthi et al. (2024) built PUB because pragmatics is still less studied than semantics. Ou et al. (2024) built DialogBench to test human-like dialogue ability. Maitra et al. (2025) showed that model dialogue acts differ from human dialogue acts. Orsini and Brunato (2025) showed that indirect speech acts remain hard for large models.

This paper asks three questions. Where does misunderstanding arise in human-AI talk. Which public results show recurring pragmatic errors. How can a speech act model reduce those errors. The paper answers these questions through secondary analysis of open resources and published experiments.

2. Research problem and aim

The core problem is simple. Current systems often optimize fluency, relevance, and task success. These goals matter, but they do not cover the full social work of language. A user may say, The room is noisy, as a complaint, a request, or a soft refusal. A fluent system may still answer the wrong act.

This gap becomes serious in tutoring, writing help, customer support, and public service chat. In these settings, a wrong act can waste time. It can also lower trust. In some domains, it may create risk. Wu et al. (2024) argued that social-pragmatic ability needs better evaluation and better tuning. Their point fits this paper's main claim.

The aim of this study is not to replace semantic or task models. The aim is to add a pragmatic layer. The proposed model reads a user turn as social action. It then plans a reply that fits the act, the relation, and the need for repair. The paper treats misunderstanding as a mismatch between user intention and system response.

3. Theoretical base

Austin (1962) separated three parts of an utterance. A speaker says something. A speaker also performs an act through that saying. The utterance may then create an effect on the hearer. This three-part view is useful for AI systems. It separates surface wording from intended action and from user outcome.

Searle (1969) gave a clearer map of speech act types. His framework includes assertives, directives, commissives, expressives, and declarations. Not every human-AI exchange needs all five types. Still, the framework helps a system decide what kind of reply belongs to a turn. A directive often needs action. An expressive may need acknowledgement first.

Grice (1975) showed that speakers often mean more than the literal sentence. Implicature matters because users rarely speak in fully direct form. Many systems still fail at this step. A literal answer can be fluent and wrong at the same time. Levinson (1983) later gathered these issues within a broad account of pragmatics.

Brown and Levinson (1987) added another key layer. Speakers manage face. They soften acts when the social risk is high. This is why indirect requests are common. It is also why politeness is not decoration. It changes meaning. Danescu-Niculescu-Mizil et al. (2013) later showed that politeness cues can be modeled in computational terms.

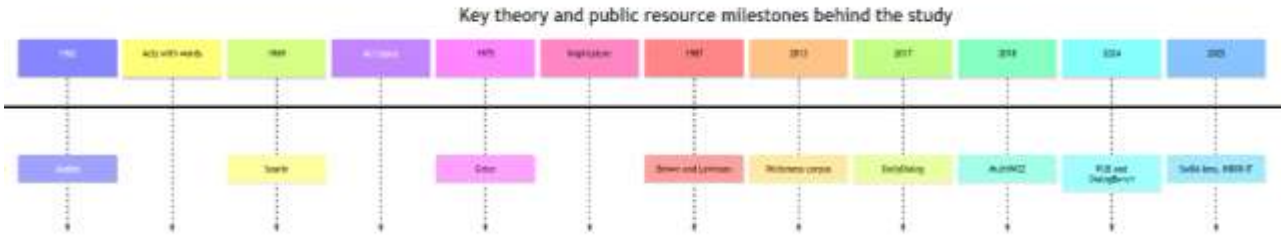

Figure 1 Key theory and benchmark milestones behind the study.

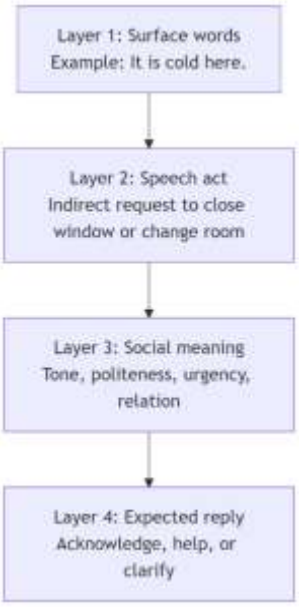

Figure 2 Speech act layers in human-AI interpretation.

4. Study design and public materials

This study uses qualitative secondary analysis. No new participants were recruited. No live model was tested for this draft. Instead, the paper brings together public corpora, benchmarks, and reported findings. The goal is to build a practical model from evidence that others can check and reuse.

The materials were chosen for four reasons. They are public. They address dialogue, pragmatics, or social meaning. They offer clear labels or clear experimental findings. They also connect to human-AI interaction. The final set includes the Switchboard Dialogue Act Corpus, DailyDialog, MultiWOZ, the Stanford Politeness Corpus, PUB, DialogBench, and INDIR-IT.

DailyDialog provides human-written daily conversation with communication intention and emotion labels (Li et al., 2017). MultiWOZ provides over 10,000 fully labeled task dialogues across domains (Budzianowski et al., 2018). The Stanford Politeness Corpus captures politeness in requests and supports near human-level classification (Danescu-Niculescu-Mizil et al., 2013). PUB adds large benchmark coverage for implicature, presupposition, reference, and deixis (Sravanthi et al., 2024). DialogBench adds dialogue evaluation across 12 tasks and 26 large models (Ou et al., 2024). INDIR-IT focuses on direct and indirect speech acts in Italian (Orsini & Brunato, 2025).

The study reads these materials through four analytic dimensions. The first is act type. The second is indirectness. The third is politeness and face work. The fourth is repair after confusion. These four dimensions shape the proposed model.

Table 1 Public corpora and benchmarks used in the study.

Resource	Public value for this study
Switchboard Dialogue Act Corpus (Fang et al., 2012; Maitra et al., 2025)	Open act labels for conversational moves and turn-by-turn interaction.
DailyDialog (Li et al., 2017)	13,118 human-written daily conversations with communication intention and emotion labels.
MultiWOZ (Budzianowski et al., 2018)	More than 10,000 task dialogues across domains for context and goal tracking.
Stanford Politeness Corpus (Danescu-Niculescu-Mizil et al., 2013)	Public request data for tone, face work, and politeness cues.
PUB (Sravanthi et al., 2024)	14 tasks, four pragmatic phenomena, and 28,000 data points for benchmark testing.
DialogBench (Ou et al., 2024)	12 dialogue tasks and tests on 26 large models for human-like dialogue evaluation.
INDIR-IT (Orsini & Brunato, 2025)	Public benchmark for direct and indirect speech act understanding.

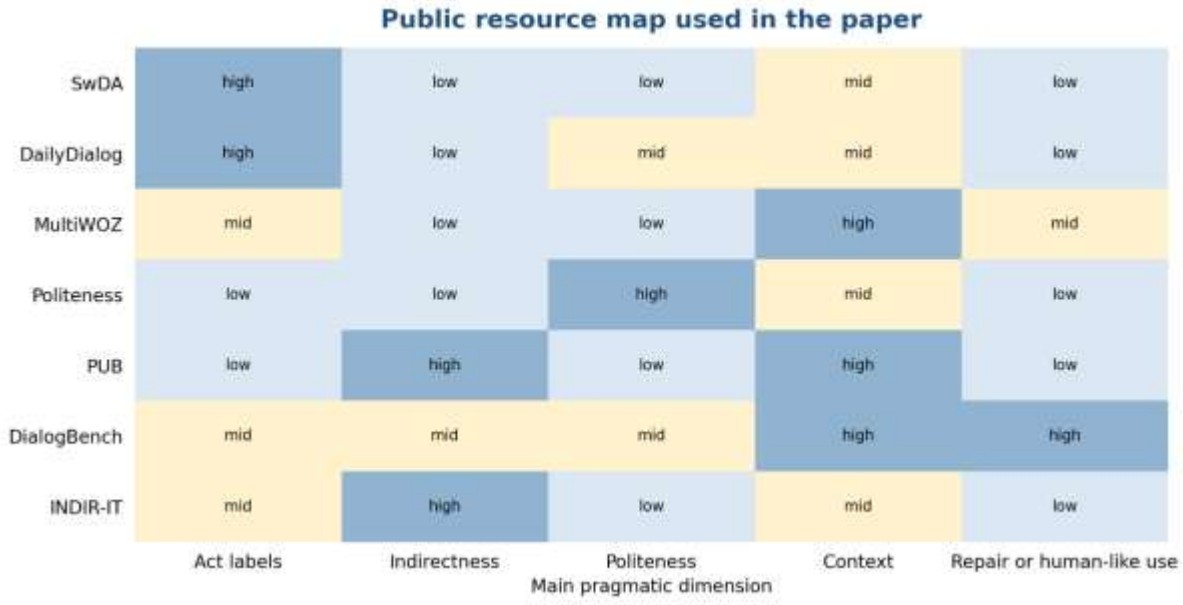

Figure 3 Public resource map used in the paper.

5. Public experimental evidence

The first body of evidence comes from pragmatic benchmarks. Sravanthi et al. (2024) reported that PUB contains 14 tasks across four pragmatic phenomena and 28,000 data points. The benchmark exists because model skill in pragmatics is not yet well studied. That finding matters here. It shows that misunderstanding is not a small edge case. It is a system design issue.

Wu et al. (2024) added a second point. They argued that multiple-choice formats are weak for social-pragmatic reasoning. They found that direct evaluation of free-form responses tracks human judgment better. They also found that preference tuning improves pragmatic ability more than standard supervised fine-tuning. This result supports a repair-aware response model, not a label-only model.

The second body of evidence comes from dialogue evaluation. Ou et al. (2024) built DialogBench with 12 tasks and tested 26 large models. They found that instruction tuning helps human-like dialogue only to a point. Most models still showed room for improvement. They also found that assistant positioning can weaken human emotional perception. This is highly relevant for systems that aim to sound helpful but still miss social fit.

The third body of evidence comes from dialogue acts. Maitra et al. (2025) used the Switchboard Dialogue Act Corpus to compare human responses and model responses. They found that dialogue act distributions differ in meaningful ways. They also noted that acts such as backchannel acknowledgement are harder for models to predict. This helps explain many awkward replies in human-AI talk. A reply may be informative, yet still feel off because it skips small but important interactional moves.

The fourth body of evidence comes from indirect speech. Orsini and Brunato (2025) reported that models handle conventional indirect speech acts better than non-conventionalized ones. Performance on harder cases depends more on model size and capacity. This is a strong warning against literal-only processing. Human speakers often use non-direct forms to reduce pressure, save face, or test the ground before a request.

The fifth body of evidence comes from politeness. Danescu-Niculescu-Mizil et al. (2013) built a politeness framework from Wikipedia and Stack Exchange requests. Their classifier used lexical and syntactic features grounded in politeness theory. It reached close to human performance across domains. This finding matters for human-AI design. It shows that tone management can be modeled in stable and testable ways.

Table 2 Main misunderstanding types and repair actions.

Failure type	Short example	Better system move
Indirect request read as statement	The room is very cold.	Offer likely help or ask one short check.
Wrong dialogue act	I guess that did not work.	Treat as complaint or repair need, not neutral comment.
Tone mismatch	Could you please revise this section.	Reply with respectful acknowledgement before action.
Reference gap	Can you do that again.	Use context to recover what that means.
Failed repair	No, I meant the last paragraph.	Update the act reading and confirm the new target.

Table 3 Public experiments and the design lesson taken from each one.

Public study	Main finding	Design lesson
Sravanthi et al. (2024)	Pragmatics needs direct benchmark attention.	Test act reading beyond semantics.
Wu et al. (2024)	Free-form pragmatic evaluation fits human judgment better.	Judge replies as interaction, not only as labels.
Ou et al. (2024)	Human-like dialogue remains limited in many models.	Include human-likeness and emotion in system review.
Maitra et al. (2025)	Model dialogue act patterns differ from human patterns.	Track missing acknowledgements and other small acts.
Orsini & Brunato (2025)	Hard indirect acts still challenge models.	Add indirectness detection and repair gates.
Danescu-Niculescu-Mizil et al. (2013)	Politeness cues can be modeled well across domains.	Use face-risk checks before final output.

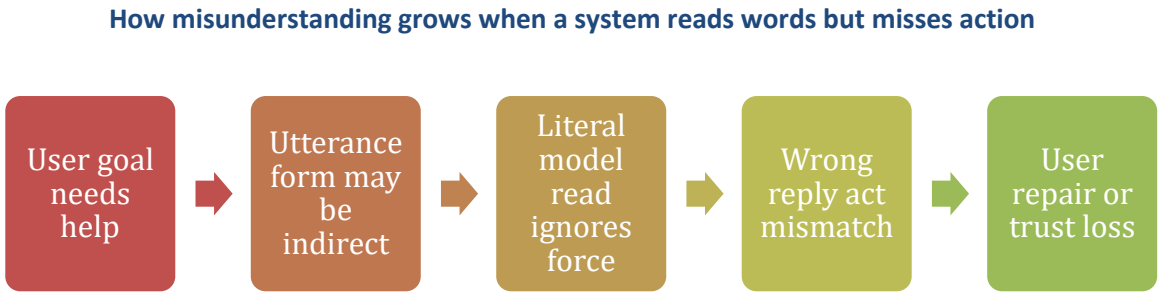

Figure 4 How misunderstanding grows when a system reads words but misses action.

6. Proposed speech act model

The proposed model has four linked stages. The first stage detects the likely speech act. It does not read only surface form. It asks what the user is doing with words. A statement may function as a request. A question may function as a challenge. An apology may also carry a request for repair.

The second stage reads context. It checks reference, deixis, prior turns, domain goals, and local constraints. This stage is important because many indirect acts depend on shared context. PUB is especially useful here because it covers implicature, presupposition, reference, and deixis (Sravanthi et al., 2024). Without this step, the system may map the right act label to the wrong situation.

The third stage checks relation and face risk. It estimates politeness, urgency, and possible social cost. Brown and Levinson (1987) showed why this matters. Danescu-Niculescu-Mizil et al. (2013) showed that these cues can be computed. A system that ignores this layer may complete the task but still sound rude, blunt, or oddly distant.

The fourth stage plans the reply. If the force is clear, the system should answer the act. If the force is not clear, the system should ask a short repair question. Repair should be brief and targeted. It should not create new burden. This design follows the main lesson from Wu et al. (2024). Pragmatic quality improves when systems are trained and judged on actual response behavior, not on narrow answer selection alone.

The model also includes a confidence gate. When the top act candidates are close, the system should avoid false certainty. It should ask one check. For example, this paragraph feels long may be a comment or a request. A better system can ask, would you like me to shorten it. That move protects accuracy and also preserves user control.

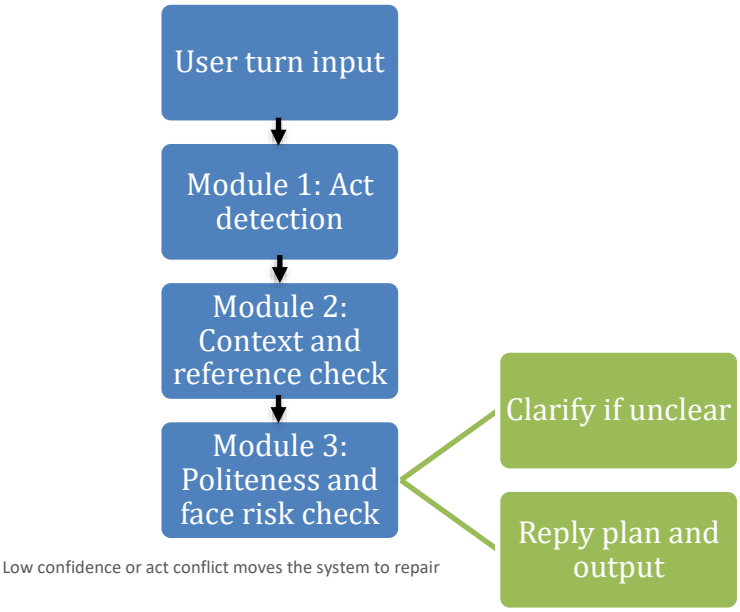

Figure 5 Proposed speech act model for better language interaction.

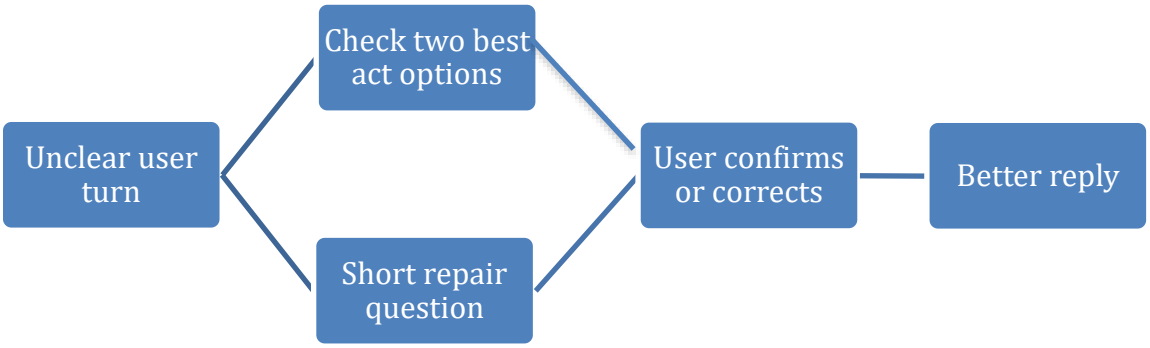

Figure 6 Repair loop when speech act force is uncertain.

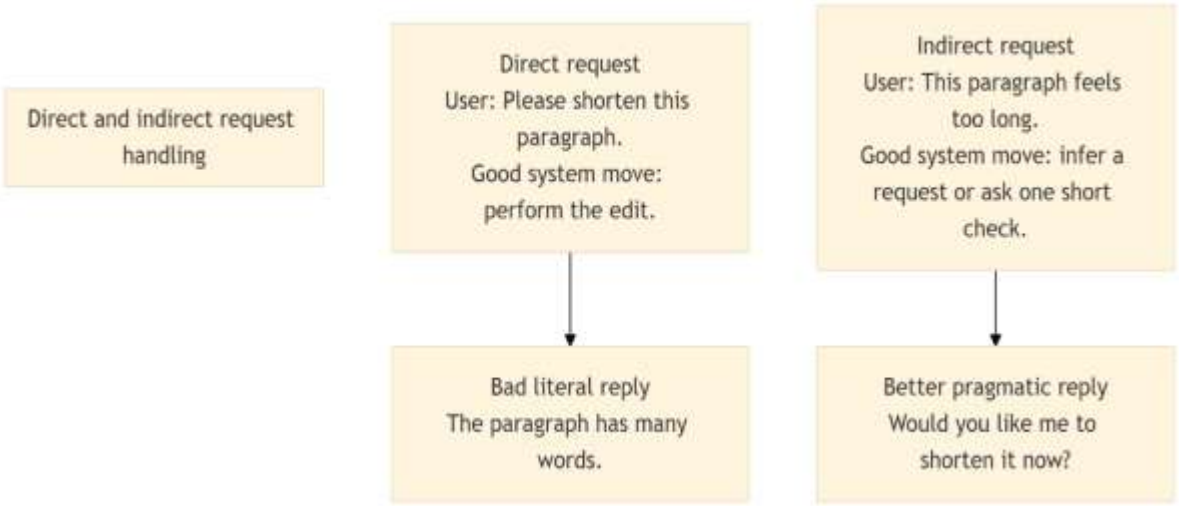

Figure 7 Direct and indirect request handling.

7. Practical value for intelligent language systems

The model fits several system types. A writing assistant can use it to separate comments from edit requests. A tutoring system can separate confusion from challenge. A customer support bot can separate complaint, request, and cancellation intent. A public service system can use repair when a user turn is vague or indirect. In all these cases, the main gain is better fit between action and reply.

The model also offers a clear training route. DailyDialog and Switchboard can support act detection. The Stanford Politeness Corpus can support tone control. MultiWOZ can support task-aware context reading. PUB and INDIR-IT can test indirectness and broader pragmatic reading. DialogBench can test whether the full system behaves in a more human-like way. The model therefore connects theory, data, and evaluation in one frame.

Another practical gain is better error analysis. Many current evaluations collapse several kinds of failure into one score. A speech act model separates them. A failure may come from act detection, context reading, tone control, or repair choice. This separation makes debugging easier. It also gives linguists a stronger role in system evaluation.

Table 4 A simple evaluation plan for future system testing.

Evaluation dimension	What to measure	Public source
Act match	Does the reply fit the user act.	SwDA, DailyDialog
Indirectness handling	Does the system recover intended force from non-direct wording.	PUB, INDIR-IT
Politeness fit	Does the reply match the social setting and level of face risk.	Stanford Politeness Corpus
Repair quality	Does the clarification reduce confusion with low extra burden.	DialogBench, live logs
Task fit	Does the system act on the right object, goal, or slot.	MultiWOZ

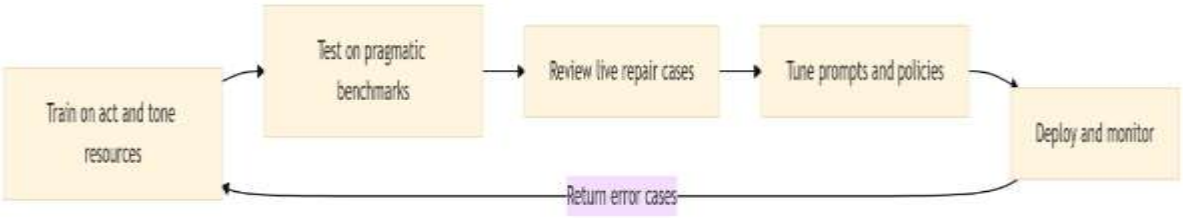

Figure 8 Evaluation cycle for a speech act aware system.

8. Discussion

This paper argues that misunderstanding in human-AI talk is often pragmatic before it is semantic. A system may know word meaning, yet still answer the wrong social action. The public evidence reviewed here supports that claim. Benchmarks of pragmatics are growing because the gap is real (Ma et al., 2025). Public experiments show the same pattern. Human-like dialogue needs more than correct facts or fluent sentences (Ou et al., 2024; Maitra et al., 2025).

Speech act theory helps because it treats language as action. That move is simple, but powerful. It tells system designers to ask not only What does this sentence mean. It also asks What is the user doing here. That shift reduces literal error. It also improves tone, acknowledgement, and repair. These small interactional gains can shape user trust over time.

The model also joins older linguistic theory with current computational work. Austin, Searle, Grice, Brown and Levinson, and Levinson offer the conceptual base. New resources such as PUB, DialogBench, and INDIR-IT offer public testing ground. The two sides work better together than alone. Theory gives the categories. Benchmarks show where the failures remain.

9. Limitations and future work

This paper has clear limits. It is a secondary study. It does not report a new live experiment. The public resources also lean toward English, although INDIR-IT adds an important non-English case. Politeness norms also vary across cultures and settings. A useful future study should test the model with multilingual users and live interaction data.

Future work should also measure repair success directly. Many systems ask clarification, but not all clarification helps. Some repair turns are too long. Some shift work back to the user. Future experiments should measure whether a short repair turn improves task success, user trust, and perceived respect. Wu et al. (2024) and Maitra et al. (2025) point in this direction, but more work is needed.

Another future step is domain tuning. Speech acts vary by setting. A writing tutor, a medical system, and a legal assistant do not face the same risks. The general model can stay the same, but the act inventory and repair rules should be adjusted for domain use.

10. Conclusion

Human-AI misunderstanding often begins when a system answers the sentence but misses the act. Public benchmarks and corpora show the same pattern from several angles. Indirect requests remain hard. Dialogue act balance remains weak. Politeness is measurable, but often underused. Repair still needs better design. These are not small style issues. They shape whether an interaction feels helpful, respectful, and clear.

The speech act model proposed here offers a practical response. It reads user turns as actions. It checks context and relation. It then replies or repairs with the user's likely intention in view. This does not solve every problem in human-AI talk. It does, however, give a strong linguistic path forward. Better systems must answer the act, not only the sentence.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

References

1. Austin, J. L. (1962). *How to do things with words*. Clarendon Press.
2. Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.
3. Budzianowski, P., Wen, T.-H., Tseng, B.-H., Casanueva, I., Ultes, S., Ramadan, O., & Gasic, M. (2018). MultiWOZ - A large-scale multi-domain Wizard-of-Oz dataset for task-oriented dialogue modelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 5016-5026). Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1547
4. Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., & Potts, C. (2013). A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 250-259). Association for Computational Linguistics. https://aclanthology.org/P13-1025/
5. Fang, A. C., Bunt, H., Cao, J., & Liu, X. (2012). Collaborative annotation of dialogue acts: Application of a new ISO standard to the Switchboard corpus. In *Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data* (pp. 61-68). Association for Computational Linguistics. https://aclanthology.org/W12-0509/
6. 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.
7. Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.
8. Li, Y., Su, H., Shen, X., Li, W., Cao, Z., & Niu, S. (2017). DailyDialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 986-995). Asian Federation of Natural Language Processing. https://aclanthology.org/I17-1099/
9. 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
10. Maitra, A., French, D., & von der Wense, K. (2025). Dialogue acts as a lens on human-LLM interaction: Analyzing conversational norms in model-generated responses. In *Proceedings of the Fourth Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCI+NLP)* (pp. 317-325). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.hcinlp-1.25
11. Orsini, M., & Brunato, D. (2025). Direct and indirect interpretations of speech acts: Evidence from human judgments and large language models. In *Proceedings of CLiC-it 2025: Eleventh Italian Conference on Computational Linguistics*. CEUR Workshop Proceedings. https://aclanthology.org/2025.clicit-1.79/
12. Ou, J., Lu, J., Liu, C., Tang, Y., Zhang, F., Zhang, D., & Gai, K. (2024). DialogBench: Evaluating LLMs as human-like dialogue systems. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. Association for Computational Linguistics. https://aclanthology.org/2024.naacl-long.341/
13. Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

14. 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* (pp. 12075-12097). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-acl.719
15. Aisha Mohamed Ahmed. (2026). Globalization and Technology in English: Structural Shifts, Digital Usage, and Cultural Implications. African Union Journal of Academic and Research Studies, 1(1), 40-55.
16. Wu, S., Yang, S., Chen, Z., & Su, Q. (2024). Rethinking pragmatics in large language models: Towards open-ended evaluation and preference tuning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 22583-22599). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.1258

