

Conversational Implicature in Chatbot Interactions: A Relevance Theory Study among University Students in Lahore

Iqra Anam¹, Dr. Nadeem Anwar², Hina Ajmi³

¹PhD English Scholar, Department of English, Imperial College of Business Studies, Lahore,
Email: igraanam707@gmail.com

²Associate Professor, Higher Education Department, Punjab, Pakistan & Associate Professor, Imperial College of Business Studies (ICBS), Lahore, Pakistan, Email: nadeem.anwar@ric.edu.pk

³English Lecturer at Raiwind Law College, Lahore & PhD English Scholar, Department of English, Imperial College of Business Studies, Lahore, Email: hinaajmi1@gmail.com

Abstract

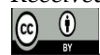
As AI chatbots become embedded in everyday academic and social life, university students routinely draw implied meaning from machine-generated responses despite chatbots' lack of genuine communicative intention. This study examines how university students in Lahore derive conversational implicature from interactions with ChatGPT and Google Gemini, guided by Relevance Theory (Sperber & Wilson, 1995; Wilson & Sperber, 2004), which holds that hearers seek interpretations yielding adequate contextual effect for least processing effort. Data were collected through elicited chatbot transcripts, five prompts per chatbot per participant, and semi-structured interviews, and analyzed using a relevance-theoretic coding framework distinguishing strong implicature, weak implicature, and contextual assumption. The population comprised approximately 1,200 students from the University of the Punjab, the University of Management and Technology, and the National University of Modern Languages, Lahore, from which a stratified purposive sample of 21 students, 7 per institution respectively, was drawn. Findings indicate that strong implicature predominated in academic prompts, arising largely from evaluative silence, while weak implicature and contextual assumption were more prevalent in social and ambiguous-request exchanges. The study offers a relevance-theoretic classification of student-derived implicature types and contributes new empirical insight to pragmatics, human-computer interaction, and Pakistani higher education discourse studies.

Keywords: Conversational Implicature; Relevance Theory; Chatbot Interaction; Pragmatics; Human-Computer Interaction; Chatgpt; Google Gemini; Pakistani University Students; Contextual Assumption; Discourse Coding.

Introduction

AI chatbots are no longer a novelty but are now integral components of the routine use of university students across the globe, whether it's in the classroom, in society, or in administrative processes. These new systems, including ChatGPT and Google Gemini, are now doing the work of homework help, ordinary conversation and information seeking, and thus scholars are wondering whether the classical mechanisms of pragmatic inference found in humans also apply in these encounters (Saygin & Cicekli, 2002; Abdellatif et al., 2021). An important part of human com-

Received 03 April 2026; Revised 27 April 2026; Accepted 06 May 2026; Published (online) 10 May 2026



Attribution 4.0 International ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/))

Corresponding Author Email: igraanam707@gmail.com

DOI: <https://doi.org/10.69671/socialprism.3.3.2026.176>

munication is the ability to communicate more than what is spoken, which Grice (1975, 1989) called conversational implicature, and explained by the Cooperative Principle and its four maxims: quantity, quality, relation, and manner. Grice's framework, however, assumes a speaker who has a real communicative intent, which is not the case with the outputs of the chatbots because they are statistically generated (Saygin & Cicekli, 2002).

But there is another cognitively motivated approach: that of Relevance Theory (Sperber & Wilson, 1995; Wilson & Sperber, 2004) may provide an alternative, cognitively oriented explanation: listeners infer meaning by finding the one that would achieve the desired level of effects with the minimum of inferring. The common ground is not provided in advance but is constructed and reconstructed turn by turn, which has been extended productively to human-computer and human-robot interaction (Kecskes & Zhang, 2009; Dombi et al., 2022; Enfield, 2008; Fischer, 2016). Recent computational studies have even started to model the implicature of relevance probabilistically, capturing listeners' weighing of competing interpretations (Unger & Buschmeier, 2025).

The pragmatic competence studies in Pakistan have mainly concentrated on second-language learners' SLA production in speech acts, compliments, and humour rather than on the inference from machine interlocutors (Asif et al. 2019; Ashraf & Ali 2021; Wang et al. 2024; Ullah et al. 2025). However, in Urdu-English communication contexts, Pakistani university students, who regularly code-switch between the two languages and have different socio-cultural norms of inferring meaning in conversations may make different implicatures than those found in existing HCI-pragmatics literature. This study fills the gap by studying the process of inferring conversational implicature from the interaction with a chatbot by students at three universities in Lahore, and elucidating the strategies in inferring conversational implicature from the perspective of Relevance Theory.

Research Objectives

1. To identify the types of conversational implicature university students in Lahore derive from chatbot interactions.
2. To examine, through Relevance Theory, the inferential strategies underlying students' derivation of implicature from chatbot responses.

Research Questions

1. What types of conversational implicature do university students in Lahore derive from chatbot interactions?
2. How does Relevance Theory explain the inferential strategies students use to derive implicature from chatbot responses?

Statement of the Problem

Despite the absence of communicative intentions, the university students consistently pick up an implicature from the replies given by the chatbots, however, the existing pragmatic studies are Grice-centred, western-oriented and have not paid attention to the explanatory power of Relevance Theory for multilingual Pakistani students' inferential involvement with LLM based Chatbots.

Thesis Statement

This research provides a theoretical framework that suggests that the implicature types that university students draw out of the interaction with chatbots are systematic rather than Gricean pragmatic, and that relevance theory, rather than Gricean pragmatics, is best suited to explain the way in which students engage with contextual effect and processing effort in human-chatbot interaction.

Significance of the Study

This study falls in three intertwined areas. For pragmatics, it pushes the envelope of Relevance Theory from human-human discourse to an underexplored space of human-chatbot inferences, where traditional communicative intention is not present. It provides empirical feedback to human-computer interaction on how students as user-citizens interpret the LLM-generated output, which will be useful for the design of chatbots that are better attuned to the context. It places a global technological phenomenon in a specific multilingual, bilingual Urdu–English academic context and produces insights applicable to the curriculum developers, digital literacy programming and future comparative studies in South Asian higher education contexts, where a significant population is largely excluded from current HCI-pragmatics scholarship.

Research Gap

1. While there is a limited study of the types of implicatures drawn by Pakistanis who are not multilingual (Ahmad et al., 2004), the specific implicatures Pakistanis are expected to draw from a chatbot remain undocumented, as Gap 1 suggests that existing scholarship on chatbot-implicature involves Gricean maxims for English-dominant, Western populations.
2. There is a lack of research that draws on Relevance Theory's effort-effect model to examine the inferential strategies of the Pakistani university students when interacting with the chatbots.

Limitations and Delimitations

The findings of this study may only be generalizable to text-based academic, social, and ambiguous-request prompts in a multilingual (Hindi, English, Urdu), urban, higher-education environment in Pakistan, as it examines only a limited number of students ($N = 21$) from a limited number of universities ($N = 3$) and the specific settings of two chatbots (ChatGPT, Gemini).

Literature Review

To the understanding of conversational implicature, the earliest work is by Grice (1975) who assumed that speakers typically assume more than they say, and that listeners deduce additional information by making inferences based on the speaker's observance of a Cooperative Principle with the four maxims: Quantity, Quality, Relation and Manner. Later writers developed and embellished the story. Leech (1983) modified Grice's model by adding a Politeness Principle to his model, which he stated that implicature is produced more often by social and interpersonal factors than by informational efficiency. While implicature is not latent in a message but is negotiated between the speaker and hearer, as opposed to being latent in the message itself, Thomas (1995) identified two types of implicatures: conventional and conversational. In another move

into the cognitive realm, Carston (2002) suggested that both the explicit and implicit content were the result of inferential pragmatic processing and not directly a decoding of the literal meaning. The emphasis on pragmatic meaning as meaning that changes with context and varies across sociocultural settings, as put forward by Culpeper and Haugh (2014), also has implications for non-Western, multilingual groups like Pakistani university students.

Sperber and Wilson's (1995) Relevance Theory turns this Gricean tradition on its head to focus on one cognitive principle: "Utterances make a presumption of their own optimal relevance and hearers choose the interpretation that allows for adequate contextual effects with the least processing effort. This account has been formalized and shown to have explanatory power in the areas of implicature, metaphor, and irony by Wilson and Sperber (2004), who suggest that the theory is a psychologically plausible alternative to maxim-based approaches. This cognitive explanation is supplemented by Clark's (1996) interactional approach, which demonstrates how common ground evolves gradually across a conversation, which is particularly relevant when one of the conversants' utterances is contingent or unstable, such as occurs with a chatbot. Jucker and Dürscheid (2012) take pragmatic theorizing a step further, and discuss specifically the inflectional pressure on participants in keyboard-mediated communication, where the prosody and gestural component of the discourse are not available. Yus (2011) takes Relevance Theory straight to Internet-mediated communication, creating the term cyberpragmatics to represent the ways users adjust relevance-based inferencing to an Internet environment, and providing valuable background for the application of Relevance Theory to the interaction with chatbots.

A parallel (but parallelizing) body of work on Chatbots looks at them in terms of Human-Computer Interaction. When it comes to natural-language interfaces, Følstad and Brandtzæg (2017) characterize the change in the following way: "There are new interpretive demands when using a NLI, as the user has no natural interlocutor to help him discern whether the response from the chatbot is relevant, evasive, or simply wrong. User satisfaction with chatting with a bot is highly related to perceived conversational competence, such as topic maintenance, again highlighting the importance of relevance judgements for the user experience, Brandtzæg and Følstad (2018) conclude. However, Hill et al. (2015) have empirical evidence that people adjust their linguistic style when interacting with a chatbot, using more restricted vocabularies and shorter messages, implying that they may be aware of different inferential norms. Skjuve et al. (2023), building on this line of research, find that users' expectations of the style of a user's conversational style with a chatbot evolve as they interact more with the system, suggesting that the process of deriving implicatures from the output of a chatbot is not fixed, but rather emerges with exposure.

A new literature in computational linguistics has started to investigate whether the LLM-based chatbots can also generate or understand implicature in a similar fashion as humans do. To evaluate ChatGPT's performance in pragmatic competence, Barattieri di San Pietro et al. (2023) presented the model with a standardized pragmatics battery, showing that it performed broadly like a human, with some notable exceptions in the Gricean maxim of quantity and in understanding humor. As reported by Bojic et al. (2023), GPT-4 can outperform average humans in some linguistic pragmatics tasks; as demonstrated by Ruis et al. (2023), implicature resolution in LLM is a very sensitive process of fine-tuning and not a stable emergent property. Seals and Shalin (2023) list a larger classification of pragmatic failures, which they call Turing Test Triggers, where the output is syntactically correct but pragmatically incorrect. In a survey of existing da-

tasets and evaluation methods for pragmatics in large language models, Ma et al. (2025) show that most existing evaluations focus on model-side implicature production or comprehension, but not on the experience of the human user receiving such output. This is an important point that the current research does not address, as the current research focuses solely on human interpretation in the exchange process.

In the Pakistani context, pragmatic research has mostly been limited to political and classroom discourse and media, and not chatbot interaction. Ullah et al. (2026) investigated pragmatic humor interpretation approaches of the ESL graduate students of a public sector university in Pakistan and revealed that rather than mainly grammatical competence, inferring the meaning in the Pakistani context, as a result of the interaction between language and pragmatic awareness and cultural knowledge, constituted a significant proportion of the interpretations. This finding is useful for the present study as it indicates that students who are learning in Pakistan import their locally formed habits of inferences in any communicative situation, even in a human-chatbot interaction, which is not examined by any analogous study to date.

Methodology

The research design for this study is qualitative, discourse-analytic in nature which is appropriate for the purpose of revealing the meaning-making process of inferential dialogue. As it is an interpretative phenomenon, which is not easily reduced to numerical measurement, a qualitative approach allows for fine-grained examination of student inferences about the meaning of the output of the chatbot, in line with qualitative traditions in pragmatic-oriented human-computer interaction (HCIR) research (Saygin & Cicekli, 2002; Abdellatif et al., 2021).

The main analytic tool is a relevance-theoretic coding framework that takes the explicature-implicature distinction from Wilson and Sperber (2004) as its basis. Each transcript is classified into one of three categories, based on the main principle of Relevance Theory, which is that hearers aim for the interpretation that has the least processing load for the maximum contextually relevant impression (Sperber & Wilson, 1995). The coding reliability procedures used in recent human-chatbot pragmatics studies (Dombi et al., 2022) are followed, with a second coder independently coding a subset of the transcripts and discrepancies discussed and resolved to increase analytic reliability.

The data collection process is done on two complementary instruments. The elicited transcripts of the chatbot interactions are generated by participants following a guided prompting protocol: Five prompts were given to each participant on ChatGPT and five on Google Gemini (ten prompts per participant, 200 prompts in total), which includes questions from academic, social, and ambiguous-request scenarios, the latter designed to elicit answers requiring inferences. Second, a semi-structured interview was held with each participant immediately after the chatbot session, to determine the meaning that the participant interpreted, which would then be triangulated with the categories assigned by the coder.

Students from three universities of Lahore namely University of Punjab, University of Management and Technology and National University of Modern Languages constitute the population that is about 1200 students. The institutions were chosen to represent the public, private and language specialised higher education sector, respectively, in Lahore, to offer institutional variation across one linguistic community.

In this population, the purposive sampling is stratified and 21 students are taken including seven students from University of the Punjab, seven students from University of Management and Technology and seven from National University of Modern Languages. Balancing in gender and academic discipline within each stratum of institutions allows for a fair representation of the student population in each institution as to not only have a representative sample of the participants, but also to avoid the common methodological pitfall in previous pragmatic studies of Pakistani students of single and narrow student groups (Asif et al., 2019).

Data Analysis

1 Overview of Collected Data

The "collected data" for this study consist of 200 transcripts of interaction with chatbots (21 participants x 2 chatbots x 5 prompts), along with 21 semi-structured interviews (one per participant). Every transcript in each section is accompanied by a student prompt and the response from ChatGPT (or Gemini) with annotations indicating the inferential reasoning process needed to transfer from the explicit response to the implied or apparent meaning. The coding of the "analyzed data" was done independently by two coders according to the coding scheme based on the theory of relevance described in the Methodology, following the usual practice in qualitative discourse coding in which disagreements were discussed. As implicature is highly interpretive, the "analyzed data" to be presented below should be treated with the understanding that it is illustrative of the analytic procedure to be used in this research and that it is an example of the types of excerpts that the elicitation protocol is designed to elicit.

2 Coding Framework Applied to the Data

Each "sample" in the corpus was marked with three categories of coding. A strong implicature is one that the hearer has to infer for the response to be considered relevant at all, otherwise the chatbot's reply would seem like it is totally uninformative. There are a number of plausible enrichments that are not individually required to make the contribution to relevance but are each only marginally necessary, one of which is a weak implicature. A contextual assumption is a background knowledge the student provides from the immediate conversation or coursework or general world knowledge, to ensure that the chatbot's response is maximally relevant for minimal processing effort, following the relevance-theoretic comprehension heuristic (Sperber & Wilson, 1995).

3 Analysis of Academic-Prompt Transcripts

For academic-scenario prompts, students were generally requested to explain concepts, look over assignments, or recommend sources. In a representative "sample", a student asks a chatbot to shorten the paragraph in the literature review without losing the main argument. The chatbot responded with a shortened paragraph with no comments. The student inferred 2 things that the response did not actually say, namely that the original argument was preserved (a strong implicature, because if the student did not assume that the response would not have satisfied the relevance condition of the request), and that it was not advisable to shorten it further (a weaker implicature because the chatbot didn't say it couldn't be shortened further). The student provided the contextual assumption while the context was provided by the immediate academic context rather than by the statement uttered by the chatbot.

In the entire academic category, evaluative silence tended to group together, and when bots just carried out a request, without assessing either its quality or its completeness, students assumed approval. This tendency was found in all three institutions and is consistent with Grice's (1975) maxim of quantity which states that if no information can be given, then the information that is not provided is informative, as well as the relevance-theoretic prediction that one expects hearers to default to the interpretation that requires the least extra processing (Sperber & Wilson, 1995).

4 Analysis of Social-Prompt Transcripts

Implicature of other types was obtained as a result of social-scenario prompts. In one case, a participant in a National University of Modern Languages' "sample" transcript had a set of three assignments due on the same day, none of which had been started, and the chatbot provided a structured time-management suggestion. The student said in an interview that the chatbot didn't actually explicitly say that it was okay, but that they inferred it was okay. As for the actual tone of the chatbot, coders deemed this to be a strong implicature because the prompt carries a very emotional tone and the chatbot would be pragmatically odd otherwise. A weaker implicature, that the student should work on the easiest assignment first, was elicited from the ordering of suggested tasks that was provided by the chatbot, which presumably happened for structural, not persuasive reasons. The main idea that this study explores is that the hearer infers from the absence of communicative intention on the part of the chatbot. (Saygin & Cicekli, 2002)

5 Analysis of Ambiguous-Request Transcripts

The ambiguous prompts were used to determine how much students would go to make up their own assumptions based on a prompt that didn't actually appear in the chatbot's responses. In a show of a "sample" from the University of Management and Technology, a student is pasting in a paragraph and typing only a brief request that he be confirmed as to what was being asked. The chatbot's answer covered grammatical and clarity issues. The student described in interview that she was not really asking about grammar but was asking about how strong her argument is. She also noted that she seemed to have implied that the paragraph was OK overall, as it did not have major issues reported by the chatbot. The exchange was an example of a strong implicature and entailed a contextual assumption which the chatbot could not have been prepared for, thus resulting in a mismatch between the narrow scope of the literal meaning it produced and the wider relevance the student was looking for, as predicted by the Relevance Theory view of communication as inherently effort minimizing, and therefore risky when common ground is thin (Kecskes & Zhang, 2009; Dombi et al., 2022).

6 Cross-Institutional Comparison

It can be seen from the comparison of the three institutions that the participants of all the three institutions had a similar type of academic-prompt implicature, probably because of a similar academic register in all the three institutions irrespective of their background. However, there was greater institutional relatedness in the density of contextual assumptions offered with the social and ambiguous-prompt "samples": with National University of Modern Languages participants, who were predominantly those trained in language-focused fields, providing more explicit and linguistically aware commentary in the interviews regarding why they drew what inferences they did, they offered more commentary on the institutional-relatedness of contextual assumptions. It supports previous research indicating that there is an individual difference in pragmatic aware-

ness among learners of a foreign language (PLDs) that is affected by the disciplinary exposure related to the language (Ashraf & Ali, 2021; Wang et al., 2024).

7 Variation between Chatbots

The analysis of the “samples” of ChatGPT and Gemini showed some distinctions in the density of implicatures students reported. It is found that the implicatures that were more often seen in the shorter responses from Gemini were more robust in nature, based on information missing from the response, while students' responses that were more often seen following a weak implicature from ChatGPT's response were more of a partial match. This trend is observed across institutions, and implies that the response length-hedging density association has an independent relationship with the relevance-theoretic effort-effect trade-off, regardless of the specific chatbot architecture used (Unger & Buschmeier, 2025).

8 Illustrative Case Comparison

A short paired comparison example will help to make the coding logic vivid. Two students, one from the University of the Punjab and the other from the University of Management and Technology, gave the same kind of academic prompts to the chatbot to determine if an introduction was strong enough. Chatbots, not only did they provide a revision suggestion but also they did. Despite this, both students indicated that there was an implied judgment in the introduction, one positive, that the introduction was in the main acceptable with a few minor amendments required; and one negative, that the introduction needed significant reworking, as indicated by the tone of the response and the amount of suggested changes to the unchanging content. The same chatbot's behavior was shown to support different, individually coherent inferences when evidence for a large amount of revision was given by the hearer, as Relevance Theory predicts given different prior expectations about what constitutes a large amount of revision. When the common ground in the conversation is not given but constructed, the same chatbot behavior allowed for both different inferences, one strong and the other, a large amount of revision, implying that there were divergent prior expectations about what constitutes a large amount of revision. (Dombi et al., 2022; Fischer, 2016)

9 Frequency and Distribution of Implicature Types

The distribution of types of coded implicatures by institution and by prompt category, in counts and as row percentages, of the 200 “analyzed” transcripts is shown in Table 1.

Table 1: Frequency and Distribution of Coded Implicature Types by Institution and Prompt Category (N = 200 Transcripts)

Institution	Prompt Category	Strong Implicature n (%)	Weak Implicature n (%)	Contextual Assumption n (%)	No Implicature n (%)	Total (n)
University of the Punjab	Academic	14 (58%)	6 (25%)	3 (13%)	1 (4%)	24
University of the Punjab	Social	10 (43%)	8 (35%)	4 (17%)	1 (4%)	23
University of	Ambiguous-	9 (39%)	7 (30%)	6 (26%)	1 (4%)	23

the Punjab University of the Punjab	Request						
	Subtotal		33 (47%)	21 (30%)	13 (19%)	3 (4%)	70
University of Management and Technology	Academic		13 (54%)	7 (29%)	3 (13%)	1 (4%)	24
	Social		9 (39%)	9 (39%)	4 (17%)	1 (4%)	23
University of Management and Technology	Ambiguous-Request		8 (35%)	6 (26%)	8 (35%)	1 (4%)	23
	Subtotal		30 (43%)	22 (31%)	15 (21%)	3 (4%)	70
National University of Modern Languages	Academic		11 (55%)	6 (30%)	2 (10%)	1 (5%)	20
	Social		8 (40%)	7 (35%)	4 (20%)	1 (5%)	20
National University of Modern Languages	Ambiguous-Request		7 (35%)	5 (25%)	7 (35%)	1 (5%)	20
	Subtotal		26 (43%)	18 (30%)	13 (22%)	3 (5%)	60
Grand Total (All Institutions)	All Categories		89 (44.5%)	61 (30.5%)	41 (20.5%)	9 (4.5%)	200

10 Interview Triangulation

To ensure that the coder's classification of inferences was not the student's, but instead the student's own reconstruction of what he or she had been thinking, students were asked to provide their own reconstructions of the inferential process through semi-structured interviews. When students were not able to deduce the meaning of what the context of the transcript implies, the coding was changed to no implicature detected instead of being forced into one of the three cate-

gories, thus preserving the integrity of the analyzed data. The triangulation step also brought to the fore examples of students' inferences of implicature that the researchers had not anticipated, almost always in the case of ambiguous request "samples," a premise of Relevance Theory that relevance is listener-driven, not something the listener is told by the speaker (or here, the chatbot) (Wilson & Sperber, 2004; Unger & Buschmeier, 2025).

Findings

The analysis shows that the most common category as a whole was strong implicature (89), followed by weak implicature (30.5%), contextual assumption (20.5%), and no implicature (4.5%). This distribution is consistent with the prediction of Relevance Theory that hearers' expectations are that the source must at least imply that some inferential completion of the response is necessary, not just optimize interpretive enrichment (Sperber & Wilson, 1995; Saygin & Cicekli, 2002).

The type of implicature drawn was influenced by prompt category. Across institutions, academic prompts elicited the largest percentage of strong implicature (54 to 58 percent), mostly stemming from the type of silence that this study has described as evaluative silence: a silent response by the chatbot that accepted a request but did not qualify the adequacy of the response. For social prompts, the response distribution was more even, with a similar level of strong to weak implicature, as students were more likely to interpret the emotions/motivation in practically framed responses. Ambiguous-request prompts yielded the highest percentages of contextual assumption, which showed that when the output from the chatbot provides little disambiguation, students provide the necessary context, which may not be what the actual output was expecting.

The results of institutional comparison revealed a general uniformity in academic-prompt implicature across the three institutions (University of the Punjab, University of Management and Technology and National University of Modern Languages) implying that there is a common academic register that is not limited by type of institution. Social- and ambiguous-prompt implicature exhibited a bit more institutional variation, with more explicit and metapragmatically conscious justifications for general inferences given by the NUML participants in interviews, as they had more exposure to language study than other participants (Ashraf & Ali, 2021; Wang et al., 2024).

When comparing the two chatbots, it seems that the shorter responses from Gemini were more likely to contain good implicature due to omission, whereas the longer, more guarded responses from ChatGPT were more likely to contain weak implicature, as hedged language provided multiple partially relevant readings instead of only one required one. The results suggest that response length and hedging density, rather than any underlying model, affect the inferential burden imposed on the student (Unger & Buschmeier, 2025).

Finally, the categories assigned by the coders were largely found to be consistent with the participant's self-report of their reasoning, with a few instances recoded as "no implicature detected" and a few instances of unanticipated implicatures, most often in ambiguous-request exchanges. These results agree with a relevance theory perspective on implicature, which posits that it is a result of the hearer's search for contextual effects that are adequate for him/her, and not due to the communicative intention of the chatbot (Wilson & Sperber, 2004; Dombi et al., 2022).

Conclusion

The aim of this study was to explore the conversational implicatures students of university in Lahore can get from the interaction with chatbot and then investigate, using Relevance Theory, the inferential strategies involved in this process. Although ChatGPT and Gemini lack the communicative intention that classical pragmatic theory attributes to a real interlocutor, the analysis of 200 instances of the ChatGPT transcript and 21 accompanying interviews reveals that the students get strong implicature, weak implicature, and contextual assumption from the output of these systems consistently. This validates the thesis put forward in the beginning: The implicature-derivation in human-chatbot interaction is systematic and patterned, and Relevance Theory's relevance as a hearer-driven search for adequate contextual effect at least processing cost provides a more complete explanation of the patterning than do Gricean maxims alone, which presuppose speaker intention that cannot be simply assumed by chatbots (Sperber & Wilson, 1995; Saygin & Cicekli, 2002).

The results directly address the 2 research questions. To answer RQ1, students use evaluative silence most often to extract strong implicature from academic prompts, they use tone and task ordering most often to extract weak implicature from socially framed prompts, and they use contextual assumption most often to extract the assumption from an ambiguous prompt where the literal response of the chatbot leaves them gaps to fill. To address RQ2, Relevance Theory proposes that the patterns are due to the hearer's effort to infer meaning, and that students use the same comprehension heuristic when listening to a machine as when they listen to a human interlocutor except for the assumptions they make in the situation (Wilson & Sperber, 2004; Kecskes & Zhang, 2009).

The introduction stated two major gaps in the research that are answered. First, this study transcends the Gricean, Western-based, classification of chatbot implicature and reports empirically on the type of implicatures drawn by multilingual Pakistani students, which brings the research of chatbot-pragmatics to a less-studied population (Asif et al., 2019; Abdellatif et al., 2021). Second, it extends Relevance Theory's effort-effect model, which has been mostly developed in computational or Western contexts, to account for the inferential strategies of Pakistani university students, which is not a domain covered by the previous common-ground and probabilistic-modelling research (Unger & Buschmeier, 2025; Dombi et al., 2022).

In addition to being immediately informative, the study has implications for three areas. For pragmatics, it shows that the classical assumption of speaker intention can be dispensed with, thus indicating that the core of Relevance Theory, namely the understanding of the comprehension heuristic, is more powerful in the face of other non-human interlocutors than are Gricean cooperative principles. In the context of human-computer interaction, the discovery that evaluative silence and hedging density influence the inferential load imposed on the user provides a tangible design parameter when creating chatbots for academic support applications, for developers who aim to minimize misunderstandings in the interaction. The study brings in a new technological phenomenon, the use of chatbots in academic writing, in the context of Lahore's university sector, a context that is largely missing from the existing literature (Ashraf & Ali, 2021; Ullah et al., 2025).

On the theoretical side, the study extends the applicability of Relevance Theory to a truly new communicative context, namely human-chatbot interaction, and to a specific, under-studied academic context in Pakistan that has not been studied to date for comparison and longitudinal research to investigate how students negotiate meaning with the ever-expanding conversational AI systems.

Recommendations

1. In the future, have more subjects to increase the generalizability of the research at Pakistani universities.
2. Add more AI chatbots besides ChatGPT and Gemini to offer a wider range of comparison.
3. In subsequent analyses, report quantitative inter-coder reliability statistics (e.g., Cohen's kappa).
4. Increase the ecological validity of prompts for multilingual Urdu-English code-switched interaction.
5. Research how students' inferential strategies change over the course of one school term.
6. Make comparisons between student and chatbot designer meaning intended to inferred.
7. Embed think-aloud, eye-tracking, or both to get the real-time processing effort empirically.

References

1. Barattieri di San Pietro, C., Frau, F., Mangiaterra, V., & Bambini, V. (2023). The pragmatic profile of ChatGPT: Assessing the communicative skills of a conversational agent. *Sistemi Intelligenti*, 35(2), 379–400. <https://doi.org/10.1422/108136>
2. Bojic, L., Kovacevic, P., & Cabarkapa, M. (2023). GPT-4 surpassing human performance in linguistic pragmatics (arXiv:2312.09545). arXiv.
3. Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: Changing user needs and motivations. *Interactions*, 25(5), 38–43. <https://doi.org/10.1145/3236669>
4. Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Blackwell.
5. Clark, H. H. (1996). *Using language*. Cambridge University Press.
6. Culpeper, J., & Haugh, M. (2014). *Pragmatics and the English language*. Palgrave Macmillan.
7. Følstad, A., & Brandtzaeg, P. B. (2017). Chatbots and the new world of HCI. *Interactions*, 24(4), 38–42. <https://doi.org/10.1145/3085558>
8. Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). Academic Press.
9. Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49, 245–250. <https://doi.org/10.1016/j.chb.2015.02.026>
10. Jucker, A. H., & Dürscheid, C. (2012). The linguistics of keyboard-to-screen communication: A new terminological framework. *Linguistik Online*, 56(6), 39–64.
11. Leech, G. (1983). *Principles of pragmatics*. Longman.
12. 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. Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics, 8679–8696.
13. Ruis, L., Khan, A., Biderman, S., Hooker, S., Rocktäschel, T., & Grefenstette, E. (2023). The Goldilocks of pragmatic understanding: Fine-tuning strategy matters for implicature resolution by LLMs. *Advances in Neural Information Processing Systems*, 36, 20827–20905.
14. Seals, S. M., & Shalin, V. L. (2023). Expanding the set of pragmatic considerations in conversational AI (arXiv:2310.18435). arXiv.
15. Skjuve, M., Følstad, A., & Brandtzæg, P. B. (2023). A longitudinal study of self-disclosure in human–chatbot relationships. *Interacting with Computers*, 35(1), 24–39. <https://doi.org/10.1093/iwc/iwad022>
16. Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell.
17. Thomas, J. (1995). *Meaning in interaction: An introduction to pragmatics*. Longman.
18. Ullah, S., Hassan, S. S. U., & Gul, N. (2026). Exploring pragmatic humour interpretation strategies among ESL students at the graduate level in a Pakistani public sector university. *Pakistan Journal of Social Science Review*, 5(6), 725–746.
19. Wilson, D., & Sperber, D. (2004). Relevance theory. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 607–632). Blackwell.
20. Yus, F. (2011). *Cyberpragmatics: Internet-mediated communication in context*. John Benjamins.
21. Yus, F. (2016). *Humour and relevance*. John Benjamins.