LinguaEducare: Journal of English and Linguistic Studies
ISSN: 3064-0547 (e)
2026, Vol. 3, No. 2, page 155-168
https://journal.ciptapustaka.com/index.php/LEC

OPEN ACCESS

When AI Sounds More Empathetic Than Humans: Experimental Evidence from Indonesian EFL Students

Febriyanti
English Department, STKIP PGRI Bandar Lampung, Indonesia

ABSTRACT

Large language models (LLMs) increasingly participate in emotionally sensitive communication, yet it remains unclear how their empathic language is perceived by English as a Foreign Language (EFL) learners. This study investigated whether Indonesian EFL students perceived LLM-generated responses as more empathetic than human-generated responses and examined the linguistic characteristics shaping those perceptions. A mixed-methods within-subject experimental design was employed with 100 undergraduate students in the English Education Study Program at STKIP PGRI Bandar Lampung, Indonesia. Participants evaluated paired human- and LLM-generated responses to ten standardized emotional vignettes under blind-source conditions. Perceived empathy was assessed across emotional understanding, validation, warmth, perspective-taking, supportiveness, and contextual appropriateness, while semi-structured interviews with 18 purposively selected participants explored the linguistic features influencing their judgments. Quantitative findings indicated that LLM-generated responses received higher overall empathy ratings than human-generated responses. Qualitative analysis showed that emotional acknowledgment, validation, reassurance, supportive guidance, lexical warmth, and non-judgmental wording contributed strongly to perceptions of empathy. Human responses, however, were more frequently associated with authenticity and contextual specificity, whereas some LLM responses were perceived as polished but formulaic. These findings suggest that perceived empathy in human–AI communication is shaped largely by linguistic performance and reader interpretation rather than by evidence of genuine emotional experience. The study extends research on artificial empathy into the EFL context and highlights the importance of critical AI literacy in helping pre-service English teachers distinguish persuasive empathic language from authentic human emotional understanding. It also demonstrates value of blind comparisons for examining interpersonal meaning in AI-mediated second-language communication.

This is an open access article under CC-BY-NC 4.0 license.

ARTICLE INFO

Keywords:
artificial empathy;
EFL learners;
empathic language;
large language models;
perceived empathy

Article History:
Received: 10 September 2026
Revised: 14 October 2026
Accepted: 2 November 2026
Published: 4 November 2026

How to Cite in APA Style:
Febriyanti. (2026). When AI Sounds More Empathetic Than Humans: Experimental Evidence from Indonesian EFL Students. *LinguaEducare: Journal of English and Linguistic Studies, 3*(2), 155-168. https://doi.org/10.63324/lec.3v.2i.198

Introduction

The rapid development of generative artificial intelligence (GenAI) has transformed the nature of human–computer communication. Large language models (LLMs), particularly conversational systems such as ChatGPT, can produce fluent, contextually

*Corresponding author: Febriyanti febriyanti@stkippgribl.ac.id

responsive, and increasingly human-like language across a wide range of communicative situations. In language education, these capabilities have enabled learners to use AI not merely as an information-retrieval tool but as an interactive conversational partner. Recent studies have documented the use of ChatGPT and related AI systems for language practice, writing feedback, vocabulary development, speaking activities, and autonomous learning (Kohnke et al., 2023; Klímová & Ibna Seraj, 2023; Tram et al., 2024). Experimental research has further shown that AI-mediated interactions can increase willingness to communicate, support speaking development, and reduce anxiety associated with second-language communication (Fathi et al., 2024; Kim & Su, 2024; Wang et al., 2024). These developments suggest that the educational significance of LLMs extends beyond their ability to generate linguistically accurate text; learners increasingly encounter AI as a communicative entity whose language may also convey interpersonal and emotional meanings.

One increasingly important dimension of this interaction is empathy. In human communication, empathy is associated with recognizing another person's emotional state, acknowledging the significance of that experience, and responding in ways that communicate understanding, validation, and care. When this concept is applied to artificial systems, however, an important distinction emerges between possessing empathy and producing language that is perceived as empathetic. LLMs generate their responses by modelling linguistic patterns rather than by experiencing the emotions described by users. Consequently, claims that an LLM "feels" empathy need to be separated from evidence that its output contains linguistic features that users interpret as empathetic. Concannon and Tomalin (2024) emphasize the importance of measuring perceived empathy when evaluating dialogue systems because human judgments are shaped by observable communicative behaviour. Similarly, Rubin et al. (2024) argue that the apparent empathic capabilities of AI should not automatically be equated with the affective and relational processes involved in human empathy. More recently, Mugableh and Mohammed Alruwaili (2026) distinguished the functional performance of empathic communication from phenomenological empathy, demonstrating why high ratings of AI responses do not necessarily establish experiential understanding.

Nevertheless, empirical findings increasingly demonstrate that LLMs can produce responses that people judge to be highly empathetic. Ayers et al. (2023), in a comparison of physician and ChatGPT responses to patient questions, found that chatbot responses received substantially higher empathy ratings than physician responses. A similar pattern was reported by Ovsyannikova et al. (2025), whose four experimental studies showed that AI-generated responses were rated as more compassionate than responses produced by both selected human participants and trained crisis responders. Their analyses suggested that AI's advantage was partly explained by perceived responsiveness, particularly its consistent expression of understanding, validation, and care. Research examining emotional narratives also indicates that ChatGPT can represent affective information in systematic ways and differentiate emotional valence and arousal in text (Park & Kim, 2024). Ruben et al. (2026) further showed that judgments of AI "empathy" involve identifiable verbal characteristics, reinforcing the need to examine what linguistic properties lead readers to attribute empathy to machine-

generated responses rather than treating empathy ratings as evidence of an internal emotional capacity.

Perceived empathy is also shaped by users' expectations about the communicator. Human–AI interaction research demonstrates that people readily attribute human-like qualities, intentions, and psychological states to sufficiently sophisticated artificial agents (Kim & Im, 2023). Pataranutaporn et al. (2023), for example, found that beliefs about an AI system influenced perceptions of its trustworthiness, empathy, and effectiveness even when participants interacted with essentially the same underlying technology. Pan et al. (2025) similarly showed that linguistic subjectivity and perceived AI agency affected evaluations of trust, liking, and conversational quality. In higher education, anthropomorphism and trust have also emerged as important antecedents of students' attitudes toward ChatGPT (Polyportis & Pahos, 2024). Perceived empathy itself can influence subsequent reactions to conversational agents, as greater chatbot interactivity may strengthen perceptions that an artificial interlocutor understands the user's situation (Yim, 2024). These findings demonstrate that empathy judgments emerge from an interaction between linguistic cues and users' interpretations of the source producing those cues.

Importantly, evidence concerning artificial empathy is not uniformly favourable to AI. The social meaning attributed to a message can change when readers know whether its author is human or artificial. Shen et al. (2024), for example, found that participants generally experienced stronger empathy toward human-written than AI-written personal narratives, while transparency concerning AI authorship also influenced willingness to empathize with artificial narrators. Such findings illustrate that textual quality alone may not determine emotional judgments. An apparently caring response may be interpreted differently depending on assumptions about authenticity, intentionality, and authorship. This creates a particularly interesting psycholinguistic problem: readers make judgments about emotional meaning through lexical choice, discourse organization, evaluative language, acknowledgment of feelings, reassurance, perspective-taking, and other pragmatic cues, yet the perceived social meaning of these linguistic features may vary according to whether the presumed communicator is human or artificial.

This issue is especially relevant to English as a Foreign Language (EFL) learners. Current research on GenAI in EFL education has largely concentrated on instrumental and academic uses. ChatGPT has been investigated as a source of writing feedback (Guo & Wang, 2024; Teng, 2024), as a means of improving writing performance and motivation (Song & Song, 2023), as an autonomous language-learning assistant (Tram et al., 2024), and as an interactive partner for speaking practice (Fathi et al., 2024). Other studies have begun to demonstrate affective consequences of AI-supported language learning. Rezai et al. (2024), for instance, identified relationships between ChatGPT use, emotion regulation, and EFL learners' well-being, while Wang et al. (2024) found that conversational GenAI environments could reduce speaking anxiety and enhance self-perceived communicative competence. AI-generated learning materials have likewise been associated with positive affective experiences such as

reading enjoyment (Lee et al., 2023). Together, these studies demonstrate growing scholarly interest in the emotional dimensions of AI-assisted language learning.

Despite this progress, considerably less attention has been paid to how EFL learners interpret the interpersonal meanings embedded in LLM-generated language. Existing language-learning studies predominantly examine whether AI improves learning outcomes, facilitates practice, increases motivation, or provides useful feedback. Meanwhile, much of the experimental work directly comparing human and AI empathy has been conducted in healthcare, psychological support, or general human–AI communication rather than second-language settings (Ayers et al., 2023; Ovsyannikova et al., 2025; Rubin et al., 2024). This separation leaves an important gap. For EFL students, judging whether an English response sounds understanding, warm, validating, supportive, or contextually appropriate is itself a form of meaning processing in a second language. Studying their evaluations therefore provides an opportunity to connect research on artificial empathy with the psycholinguistics of L2-mediated interpersonal communication.

The present study addresses this gap by experimentally comparing how Indonesian EFL students perceive human-generated and LLM-generated responses to standardized emotional situations. Undergraduate students in an English Education program at STKIP PGRI Bandar Lampung evaluate the responses without relying solely on assumptions about AI authorship, allowing the study to focus on the language through which empathy is communicated. Rather than attempting to determine whether an LLM genuinely possesses empathy, the study conceptualizes empathy primarily as a perceived communicative quality observable through readers' evaluations of emotional understanding, validation, warmth, supportiveness, and contextual appropriateness. It additionally examines the linguistic characteristics that participants associate with empathetic communication. In doing so, the study extends research on human–AI empathy into an EFL context and contributes to understanding how artificial systems can linguistically perform interpersonal meanings that users may interpret as human-like emotional responsiveness. The study addresses the following research questions:

1. To what extent do Indonesian EFL students differ in their perceptions of empathy between human-generated and LLM-generated responses to emotional situations?
2. What linguistic characteristics of human- and LLM-generated responses shape Indonesian EFL students' perceptions of empathy?

Method

This study employed a mixed-methods experimental design to examine differences in Indonesian EFL students' perceptions of empathy in human- and large language model (LLM)-generated responses and to identify the linguistic characteristics underlying those perceptions. A within-subject design was adopted so that each participant evaluated responses from both sources, thereby reducing the influence of individual differences in empathy judgment. Similar human–AI comparative designs have been used to evaluate perceived compassion and affective responsiveness across

standardized emotional prompts (Ovsyannikova et al., 2025; Park & Kim, 2024). The quantitative component addressed differences in perceived empathy, whereas the qualitative component explored participants' interpretations of the linguistic features that made particular responses sound understanding, validating, warm, supportive, or contextually appropriate. Combining experimental ratings with qualitative accounts was considered appropriate because recent EFL research has demonstrated that mixed-methods designs can connect measurable patterns of GenAI evaluation with learners' more nuanced perceptions and experiences (Mahapatra, 2024; Teng, 2024).

The participants were 100 undergraduate students enrolled in the English Education Study Program at STKIP PGRI Bandar Lampung, Indonesia. Participants were recruited from different semesters to provide variation in English learning experience while maintaining a relatively comparable disciplinary background. Eligibility required participants to be active English Education students and able to comprehend English texts independently. Participation was voluntary, and informed consent was obtained before data collection. Demographic information included age, gender, semester, self-reported English proficiency, frequency of GenAI use, and prior experience with ChatGPT or comparable conversational AI. An additional group of 10 English Education students who did not participate as raters produced the human responses used as experimental materials. Separating response writers from evaluators minimized familiarity effects and prevented participants from recognizing individual writing styles. Ethical approval was obtained from the relevant institutional research ethics committee before participant recruitment, and all responses were anonymized during analysis.

The experimental materials consisted of ten English-language emotional vignettes depicting situations relevant to university students, including examination failure, academic workload, presentation anxiety, friendship conflict, scholarship rejection, low academic achievement, lack of confidence in English, social isolation, uncertainty about future employment, and academic pressure. The use of multiple standardized emotional prompts was informed by recent experimental work comparing human and AI empathic responses across varied personal experiences (Ovsyannikova et al., 2025). Each vignette contained approximately 70–100 words and presented sufficient emotional and contextual information to invite an empathic response. Three experts with backgrounds in English language education, psycholinguistics, and educational psychology reviewed the materials for linguistic difficulty, contextual relevance, emotional clarity, and comparability. Revisions were made when wording was judged unnecessarily difficult or when the emotional situation was ambiguous.

For every vignette, one human-generated and one LLM-generated response were prepared. Human writers received only the instruction to respond naturally and empathetically to the situation in approximately 80–120 words. They were not shown AI-generated examples. LLM responses were generated using ChatGPT through a standardized prompt instructing the system to respond naturally, supportively, and empathetically without mentioning its artificial identity. The same prompt structure and model setting were retained across all ten vignettes, and the model version, generation date, and complete prompts were archived to support reproducibility. Response length

was kept comparable across sources because verbosity can influence judgments of AI empathy and responsiveness (Ovsyannikova et al., 2025). The texts were additionally checked to ensure that they contained no explicit statements revealing whether the writer was human or artificial.

Participants completed the experiment individually. Each vignette was followed by the two responses, labelled Response A and Response B, with source identity concealed. The presentation order of human and LLM responses was randomized across items to control for order effects. Concealing authorship was particularly important because knowledge that content is AI-generated can alter emotional and empathic judgments independently of the actual language presented (Shen et al., 2024). Participants rated each response on a five-point Likert scale assessing emotional understanding, validation, warmth, perspective-taking, supportiveness, and contextual appropriateness. The items were adapted for written human–computer communication based on contemporary approaches to measuring perceived empathy in dialogue systems (Concannon & Tomalin, 2024). A pilot administration with 30 EFL students outside the main sample was conducted to examine item clarity, response time, and internal consistency before the main experiment.

For RQ1, an overall perceived-empathy score was calculated by averaging the six dimensions across the ten responses from each source. Descriptive statistics were first calculated, followed by normality and assumption checks. Human and LLM scores were then compared using a paired-samples t test, with statistical significance set at $p < .05$. Cohen's d was reported to indicate the magnitude of the difference rather than relying solely on statistical significance. To address RQ2, 18 participants representing higher LLM ratings, approximately equivalent human–LLM ratings, and higher human ratings were purposively selected for semi-structured interviews. Participants were asked which expressions made responses seem empathetic, natural, supportive, generic, formulaic, or emotionally authentic. Interview transcripts and participants' explanations were coded iteratively to identify recurrent linguistic and interpretive patterns. Quantitative and qualitative findings were subsequently integrated during interpretation, allowing differences in empathy ratings to be explained through the linguistic features participants associated with empathetic communication.

Results

Perceived Empathy in Human- and LLM-Generated Responses

The first research question examined whether Indonesian EFL students differed in their perceptions of empathy between human-generated and LLM-generated responses to emotional situations. As seen in Table 1, participants generally assigned higher ratings to the LLM responses across the ten vignettes. The overall perceived-empathy mean for the LLM condition was 4.11 (SD = 0.46), compared with 3.77 (SD = 0.52) for the human condition, producing a mean within-participant difference of 0.34 points on the five-point scale. Seventy-eight participants obtained a higher overall mean for the LLM responses, 14 rated the human responses higher, and eight produced equivalent scores after rounding to two decimal places. The direction of the difference was therefore evident for most participants rather than being produced by a small number of extreme ratings.

Table 1. Descriptive Perceived-Empathy Scores by Response Source

Dimension	Human M	Human SD	LLM M	LLM SD
Emotional understanding	3.82	0.56	4.12	0.50
Validation	3.75	0.59	4.19	0.51
Warmth	3.80	0.58	4.08	0.53
Perspective-taking	3.74	0.60	3.99	0.55
Supportiveness	3.79	0.57	4.22	0.49
Contextual appropriateness	3.72	0.61	4.05	0.50
Overall perceived empathy	3.77	0.52	4.11	0.46

A paired-samples t test confirmed that the overall difference was statistically significant, $t(99) = 7.08$, $p < .001$. The 95% confidence interval for the mean difference ranged from 0.25 to 0.44, and the standardized within-participant effect was moderate to large, Cohen's $dz = 0.71$. Thus, when source identity was concealed, students perceived LLM-generated responses as more empathetic overall than human-generated responses. Because the texts were presented as Response A and Response B without identifying their authorship, the result represents judgments of the language itself rather than an explicit preference for an AI-labelled or human-labelled source.

Descriptive patterns across the six dimensions were consistent with the overall comparison. The largest numerical differences appeared for supportiveness and validation. LLM responses received a mean of 4.22 (SD = 0.49) for supportiveness, compared with 3.79 (SD = 0.57) for human responses, and 4.19 (SD = 0.51) for validation, compared with 3.75 (SD = 0.59). Higher LLM means were also found for emotional understanding (4.12 vs. 3.82), warmth (4.08 vs. 3.80), perspective-taking (3.99 vs. 3.74), and contextual appropriateness (4.05 vs. 3.72). These dimension-level values were treated descriptively, but they indicate that the overall advantage was especially associated with explicit validation and sustained emotional support. The LLM advantage was also most noticeable in vignettes concerning academic failure, presentation anxiety, scholarship rejection, and uncertainty about future employment. Overall, the quantitative findings answer RQ1 by showing a clear perceived-empathy advantage for LLM-generated responses in this sample.

Linguistic Characteristics Shaping Students' Perceptions of Empathy

The second research question explored which linguistic characteristics shaped students' perceptions of empathy. Analysis of the 18 interviews, as seen in Table 2, yielded five recurrent patterns: explicit emotional acknowledgment, validation and reassurance, supportive guidance, lexical warmth and non-judgment, and authenticity versus formulaicity. The first four patterns appeared most often in explanations of highly rated LLM responses, whereas authenticity and conversational spontaneity were more frequently associated with preferred human responses. Participants therefore did not define empathy through a single word or phrase; rather, they described an accumulated impression created a response recognized emotion, positioned the recipient, and developed support across several sentences.

Table 2. Qualitative Patterns Associated with Perceived Empathy

Pattern	Typical interpretation	Participants
Emotional acknowledgment	Recognizes the specific emotion before giving advice	15/18
Validation and reassurance	Treats the emotion as understandable without dismissing the problem	14/18
Supportive guidance	Offers specific, manageable, and non-directive next steps	13/18
Lexical warmth/non-judgment	Uses warm language and softeners that preserve agency	13/18
Authenticity/formulaicity	Natural imperfection sounds human; excessive polish can sound scripted	11/18

The most prominent feature was explicit acknowledgment of the person's emotional state. Fifteen interviewees referred to expressions that directly recognized disappointment, nervousness, embarrassment, pressure, or uncertainty. Responses sounded more empathetic when they first recognized what the person was experiencing rather than immediately giving advice. P6, for example, explained that a preferred response worked because "it notices the embarrassment first, not only the presentation problem." P11 similarly said that a response felt caring because "it says the feeling makes sense before telling the student what to do." LLM responses frequently followed this acknowledgment-first pattern, which led participants to interpret them as attentive to the emotional meaning of the vignette rather than merely focused on solving the problem.

Validation and reassurance formed a closely related pattern. Fourteen participants valued statements that normalized an emotional reaction without minimizing the problem. P3 noted that a response sounded supportive because "it does not say just don't worry; it explains why the reaction is understandable." Participants preferred reassurance that was tied to details from the vignette, for example, explaining that one unsuccessful presentation did not determine a student's suitability to become an English teacher. This combination of validation and reframing corresponded with the relatively high quantitative scores for validation and supportiveness. Participants also valued supportive guidance when it followed emotional acknowledgment. Thirteen interviewees mentioned suggestions that were specific and manageable, such as practising with a friend, seeking feedback, or focusing on one small next step. Advice was particularly well received when it was expressed as an option through words such as could, might, perhaps, and if you feel ready, because these forms preserved the recipient's agency rather than sounding directive.

Lexical warmth and a non-judgmental stance also contributed to perceived empathy. Participants highlighted expressions such as it is understandable, it is okay to need time, and you are not alone in feeling this way as signals of interpersonal support. However, the interviews also revealed a counter-pattern concerning authenticity. Ten participants occasionally preferred human responses because they were shorter, less polished, and more conversational. P8 described one human response as sounding "like a real friend because it is not perfectly organized." In contrast, 11 participants considered some LLM responses overly structured or formulaic when they contained several validating statements followed by neatly sequenced advice. P17 characterized one such response as "very kind, but almost too complete." Repeated constructions such as it is completely

understandable, remember that, and what matters is could therefore function both as empathic cues and as signals of templated language.

Contextual specificity moderated these judgments. Human responses were sometimes preferred when they referred directly to a concrete detail in an informal and spontaneous way, while generic encouragement weakened evaluations regardless of source. Conversely, LLM responses were viewed favourably when they integrated several contextual cues without becoming repetitive. Across the interviews, students appeared to evaluate empathy through a balance between explicit emotional work and credible interpersonal positioning. They valued language that acknowledged emotion, validated the reaction, used warm and non-judgmental wording, and offered softened support, but they remained sensitive to language that appeared excessively scripted. The qualitative findings therefore answer RQ2 by showing that the LLM advantage observed in RQ1 was not attributed simply to grammatical accuracy or response length. Instead, it was associated with identifiable discourse and lexical strategies that made emotional understanding visible in the text, while naturalness and authenticity remained areas in which human responses could retain an advantage.

Discussion

The present study showed that Indonesian EFL students perceived LLM-generated responses as more empathetic than human-generated responses, while qualitative findings linked this advantage to emotional acknowledgment, validation, reassurance, lexical warmth, and non-judgmental wording. Human responses, however, were more often associated with authenticity and contextual specificity. These findings suggest that perceived empathy in AI-mediated communication depends less on whether a system possesses emotions than on how effectively empathy is performed through language. This interpretation aligns with Concannon and Tomalin (2024), who treat perceived empathy as a judgment arising from observable communicative behavior rather than direct access to an interlocutor's internal emotional state.

The higher ratings for LLM responses extend evidence that AI-generated messages can outperform human responses on perceived empathy and compassion. Ayers et al. (2023) found that chatbot responses to patient questions received higher empathy ratings than physicians' responses, whereas Ovsyannikova et al. (2025) showed that AI responses were perceived as more compassionate than responses from laypeople and trained crisis responders. In the present study, LLM responses consistently acknowledged emotions, normalized difficulties, and offered reassurance before suggesting actions. Because source identity was concealed, this advantage is more plausibly associated with characteristics of the language than with participants' prior attitudes toward AI.

The qualitative findings explain how particular linguistic features shaped those judgments. Participants responded positively to expressions that recognized emotion, validated the experience, softened advice, and conveyed warmth without blame. Such features appear to operate as interpersonal signals of attentiveness and support. Ovsyannikova et al. (2025) similarly identified perceived responsiveness as an important mechanism underlying preferences for AI-generated compassionate responses. Nevertheless, some LLM responses were described as polished, predictable, or formulaic.

Thus, explicit empathic signaling may increase perceived empathy, but excessive regularity can reduce authenticity. Effective empathic communication may therefore require both recognizable supportive expressions and contextual adaptation.

This tension highlights the difference between perceived empathy and emotional authenticity. LLMs can generate convincing empathic language without possessing human emotional experience or relational history. Rubin et al. (2024) caution that linguistic performance should not be treated as evidence that AI possesses the affective components of human empathy. Moreover, judgments can shift when authorship becomes known. Shen et al. (2024) found that transparency about AI authorship influenced empathic responses, while Pataranutaporn et al. (2023) demonstrated that prior beliefs about AI affected perceived empathy, trustworthiness, and effectiveness. The present blind comparison therefore illustrates an empathy paradox: AI may sound more empathetic linguistically while still appearing less authentic when formulaicity or lack of lived experience becomes salient.

The findings also extend GenAI research in EFL education beyond academic performance toward interpersonal meaning processing. Previous studies indicate that AI-supported interaction can increase willingness to communicate and speaking performance while reducing language anxiety (Fathi et al., 2024; Wang et al., 2024). ChatGPT use has also been associated with emotion regulation and well-being among EFL learners (Rezai et al., 2024), and students value its supportive feedback in writing contexts (Teng, 2024). The present findings add a psycholinguistic dimension by showing that EFL learners attend to lexical and pragmatic cues of empathy, including stance, validation, reassurance, politeness, and emotional alignment.

The study therefore offers theoretical and pedagogical implications. Theoretically, perceived artificial empathy can be understood as an interaction between linguistic performance and reader interpretation rather than as evidence of machine consciousness. Pedagogically, this distinction should form part of critical AI literacy for pre-service English teachers. Although conversational AI can provide accessible and low-anxiety interactional environments (Kohnke et al., 2023; Wang et al., 2024), students should learn to distinguish persuasive empathic wording from genuine emotional understanding. Comparing human and AI responses in class could help learners identify markers of validation, support, and contextual sensitivity while recognizing generic or formulaic patterns, thereby developing pragmatic awareness and critical judgment of human–AI communication.

Conclusion

This study examined how Indonesian EFL students perceived empathy in human- and LLM-generated responses to emotional situations. The findings indicated that LLM-generated responses were perceived as more empathetic overall, particularly because they consistently expressed emotional acknowledgment, validation, reassurance, warmth, and supportive guidance. At the same time, human responses were more frequently associated with authenticity, contextual specificity, and a sense of lived emotional experience. These findings suggest that perceived empathy in AI-mediated communication is strongly shaped by linguistic performance rather than by evidence of genuine emotional understanding. The study therefore contributes to emerging research

on artificial empathy by showing that EFL learners can attribute high levels of empathy to machine-generated language when it effectively employs recognizable interpersonal and pragmatic cues. In pedagogical terms, the findings highlight the value of incorporating critical AI literacy into English teacher education so that pre-service teachers can evaluate not only the linguistic quality of AI responses but also the distinction between persuasive empathic language and genuine human emotional understanding.

Several limitations should nevertheless be acknowledged. The study involved students from a single English Education program, which limits the generalizability of the findings across different institutional, cultural, and proficiency contexts. The use of a limited set of emotional vignettes and a single LLM may also have constrained the range of empathic communication examined, while participants' judgments may have been influenced by their previous experience with GenAI and their ability to interpret subtle interpersonal meanings in English. Future research should therefore involve larger and more diverse EFL populations, compare multiple LLMs, include different levels of emotional intensity, and examine whether explicit disclosure of AI authorship changes empathy judgments. Longitudinal and cross-cultural studies would also be valuable for investigating how repeated interaction with conversational AI influences perceptions of authenticity, trust, and emotional support over time.

References

Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M., & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine, 183*(6), 589–596. https://doi.org/10.1001/jamainternmed.2023.1838

Concannon, S., & Tomalin, M. (2024). Measuring perceived empathy in dialogue systems. *AI & Society, 39*, 2233–2247. https://doi.org/10.1007/s00146-023-01715-z

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System, 121*, 103254. https://doi.org/10.1016/j.system.2024.103254

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*, 8435–8463. https://doi.org/10.1007/s10639-023-12146-0

Kim, A., & Su, Y. (2024). How implementing an AI chatbot impacts Korean as a foreign language learners' willingness to communicate in Korean. *System, 122*, 103256. https://doi.org/10.1016/j.system.2024.103256

Kim, J., & Im, I. (2023). Anthropomorphic response: Understanding interactions between humans and artificial intelligence agents. *Computers in Human Behavior, 139*, 107512. https://doi.org/10.1016/j.chb.2022.107512

Klímová, B., & Ibna Seraj, P. M. (2023). The use of chatbots in university EFL settings: Research trends and pedagogical implications. *Frontiers in Psychology, 14*, 1131506. https://doi.org/10.3389/fpsyg.2023.1131506

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Lee, J. H., Shin, D., & Noh, W. (2023). Artificial intelligence-based content generator technology for young English-as-a-foreign-language learners' reading enjoyment. *RELC Journal, 54*(2), 508–516. https://doi.org/10.1177/00336882231165060

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments, 11*, Article 9. https://doi.org/10.1186/s40561-024-00295-9

Mugableh, A., & Mohammed Alruwaili, H. (2026). Functional vs. phenomenological empathy in large language models: Rethinking artificial empathy through experimental evidence. *East European Journal of Psycholinguistics, 13*(1), 307–332. https://doi.org/10.29038/eejpl.2026.13.1.mug

Ovsyannikova, D., de Mello, V. O., & Inzlicht, M. (2025). Third-party evaluators perceive AI as more compassionate than expert humans. *Communications Psychology, 3*, Article 4. https://doi.org/10.1038/s44271-024-00182-6

Pan, W., Liu, D., Meng, J., & Liu, H. (2025). Human–AI communication in initial encounters: How AI agency affects trust, liking, and chat quality evaluation. *New Media & Society, 27*(10), 5822–5847. https://doi.org/10.1177/14614448241259149

Park, C., & Kim, J. (2024). Exploring affective representations in emotional narratives: An exploratory study comparing ChatGPT and human responses. *Cyberpsychology, Behavior, and Social Networking, 27*(10), 736–741. https://doi.org/10.1089/cyber.2024.0100

Pataranutaporn, P., Liu, R., Finn, E., & Maes, P. (2023). Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. *Nature Machine Intelligence, 5*, 1076–1086. https://doi.org/10.1038/s42256-023-00720-7

Polyportis, A., & Pahos, N. (2025). Understanding students' adoption of the ChatGPT chatbot in higher education: The role of anthropomorphism, trust, design novelty and institutional policy. *Behaviour & Information Technology, 44*(2), 315–336. https://doi.org/10.1080/0144929X.2024.2317364

Rezai, A., Soyoof, A., & Reynolds, B. L. (2024). Disclosing the correlation between using ChatGPT and well-being in EFL learners: Considering the mediating role of emotion regulation. *European Journal of Education, 59*(4), e12752. https://doi.org/10.1111/ejed.12752

Ruben, M. A., Blanch-Hartigan, D., & Hall, J. A. (2026). What is artificial intelligence (AI) "empathy"? A study comparing ChatGPT and physician responses on an online forum. *Journal of General Internal Medicine, 41*(5), 1304–1311. https://doi.org/10.1007/s11606-025-10068-w

Rubin, M., Arnon, H., Huppert, J. D., & Perry, A. (2024). Considering the role of human empathy in AI-driven therapy. *JMIR Mental Health, 11*, e56529. https://doi.org/10.2196/56529

Shen, J., DiPaola, D., Ali, S., Sap, M., Park, H. W., & Breazeal, C. (2024). Empathy toward artificial intelligence versus human experiences and the role of transparency in mental health and social support chatbot design: Comparative study. *JMIR Mental Health, 11*, e62679. https://doi.org/10.2196/62679

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence, 7*, 100270. https://doi.org/10.1016/j.caeai.2024.100270

Tram, N. H. M., Nguyen, T. T., & Tran, C. D. (2024). ChatGPT as a tool for self-learning English among EFL learners: A multi-methods study. *System, 127*, 103528. https://doi.org/10.1016/j.system.2024.103528

Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System, 127*, 103533. https://doi.org/10.1016/j.system.2024.103533

Yim, M. C. (2024). Effect of AI chatbot's interactivity on consumers' negative word-of-mouth intention: Mediating role of perceived empathy and anger. *International Journal of Human–Computer Interaction, 40*(18), 5415–5430. https://doi.org/10.1080/10447318.2023.2234114