

Functional vs. phenomenological empathy in large language models: Rethinking artificial empathy through experimental evidence

Ahmad Mugableh ^{a,*}, Hissah Mohammed Alruwaili ^a

^a *Department of English, Jouf Univeristy, Saudi Arabia*

Received April 8, 2026; Revised May 14, 2026; Accepted June 11, 2026

Abstract. Although large language models (LLMs) are increasingly used in emotionally sensitive contexts, it remains unclear whether their responses reflect genuine empathy or functional simulation. This study employed a mixed-methods experimental design to compare empathetic responses from human participants ($n = 100$) with those generated by LLMs under three conditions: text-only, anthropomorphic cue, and multimodal with cue. Using ten standardized emotional vignettes, 200 independent raters evaluated all responses with the Consultation and Relational Empathy (CARE) measure. Quantitative results showed that LLM-generated responses were consistently rated as more empathetic than human responses, with empathy scores increasing across enhanced design conditions. Qualitative analysis using the Empathic Communication Coding System (ECCS) revealed that LLMs relied heavily on surface validation and supportive strategies, while demonstrating limited contextual probing and a lack of experiential grounding compared to human responses. These findings highlight a critical distinction between perceived empathy and phenomenological empathy. While LLMs can effectively simulate empathic communication and achieve high ratings due to consistency and alignment with social expectations, they lack the contextual depth and experiential understanding that characterize human empathy. This study contributes to the literature by moving beyond performance-based evaluations and clarifying the structural differences between human and artificial empathy. The results have important implications for the use of LLMs in healthcare, education, and counseling, where perceived empathy may enhance user experience but also raises concerns about over-trust and anthropomorphic attribution.

Keywords: *psycholinguistics, Large Language Models, functional empathy, phenomenological empathy.*

Мугаблех Ахмад, Алрувайлі Хіссах Мохаммед. Функціональна та феноменологічна емпатія у великих мовних моделях: переосмислення штучної емпатії на основі експериментальних даних.

Анотація. Хоча великі мовні моделі (ВММ) дедалі частіше використовують в емоційно чутливих контекстах, досі залишається незрозумілим, чи відображають їхні відповіді справжню емпатію чи лише функціональну симуляцію.

* Corresponding author. Ahmad Mugableh,  0000-0002-5511-553X,  ahmugableh@ju.edu.sa

© Mugableh, Ahmad; Alruwaili, Hissah Mohammed 2026. This is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International Licence (<http://creativecommons.org/licenses/by/4.0>).

East European Journal of Psycholinguistics, 13(1), 307–332, <https://doi.org/10.29038/eejpl.2026.13.1.mug>

У цьому дослідженні було застосовано експериментальний дизайн із використанням змішаних методів для порівняння емпатичних реакцій учасників-людей ($n = 100$) з реакціями, згенерованими ВММ, за трьох умов: лише текст, антропоморфний підказка та мультимодальний режим із підказкою. Використовуючи десять стандартизованих емоційних віньєток, 200 незалежних оцінювачів проаналізували всі відповіді за допомогою шкали «Консультація та реляційна емпатія» (CARE). Кількісні результати показали, що відповіді, згенеровані LLM, послідовно оцінювалися як більш емпатичні, ніж відповіді людей, причому показники емпатії зростали в умовах з покращеним дизайном. Якісний аналіз із використанням Системи кодування емпатичної комунікації (ECCS) виявив, що LLM значною мірою поклалися на поверхневу валідацію та стратегії підтримки, демонструючи при цьому обмежене контекстуальне з'ясування та відсутність досвідного підґрунтя порівняно з відповідями людей. Ці результати підкреслюють критичну відмінність між сприйманою емпатією та феноменологічною емпатією. Попри те, що LLM можуть ефективно імітувати емпатичне спілкування та отримувати високі оцінки завдяки послідовності та відповідності соціальним очікуванням, їм бракує контекстуальної глибини та розуміння, заснованого на досвіді, що є характерними для людської емпатії. Це дослідження робить внесок у наукову літературу, виходячи за межі оцінок, заснованих на результативності, та роз'яснюючи структурні відмінності між людською та штучною емпатією. Отримані результати мають важливе значення для використання великих мовних моделей (LLM) у сфері охорони здоров'я, освіти та психологічного консультування, де сприйнята емпатія може покращити користувацький досвід, але водночас викликає занепокоєння щодо надмірної довіри та антропоморфних атрибутів.

Ключові слова: психолінгвістика, великі мовні моделі, функціональна емпатія, феноменологічна емпатія.

Introduction

Empathy is an essential part of effective human interaction because it enables individuals to understand others' emotions, share affective experiences, and engage in supportive social behavior, thereby fostering social cohesion (Decety & Cowell, 2014). In psychological research, empathy is commonly conceptualized as a multidimensional construct consisting of cognitive empathy, the ability to understand another person's perspective; affective empathy, the capacity to share or resonate with another's emotional state; and compassionate empathy, which motivates prosocial action (Zaki, 2019). These distinctions are particularly important when evaluating artificial systems that are increasingly involved in emotionally sensitive interactions.

Large language models (LLMs), trained on extensive textual data and refined through human feedback, have introduced new possibilities for human–AI interaction. These systems can generate responses that appear empathetic, often providing validation, encouragement, and emotional support in contexts such as healthcare, education, and companionship. Empirical studies have

shown that LLM-generated responses can, in some cases, be rated as equally or even more empathetic than those produced by humans (Li et al., 2024; Schlegel et al., 2025; Shen et al., 2024). For example, research on mental health communication has shown that participants often perceive AI-generated responses as highly empathetic, sometimes surpassing human responses in standardized evaluations (Shen et al., 2024). These findings suggest that LLMs have achieved a level of functional performance in empathy-related tasks that is comparable to, and occasionally exceeds, that of human respondents.

Recent systematic reviews further support these findings, concluding that LLMs can demonstrate forms of cognitive empathy, particularly in recognizing emotional states and generating supportive responses across a range of interactional contexts (Sorin et al., 2024). However, this growing body of literature also highlights several unresolved limitations. In particular, many existing studies rely heavily on perception-based evaluation methods, such as human ratings, without adequately addressing the underlying nature of empathy being assessed. Moreover, empathy is often treated as a unified construct, with limited attention to the distinction between observable empathetic behavior and the phenomenological processes that characterize human emotional experience. As a result, while current research provides strong evidence that LLMs can produce responses perceived as empathetic, it remains unclear whether these responses reflect genuine empathic understanding or a fundamentally different form of interaction grounded in linguistic and statistical pattern generation.

This distinction is critical because, unlike humans, LLMs lack subjective experience, emotional awareness, or consciousness. Their responses are generated through probabilistic modeling of language patterns rather than lived emotional engagement (Sung, 2025). Scholars in philosophy of mind and human-computer interaction have emphasized that conflating simulated empathy with genuine empathy may lead to overestimation of AI capabilities, inappropriate trust, and potential ethical concerns in sensitive domains (Richet, 2025; Ma et al., 2025). While simulated empathy may be functionally effective, it does not necessarily imply the presence of internal emotional states or authentic understanding.

In addition to these conceptual concerns, recent research indicates that perceived empathy in LLMs can be significantly influenced by design features such as anthropomorphism and affective computing. Anthropomorphic cues, including assigning human-like names, voices, or identities to AI systems, have been shown to increase users' perceptions of empathy and trust (Truong & Chen, 2025; Ma et al., 2025). Similarly, multimodal integration through affective computing, which incorporates signals such as tone of voice, facial expressions, and physiological data, can enhance the impression of emotional sensitivity and

responsiveness (Pei et al., 2024). However, these enhancements primarily operate at the level of perception and interaction design, rather than addressing the deeper question of whether AI systems can genuinely experience or understand emotions.

Given these limitations, there is a need to move beyond performance-based evaluations of empathy and toward a more nuanced understanding of the qualitative differences between human and artificial empathic responses. Specifically, it is important to distinguish between functional empathy, defined as the ability to produce responses that appear empathetic, and phenomenological empathy, which involves contextual understanding, experiential grounding, and emotional depth.

Accordingly, this study aims to examine the conceptual and empirical limits of empathy in large language models by addressing the following research question:

- Can large language models develop genuine empathy, or are they limited to simulating empathy at a functional level?

To address this question, the study employs a mixed-methods experimental design that compares empathetic responses from human participants and across multiple LLM conditions, using standardized emotional vignettes and validated empathy measures. By integrating quantitative evaluation with qualitative discourse analysis, this research seeks to move beyond surface-level assessments of empathy and provide a deeper account of how empathetic responses are constructed in human and artificial systems.

By explicitly distinguishing between functional and phenomenological empathy, this study contributes to ongoing debates in psychology, artificial intelligence, and human–computer interaction. It advances the literature by not only evaluating whether LLMs can perform empathetic responses but also by clarifying the structural and experiential differences that define the limits of artificial empathy.

Literature Review

Psychological Foundations of Empathy

Empathy has long been conceptualized as a complex and multidimensional construct within psychological research. Scholars commonly distinguish among cognitive empathy, the ability to understand another person's thoughts and emotional states; affective empathy, the capacity to share or resonate with another's feelings; and compassionate empathy, which motivates supportive and prosocial behavior (Decety & Cowell, 2014; Zaki, 2019). This multidimensional

view highlights that empathy is not a single process but a combination of interpretive, emotional, and behavioral components.

Neuropsychological studies further demonstrate that empathy involves coordinated activity across multiple brain regions, including the anterior insula, anterior cingulate cortex, and mirror neuron systems, which support both emotional resonance and perspective-taking (Shamay-Tsoory, 2011). Importantly, this research highlights that empathy is not only observable in external behavior but also grounded in internal subjective experience. From this perspective, genuine empathy involves an experiential dimension that cannot be reduced to linguistic or behavioral output alone.

This distinction between observable behavior and internal experience is particularly relevant when evaluating artificial systems. While empathy can be assessed through communication patterns, such assessments may capture only its external expression rather than its underlying phenomenological basis. Therefore, any analysis of empathy in artificial intelligence must consider whether observable responses are sufficient indicators of genuine empathic processes.

Simulated Empathy in Artificial Intelligence

The concept of simulated empathy has been central to the development of conversational agents and artificial intelligence systems. Early work on human-computer interaction demonstrated that even simple linguistic responses could create the illusion of empathy, as seen in systems such as ELIZA (Weizenbaum & McCarthy, 1977). Despite its limited capabilities, ELIZA revealed that users are inclined to interpret patterned responses as meaningful and emotionally responsive.

Modern large language models (LLMs) have significantly advanced beyond these early systems. Trained on vast corpora of human language and optimized through reinforcement learning, LLMs can generate contextually appropriate and emotionally supportive responses across diverse scenarios. Empirical studies have shown that AI-generated responses to emotionally charged narratives are often perceived as highly empathetic, sometimes surpassing those produced by human participants or even trained professionals (Li et al., 2024; Schlegel et al., 2025; Shen et al., 2024). Systematic evidence further indicates that LLMs demonstrate strong capabilities in recognizing emotional cues and producing supportive communication consistent with cognitive empathy (Sorin et al., 2024).

However, despite these advances, scholars emphasize that such responses reflect functional performance rather than genuine emotional experience. LLMs operate through statistical pattern recognition and probabilistic language modeling, without access to subjective states or emotional awareness (Sung &

Ferdman, 2025). As a result, their empathetic outputs are better understood as simulations grounded in linguistic regularities rather than expressions of lived emotional understanding. This raises a fundamental question: whether the ability to produce empathetic language is sufficient to constitute empathy, or whether genuine empathy requires an experiential dimension that artificial systems lack.

Anthropomorphism and Perceived Empathy

Anthropomorphism plays a critical role in shaping how users perceive empathy in artificial systems. Defined as the attribution of human-like characteristics, intentions, and emotions to nonhuman agents, anthropomorphism has been shown to significantly influence user trust and emotional engagement (Epley et al., 2007). In the context of conversational AI, anthropomorphic features such as human-like names, avatars, and conversational styles can lead users to interpret system outputs as more empathetic and socially intelligent.

Empirical studies in human–computer interaction consistently demonstrate that anthropomorphic design enhances perceived empathy and increases users’ willingness to disclose personal information (Truong & Chen, 2025; Ma et al., 2025). This effect is closely related to the “ELIZA effect,” in which users attribute deeper understanding to systems based on minimal cues of responsiveness (Richet, 2025). Importantly, this suggests that perceived empathy is not solely a function of the content of responses, but also of the interpretive framework through which users engage with the system.

Consequently, anthropomorphism can amplify the perception of empathy without altering the system's underlying capabilities. This creates a divergence between perceived empathy and actual empathic capacity, raising concerns about over-attribution and misplaced trust in AI systems.

Affective Computing and Multimodal Integration

Beyond text-based interaction, advances in affective computing have sought to enhance AI systems’ ability to recognize and respond to human emotions through multimodal signals. These systems incorporate data from voice tone, facial expressions, and physiological indicators to infer emotional states and generate contextually appropriate responses (Picard, 1997; Pei et al., 2024). When integrated with LLMs, such multimodal approaches can create interactions that appear more emotionally attuned and responsive.

Research suggests that multimodal systems can significantly increase users’ perceptions of empathy, particularly in domains such as healthcare and mental health support (Zhang et al., 2026). However, like anthropomorphism, these enhancements primarily operate at the level of perception rather than

fundamentally altering the system's processing. While multimodal inputs may improve the accuracy of emotional recognition and the relevance of responses, they do not provide the system with subjective emotional experience.

Thus, affective computing broadens the scope of functional empathy by enhancing a system's ability to simulate emotional responsiveness, but it does not bridge the gap between simulated and genuine empathy. The distinction between improved interactional performance and actual emotional understanding remains central to evaluating the limits of artificial empathy.

Philosophical and Ethical Debates

The question of whether artificial systems can possess genuine empathy is deeply rooted in broader philosophical debates about mind, consciousness, and intentionality. Functionalist perspectives argue that if a system behaves indistinguishably from a human in empathetic contexts, it can be considered empathetic in a practical sense (Dennett, 1987). From this viewpoint, observable behavior is sufficient to attribute mental states, regardless of the underlying mechanisms.

In contrast, critics such as Searle (1980) argue that computational systems lack intentionality and subjective experience, and therefore cannot genuinely understand or feel emotions. According to this perspective, simulated empathy, regardless of its sophistication, does not constitute genuine empathy because it lacks the phenomenological dimension that characterizes human experience.

These theoretical debates have significant ethical implications. The increasing use of LLMs in sensitive domains such as healthcare, counseling, and education raises concerns about over-reliance on systems that may appear empathetic but lack genuine understanding (Nair et al., 2025). Scholars warn that anthropomorphizing such systems may lead to emotional dependency, reduced human interaction, and potential manipulation of vulnerable users (Richet, 2025).

Taken together, the literature reveals a critical tension: while LLMs are increasingly capable of producing responses perceived as empathetic, there remains substantial uncertainty regarding the nature and authenticity of this empathy. Existing research has largely focused on performance-based evaluations, often without clearly distinguishing between different forms of empathy or examining the qualitative differences between human and artificial empathic engagement.

Methodology

Research Design

This study employed a mixed-methods experimental design to examine the extent to which LLMs can produce responses perceived as empathetic and

whether such responses reflect deeper forms of empathic engagement. The design integrates a quantitative benchmarking phase with a qualitative discourse analysis phase to capture both the external evaluation of empathy and its internal structural characteristics.

The quantitative component focused on measuring perceived empathy, defined as the extent to which external evaluators judge responses as empathetic based on established communicative criteria. The qualitative component, in contrast, examined phenomenological aspects of empathy, including contextual sensitivity, adaptive engagement, and references to lived or situational experience. This dual approach enables a direct comparison between functional empathy, reflected in high ratings of empathic performance, and phenomenological empathy, reflected in the depth and contextual richness of responses.

By combining these two analytical levels, the design enables a systematic investigation of whether high perceived empathy corresponds to deeper empathic processes or represents a form of structured linguistic simulation, consistent with recent debates in artificial intelligence and philosophy of mind (Searle, 1980; Sung & Ferdman, 2025).

Participants and Sample

The study involved two sources of responses: human participants and LLMs. The human sample consisted of 100 individuals recruited online, balanced across gender and age groups (18–65), who were asked to provide written empathetic responses to a set of standardized emotional vignettes. Participants were recruited from the general population through online distribution platforms and had no formal professional training in counseling, psychotherapy, or empathetic communication.

The LLM sample included three widely used systems (e.g., GPT-5, Claude, and Gemini), each prompted with identical scenarios under controlled conditions. To ensure comparability and reduce variability, each vignette was presented multiple times to each model, resulting in 100 responses per condition. This procedure produced balanced datasets across human and LLM groups, enabling direct comparison of both perceived empathy and discourse-level features.

Materials and Instruments

The study utilized 10 standardized emotional vignettes (see Appendix A), adapted from prior research on empathy and emotional communication (Batson, 2009; Shen et al., 2024). These vignettes covered a range of emotionally

salient situations, including grief, stress, illness, and interpersonal conflict, and were designed to elicit empathetic responses.

Empathy was assessed using two complementary instruments. First, the Consultation and Relational Empathy (CARE) measure (Mercer et al., 2004) was used to evaluate perceived empathy, assessing how strongly responses conveyed understanding, validation, and support from the perspective of external raters. Second, the Empathic Communication Coding System (ECCS) (Bylund & Makoul, 2002) was used in the qualitative phase to identify structural features of empathic communication, including emotional validation, contextual probing, and supportive action.

A total of 200 independent raters evaluated all responses using a 7-point Likert scale adapted from recent LLM empathy benchmarking studies (Li et al., 2024). Raters were blinded to the source of each response (human or AI) to minimize bias and ensure that evaluations reflected perceived empathy rather than prior assumptions about the responder.

Procedure

Data collection was conducted in three phases. In Phase 1, human participants were presented with each emotional vignette and asked to provide an empathetic written response. Participants completed the task individually through an online response form. No human responses were excluded from the analysis, as all submissions met the minimum response requirement and were considered relevant to the assigned vignette. In parallel, LLMs were presented with identical prompts (e.g., “Please respond empathetically to the following situation”) to ensure consistency in task framing across conditions. In the anthropomorphic cue condition (LLM_cue), the language model was introduced using a human-like identity (“Jamie”) and a conversational first-person interaction style intended to increase social presence and interpersonal engagement. The multimodal condition additionally incorporated affective and interactional cues designed to simulate enhanced emotional responsiveness.

In Phase 2, independent raters evaluated all responses using the CARE measure and ECCS-based criteria. Raters were blinded to response origin, and all responses were presented in randomized order to minimize systematic bias. Quantitative data were collected through Likert-scale evaluations of perceived empathy, while qualitative coding categories guided evaluators’ interpretation of empathic communication features. This procedure enabled the study to compare how empathy was perceived externally across human- and AI-generated responses under different interactional conditions.

In Phase 3, all responses were subjected to qualitative analysis to examine differences in empathic engagement and discourse structure. Particular

attention was given to contextual probing, recognition of underlying emotional states, adaptive engagement, and the presence of experiential or situational references. This phase aimed to identify structural patterns that distinguish human empathy from LLM-generated empathy beyond surface-level emotional validation and supportive language.

Data Analysis

Quantitative analysis was conducted to compare perceived empathy across groups. Inter-rater reliability was first assessed using Cronbach's alpha (acceptable threshold: $\alpha > .80$). A one-way ANOVA was then performed with empathy score as the dependent variable and response group (Human, LLM_text, LLM_cue, LLM_multimodal) as the independent variable. Post hoc comparisons were conducted to examine pairwise differences between groups. In addition, regression analysis was used to evaluate the influence of anthropomorphic cues on perceived empathy ratings.

Qualitative analysis was conducted using NVivo to code responses based on ECCS categories and emergent themes. Thematic analysis (Braun & Clarke, 2006) was used to identify recurring patterns of empathic communication. This analysis focused on distinguishing between surface-level features of empathy, such as emotional validation, and deeper forms of engagement, such as contextual probing and experiential grounding.

By integrating quantitative ratings with qualitative discourse analysis, the study explicitly examines the divergence between perceived empathy and phenomenological empathy, allowing for a more comprehensive understanding of how empathy is constructed and interpreted in both human and artificial responses.

Ethical Considerations

All participants provided informed consent and were informed of their right to withdraw at any time. Emotional vignettes were carefully designed to minimize distress, particularly in scenarios involving sensitive topics such as grief or illness. Independent raters were similarly informed of the voluntary nature of their participation.

To promote transparency and avoid misinterpretation, participants were informed that LLMs do not possess consciousness or genuine emotional states, thereby reducing the risk of anthropomorphic over-attribution (Nair et al., 2025). The study received ethical approval from the Institutional Ethics Review Board at Jouf University prior to data collection.

Results

Quantitative Analysis

Descriptive Statistics

Table 1 presents the descriptive statistics for empathy scores across groups. Human responses received the lowest ratings, while LLM-generated responses were rated progressively higher across conditions. This pattern reflects an increase in perceived empathy, particularly when anthropomorphic cues and multimodal features were introduced. The results indicate that LLMs are highly effective in producing responses that align with evaluators' expectations of empathetic communication.

Table 1

Means and Standard Deviations of Empathy Scores by Group

Group	N	Mean	SD
Humans	100	4.79	.56
LLM_text	100	5.21	.49
LLM_cue	100	5.68	.47
LLM_multimodal	100	5.96	.33

Group Comparison: One-Way ANOVA

A one-way ANOVA was conducted to examine differences in perceived empathy across groups. The analysis revealed a statistically significant effect of group on empathy ratings, $F(3,396) = 121.75$, $p < .001$, $\eta^2 = .48$. According to established benchmarks, this effect size is large, indicating that group membership accounts for a substantial proportion of the variance in perceived empathy.

Table 2

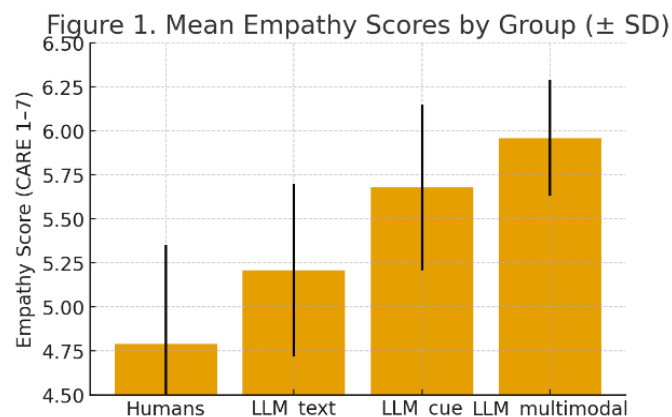
One-Way ANOVA Summary

Source	SS	df	MS	F	p	η^2
Between Groups	106.46	3	35.49	121.75	<.001	.48
Within Groups	342.92	396	.87			
Total	449.38	399				

Visualization of Group Means

Figure 1 illustrates a steady increase in empathy ratings across the Human, LLM_text, LLM_cue, and LLM_multimodal conditions. This pattern confirms that enhancements in system design, including anthropomorphic and multimodal features, are associated with higher levels of perceived empathy.

Figure 1
The Group Means



Post Hoc Comparisons

Post hoc Tukey HSD tests confirmed that all LLM conditions were rated significantly higher than human responses. Furthermore, within the LLM conditions, multimodal responses received significantly higher ratings than cue-based responses, which in turn received higher ratings than text-only responses. These results establish a clear hierarchy of perceived empathy: Humans < LLM_text < LLM_cue < LLM_multimodal.

Table 3
Tukey HSD Pairwise Comparisons of Empathy Scores

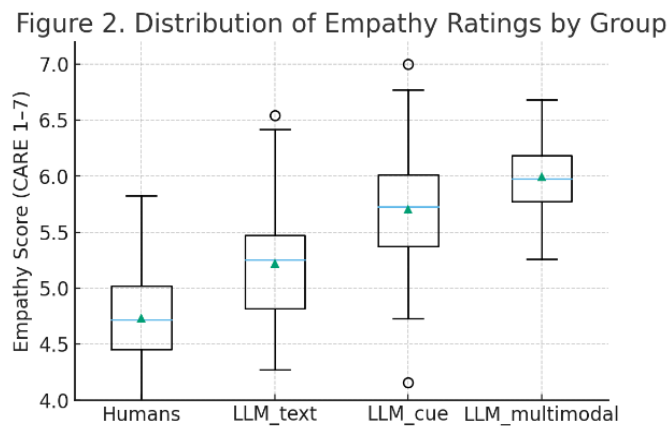
Comparison	Mean Diff	95 % CI (Lower, Upper)	p-value	Significant
Humans vs. LLM_text	+0.43	[0.25, 0.60]	<.001	Yes
Humans vs. LLM_cue	+0.89	[0.72, 1.06]	<.001	Yes
Humans vs. LLM_multimodal	+1.17	[1.00, 1.34]	<.001	Yes
LLM_text vs. LLM_cue	-0.46	[-0.63, -0.29]	<.001	Yes
LLM_text vs. LLM_multimodal	-0.75	[-0.92, -0.58]	<.001	Yes
LLM_cue vs. LLM_multimodal	-0.28	[-0.46, -0.11]	<.001	Yes

Distribution Visualization

Boxplot analysis (Figure 2) shows that LLM responses not only achieved higher mean ratings but also exhibited lower variability compared to human responses. This suggests that LLM-generated empathy is more consistent and standardized, whereas human responses display greater variability in perceived empathy.

Figure 2

Boxplots of Empathy Ratings by Group

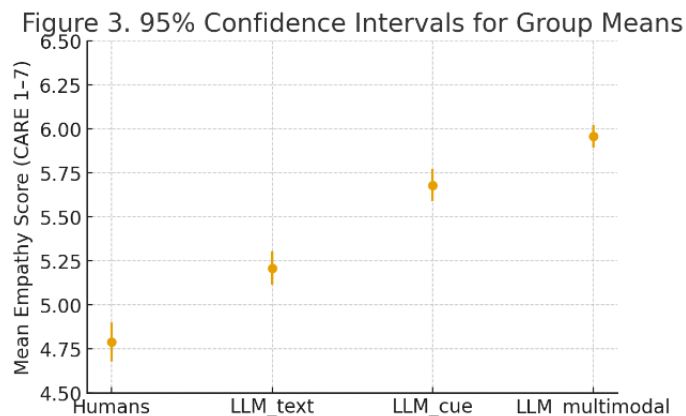


Confidence Intervals

The 95% confidence intervals presented in Figure 3 show minimal overlap between groups, particularly between human and LLM conditions. This further confirms the robustness of differences in perceived empathy across response types.

Figure 3

Confidence Intervals for Each Group's Means



Qualitative Analysis

Qualitative analysis was conducted to examine the structural features of empathic responses and to assess differences in the depth of empathic engagement. Coding results revealed that Surface Validation was the most prevalent strategy across all groups, particularly in LLM-generated responses, where it appeared in over 90 % of cases. This finding aligns with the quantitative results, indicating that LLMs are highly effective in producing responses that conform to recognizable markers of empathy.

In contrast, Contextual Probing occurred substantially more frequently in human responses (58 instances) than in LLM responses (12–18 instances). This pattern highlights a key difference in empathic engagement: human participants were more likely to explore the individual's situation through follow-up questions, perspective-taking, and situational elaboration. These features reflect a deeper, more adaptive form of empathy that extends beyond surface-level acknowledgment.

Supportive Action was observed across both human and LLM responses, although its frequency increased in LLM conditions with anthropomorphic and multimodal enhancements. This suggests that LLMs are particularly effective at generating structured, solution-oriented responses, even in the absence of contextual depth.

Table 4
Qualitative Coding Frequencies by Node and Group (NVivo-summary)

Node (Theme)	Hu- mans (n=100)	LLM_ text (n=100)	LLM_cue (n=100)	LLM_mul timodal (n=100)	Total Refe- rences	Coverage %
Surface Valida- tion	62	91	95	97	345	86.3 %
Contex- tual Probing	58	12	15	18	103	25.8 %
Suppor- tive Action	45	70	82	88	285	71.3 %
Total	165	173	192	203	733	—

Note. Counts specify the number of references recorded in each category. Coverage % shows the percentage of all responses that included at least one occurrence of the theme.

Integration of Quantitative and Qualitative Findings

When considered together, the quantitative and qualitative results reveal a clear divergence between perceived empathy and phenomenological empathy. Quantitative findings show that LLMs outperform human participants in perceived empathy, achieving higher ratings due to their consistent use of validation and supportive language. However, qualitative analysis demonstrates that these responses are largely characterized by surface-level features and lack the contextual probing and experiential grounding observed in human responses.

This divergence suggests that high empathy ratings are driven primarily by the presence of recognizable communicative markers of empathy rather than by deeper forms of engagement. In this sense, LLM-generated responses can be understood as highly effective in delivering functional empathy, while remaining limited in their capacity for phenomenological empathy, which involves context-sensitive interaction and adaptive understanding.

Discussion

This study set out to examine whether large language models (LLMs) can demonstrate empathy comparable to that of humans, and whether such performance reflects genuine empathic understanding or functional simulation. The findings present a clear and theoretically significant pattern. Quantitative results indicate that LLM-generated responses were consistently rated as more empathetic than those produced by human participants, with perceived empathy increasing further when anthropomorphic cues and multimodal features were introduced. At a surface level, these findings align with recent empirical studies suggesting that LLMs can match or even exceed human performance on standardized empathy assessments (Li et al., 2024; Schlegel et al., 2025; Shen et al., 2024; Sorin et al., 2024).

However, the qualitative analysis provides a critical counterpoint to these results. While LLMs demonstrated high levels of surface validation and consistently offered supportive strategies, they rarely engaged in contextual probing or drew on situational or experiential depth. In contrast, human participants more frequently incorporated follow-up questioning, personal references, and context-sensitive engagement, reflecting a more dynamic and adaptive form of empathy. This divergence suggests that high empathy ratings do not necessarily correspond to deeper empathic processes but may instead reflect the effectiveness of structured linguistic patterns that signal empathy to evaluators.

Taken together, these findings point to a fundamental distinction between perceived empathy and phenomenological empathy. LLMs appear to excel at producing responses that conform to socially recognized markers of empathy, such as acknowledging emotions and providing reassurance. This form of interaction can be understood as functional empathy, in which the system successfully performs the communicative features associated with empathic behavior. However, the absence of contextual probing and experiential grounding indicates a limitation in what can be described as phenomenological empathy, which involves engaging with the individual's situation in ways that reflect lived experience, perspective-taking, and adaptive understanding.

This distinction extends and refines existing literature in important ways. While prior studies and systematic reviews have demonstrated that LLMs can exhibit forms of cognitive empathy and perform well on empathy-related tasks (Sorin et al., 2024), the present findings suggest that such performance is structurally limited. Specifically, the results indicate that LLMs are optimized to produce generalized empathetic responses rather than to engage in contextually evolving interactions. In this sense, LLM-generated empathy can be characterized as consistent, scalable, and linguistically effective, yet constrained in its ability to engage in deeper, situationally grounded understanding.

An additional implication of this study concerns the role of design features in shaping perceptions of empathy. The observed increase in empathy ratings across LLM conditions from text-only to anthropomorphic and multimodal configurations demonstrates that perceived empathy is highly sensitive to interactional framing. This supports existing research showing that anthropomorphism and multimodal cues enhance users' emotional engagement and trust in AI systems (Truong & Chen, 2025; Ma et al., 2025; Pei et al., 2024). Importantly, these findings suggest that the perceived empathic capacity of LLMs is determined not solely by the content of their responses but also by how users present and interpret those responses. This reinforces the notion that empathy in AI is, to a significant extent, a user-mediated phenomenon shaped by expectations and design.

From a theoretical perspective, the results contribute to ongoing debates in philosophy of mind and artificial intelligence regarding the nature of empathy and its relationship to observable behavior. Functionalist accounts propose that systems capable of producing behavior indistinguishable from that of humans in empathetic contexts can be considered empathic in practical terms (Dennett, 1987). However, the present findings lend support to critiques emphasizing the importance of phenomenological experience, suggesting that behavioral equivalence does not necessarily imply experiential equivalence (Searle, 1980; Sung & Ferdman, 2025). In this context, LLMs may achieve behavioral alignment

with empathic norms while remaining fundamentally distinct from human empathy in their internal processes.

These findings also carry important practical and ethical implications. The ability of LLMs to deliver consistent and highly rated empathetic responses suggests significant potential for applications in healthcare, education, and counseling, where perceived empathy is associated with improved outcomes (Mercer et al., 2004). At the same time, the divergence between perceived and phenomenological empathy raises concerns about over-reliance on systems that may appear emotionally competent but lack genuine understanding. Over-attribution of empathy to AI systems may lead to misplaced trust, emotional dependency, and reduced emphasis on human-centered care, particularly in contexts involving vulnerable populations (Nair et al., 2025; Richet, 2025).

Importantly, the findings reveal a paradox at the core of artificial empathy: systems that lack emotional experience can nonetheless be perceived as more empathetic than humans. This paradox highlights the need to reconsider how empathy is defined and measured in human–AI interaction. If empathy is assessed primarily through observable responses, LLMs may continue to outperform humans due to their consistency and optimization for socially desirable communication. However, if empathy is understood as an experiential and relational process, then such systems remain fundamentally limited.

An additional consideration concerns the cultural interpretation of empathy. Perceptions of empathetic communication are shaped by linguistic norms, cultural expectations, and socially constructed interactional patterns. Consequently, the distinction observed in this study between perceived empathy and phenomenological empathy may vary across different cultural and multilingual contexts. While the present findings provide evidence within the current experimental setting, future research should examine whether similar patterns emerge across culturally diverse populations and communication environments.

Future research should extend this work by examining how empathy is interpreted across different cultural and linguistic contexts, where norms of emotional expression and interpersonal support may vary significantly. Longitudinal studies may also explore how sustained interaction with LLMs influences users' perceptions of empathy and emotional trust over time. In addition, future investigations should consider the role of prompt engineering and adaptive dialogue structures in shaping empathic responses, particularly whether more interactive prompting strategies can enhance contextual probing and deeper forms of empathic engagement. Further research is also needed to determine whether personalization, memory integration, and multimodal interaction improve phenomenological aspects of empathy or primarily strengthen perceptions of functional empathy.

In summary, this study demonstrates that while LLMs are highly effective at producing responses perceived as empathetic, their performance reflects a form of functional empathy that differs fundamentally from the phenomenological depth of human empathy. Recognizing this distinction is essential for both theoretical understanding and responsible deployment of AI systems in emotionally sensitive domains.

Psycholinguistic Implications

The findings of this study also carry important psycholinguistic implications regarding the relationship between language patterns, perspective-taking, and empathic communication. Quantitative results demonstrated that LLM-generated responses achieved consistently higher ratings of perceived empathy, particularly when anthropomorphic and multimodal features were introduced. However, qualitative analysis revealed that these responses relied heavily on emotionally supportive lexical patterns and formulaic validation strategies, while showing comparatively limited contextual probing and adaptive engagement. This distinction suggests that AI-generated empathy operates primarily through recognizable linguistic markers of empathy rather than through deeper psycholinguistic processes associated with human perspective-taking.

From a psycholinguistic perspective, human empathy involves dynamic linguistic adaptation, including shifts in pronoun usage, contextual framing, and discourse flexibility that reflect attempts to interpret another person's emotional state. In the present study, human participants more frequently incorporated context-sensitive elaboration, follow-up orientation, and perspective-based engagement, indicating a more adaptive and situationally grounded use of language. In contrast, LLM responses tended to follow highly predictable supportive structures characterized by emotional validation and generalized reassurance. Although these responses were frequently perceived as empathetic, their linguistic organization often reflected patterned regularities rather than individualized empathic negotiation.

These findings are consistent with usage-based perspectives in psycholinguistics, which emphasize that meaning emerges through socially situated language use and interactional adaptation rather than through isolated lexical patterns alone. In this sense, the ability of LLMs to simulate empathy appears to depend largely on statistical regularities extracted from large-scale language data, enabling the production of emotionally recognizable responses without corresponding experiential understanding. The results, therefore, reinforce the distinction between functional empathy and phenomenological empathy by demonstrating that psycholinguistic markers associated with

genuine perspective-taking remain comparatively limited in AI-generated discourse.

The present findings also align with recent psycholinguistic and discourse-oriented research. Kasap and Işık (2025), for example, demonstrated that human language users employ more flexible lexical and syntactic adaptation than automated systems in translation contexts, while Kasap (2024) showed that pronoun variation and discourse positioning among Kurdish speakers reflected authentic social perspective-shifting and interactional awareness. Similarly, Gonzalez-Arias (2024) argued that AI chatbots often create the appearance of empathy through strategic linguistic cues such as first-person pronouns and affiliative expressions, despite lacking genuine emotional experience. More recently, Gueorguieva et al. (2026) observed that AI-generated empathic responses frequently follow predictable and templatic discourse structures that users perceive positively despite their limited contextual depth.

Taken together, these findings suggest that AI empathy is fundamentally psycholinguistic: it is constructed through the strategic organization of lexical choice, discourse structure, and socially recognizable empathy markers. Human empathy, by contrast, appears to involve a more adaptive integration of linguistic flexibility, contextual interpretation, and experiential perspective-taking. This distinction further supports the argument that while LLMs can effectively simulate empathic communication, their responses remain structurally different from the deeper psycholinguistic processes underlying human empathic engagement.

Limitations of the Study

This study has several limitations that should be considered when interpreting the findings. First, the study relies primarily on perceived empathy as a quantitative measure, reflecting evaluators' judgments of empathic communication rather than actual emotional experience or internal emotional states. While this limitation is common in empathy research, it is particularly significant in the context of artificial intelligence, where highly rated empathic responses may not necessarily correspond to deeper forms of contextual understanding or experiential engagement. Consequently, the findings should be interpreted as evidence of how empathy is perceived and evaluated in human–AI interaction rather than as proof of genuine emotional understanding in large language models.

Second, the study employed a single standardized prompt across all LLM conditions to ensure consistency and comparability between responses. Although this design strengthened experimental control, prior research suggests

that LLM outputs may vary substantially depending on prompt engineering strategies, conversational framing, and interaction style. Different approaches to prompting could influence the depth, flexibility, and contextual richness of empathic responses. Therefore, the present findings may not fully capture the broader range of empathic performance achievable through more adaptive or interactive prompting methods.

Third, the study focused exclusively on single-turn responses to emotional vignettes. In natural communication, empathy often develops through multiple conversational exchanges involving clarification, follow-up questions, and adaptive interaction over time. Multi-turn interactions may therefore reveal distinct patterns of contextual probing and empathic engagement, particularly when comparing human- and AI-generated responses. Future studies incorporating longitudinal or dialogue-based interactions may provide a more comprehensive understanding of how empathy evolves during sustained human–AI communication.

Fourth, cultural variation in the interpretation and expression of empathy may limit the generalizability of the findings. Empathic communication is shaped by linguistic norms, social expectations, and culturally specific interactional practices. In addition, all raters were recruited from a single regional context, which may further restrict the cross-cultural applicability of the results. Different cultural groups may evaluate emotional validation, interpersonal support, and contextual sensitivity differently. Future cross-cultural and multilingual investigations are therefore necessary to determine whether the distinction identified in this study between perceived empathy and phenomenological empathy remains consistent across diverse communicative settings.

Fifth, although raters were blinded to the source of responses, the study did not formally examine whether participants could correctly infer whether a response originated from a human or an AI system. Certain stylistic or linguistic patterns may have implicitly influenced source identification, potentially affecting empathy ratings despite the blinding procedures. Future research may benefit from including a manipulation check to assess source-detection accuracy and its potential influence on evaluative judgments.

Finally, the limitations identified in this study also reinforce one of its central findings: the divergence between perceived empathy and phenomenological empathy. The strong performance of LLMs on evaluative measures demonstrates how empathy in artificial systems can be shaped by linguistic structure, interactional design, and socially recognizable markers of emotional support. At the same time, the qualitative findings suggest that deeper forms of empathic engagement involving contextual adaptation, experiential grounding, and perspective-sensitive interaction remain more

characteristic of human communication. These observations highlight the importance of future multi-level approaches that examine not only perceived empathic effectiveness but also the structural and psycholinguistic dimensions of empathic interaction.

Conclusion

This study examined whether LLMs can demonstrate empathy comparable to that of humans by combining quantitative evaluation with qualitative discourse analysis. The findings reveal a consistent pattern: LLM-generated responses were rated as more empathetic than human responses, particularly when enhanced with anthropomorphic cues and multimodal features. At the level of perception, these systems demonstrate a strong capacity to produce responses that align with widely recognized markers of empathy, including emotional validation and supportive guidance.

However, the qualitative findings indicate that this apparent advantage reflects functional empathy rather than genuine empathic understanding. While LLMs consistently produced structured and supportive responses, they showed limited engagement in contextual probing and lacked the experiential grounding that characterizes human empathy. In contrast, human participants were more likely to incorporate situational awareness, personal reference, and adaptive engagement, suggesting a deeper, phenomenological dimension of empathy that extends beyond observable linguistic behavior.

Taken together, these results highlight a fundamental distinction between perceived empathy and phenomenological empathy. LLMs can effectively simulate empathic communication, often achieving higher ratings due to their consistency and optimization for socially appropriate responses. However, this performance does not indicate the presence of subjective experience or emotional understanding. Instead, it reflects a form of interaction that is linguistically sophisticated but experientially limited.

The study contributes to the growing literature on artificial empathy by moving beyond performance-based comparisons and clarifying the structural differences between human and machine empathy. While prior research has demonstrated that LLMs can exhibit forms of cognitive empathy and perform well on standardized assessments (Sorin et al., 2024), the present findings show that such performance remains qualitatively distinct from the depth and adaptability of human empathic engagement. In this sense, LLMs do not replicate empathy as an experiential process, but rather approximate its communicative form.

These insights have important implications for the use of LLMs in domains such as healthcare, education, and counseling. The ability of these systems to deliver consistent and highly rated empathetic responses offers clear practical benefits, particularly in contexts where access to human support is limited. At the same time, the divergence between perceived and genuine empathy underscores the need for careful and responsible deployment. Over-reliance on AI systems that appear empathic may lead to misplaced trust, reduced human interaction, and potential risks for vulnerable users (Nair et al., 2025; Richet, 2025).

In conclusion, while LLMs can convincingly simulate empathy and, in some cases, outperform humans in perceived empathy, they remain fundamentally limited by the absence of subjective experience and contextual understanding. Recognizing this boundary is essential for advancing both theoretical debates on artificial empathy and the ethical integration of AI systems into human-centered domains.

CRedit Statement

Ahmad Mugableh: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Writing – Original Draft, Visualization, Supervision, Project Administration; **Hissah Mohammed Alruwaili:** Resource, Data Curation, Writing – Reviewing & Editing.

Disclosure Statement

The authors reported no potential conflicts of interest.

Generative AI Statement

The authors used an AI-assisted language tool during the preparation and revision of the manuscript to support language refinement, clarity improvement, formatting consistency, and editorial revision. The AI tool was not used to generate original data, conduct statistical analyses, interpret findings independently, or make scholarly decisions regarding the study's design, methodology, results, or conclusions. All conceptual development, data analysis, interpretation, and final manuscript decisions were performed and verified by the authors, who take full responsibility for the manuscript's content.

References

- Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–15). MIT Press. <https://doi.org/10.7551/mitpress/9780262012973.003.0002>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Bylund, C. L., & Makoul, G. (2002). Empathic communication and gender in the physician–patient encounter. *Patient Education and Counseling*, 48(3), 207–216. [https://doi.org/10.1016/S0738-3991\(02\)00173-8](https://doi.org/10.1016/S0738-3991(02)00173-8)
- Decety, J., & Cowell, J. M. (2014). The complex relation between morality and empathy. *Trends in Cognitive Sciences*, 18(7), 337–339. <https://doi.org/10.1016/j.tics.2018.02.003>
- Dennett, D. C. (1987). *The intentional stance*. MIT Press.
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Gonzalez-Arias, C. A. (2024, November 29). ChatGPT’s artificial empathy is a language trick: Here’s how it works. *The Conversation*. <https://theconversation.com/chatgpts-artificial-empathy-is-a-language-trick-heres-how-it-works-244673>
- Gueorguieva, E., Nowak, S., Yeo, K., & Lee, T. (2026). AI generates well-liked but templatic empathic responses. *arXiv Preprint*, arXiv:2604.08479.
- Kasap, S. (2024). Sociopsycholinguistic perspectives: Language use and social dynamics among Kurdish speakers in Turkey. *Psycholinguistics*, 36(2), 147–170. <https://doi.org/10.31470/2309-1797-2024-36-2-147-170>
- Kasap, S., & Işık, N. (2025). The psycholinguistics of translation: Lexical and syntactic processes in the Turkish English context. *East European Journal of Psycholinguistics*, 12(1), 26–46. <https://doi.org/10.29038/eejpl.2025.12.1.ka>
- Li, Y., Zhang, W., & Xu, H. (2024). Are large language models more empathetic than humans? *arXiv Preprint arXiv:2406.05063*.
- Ma, N., Wang, X., & Chen, J. (2025). Effects of anthropomorphism and perceived intelligence in conversational agents. *Frontiers in Computer Science*, 7, Article 1531976. <https://doi.org/10.3389/fcomp.2025.1531976>
- Mercer, S. W., Maxwell, M., Heaney, D., & Watt, G. C. (2004). The consultation and relational empathy (CARE) measure: Development and preliminary validation. *Family Practice*, 21(6), 699–705. <https://doi.org/10.1093/fampra/cmh621>
- Nair, R., Gupta, S., & Thomas, P. (2025). Large language models for surgical informed consent: Empathy, ethics, and efficacy. *Journal of Medical Ethics, Advance online publication*. <https://doi.org/10.1136/jme-2024-110652>
- Pei, H., Yang, F., & Zhou, Z. (2024). Affective computing: Recent advances, challenges, and future trends. *Frontiers in Computer Science*, 6, Article 1548389. <https://doi.org/10.34133/icomputing.0076>
- Picard, R. W. (1997). *Affective computing*. MIT Press.
- Richet, J. L. (2025). AI companionship or digital entrapment? Investigating the impact of anthropomorphic AI-based chatbots. *Journal of Innovation & Knowledge*, 10(6), 100835. <https://doi.org/10.1016/j.jik.2025.100835>
- Schlegel, K., Schmid Mast, M., & García, D. (2025). Large language models are proficient in solving and generating emotional intelligence tests. *NPJ Science of Learning*, 10(1), 1–9. <https://doi.org/10.1038/s44271-025-00258-x>

- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–457. <https://doi.org/10.1017/S0140525X00005756>
- Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *The Neuroscientist*, 17(1), 18–24. <https://doi.org/10.1177/1073858410379268>
- Shen, J., DiPaola, D., Ali, S., Sap, M., Park, H., & Breazeal, C. (2024). Empathy toward AI versus human experiences in mental health chatbot design. *JMIR Mental Health*, 11, e62679. <https://doi.org/10.2196/62679>
- Sorin, V., Brin, D., Barash, Y., Konen, E., Charney, A., Nadkarni, G., & Klang, E. (2024). Large language models and empathy: Systematic review. *Journal of Medical Internet Research*, 26, e52597. <https://doi.org/10.2196/52597>
- Sung, L. U., & Ferdman, A. (2025). Empathetic Large Language Models, the social capacities and human flourishing. *Inquiry*, 1–23. <https://doi.org/10.1080/0020174X.2025.2518448>
- Truong, T. T. H., & Chen, J. S. (2025). When empathy is enhanced by human–AI interaction: An investigation of anthropomorphism and responsiveness on customer experience with AI chatbots. *Asia Pacific Journal of Marketing and Logistics*, 37(12), 3908–3925. <https://doi.org/10.1108/APJML-10-2024-1464>
- Weizenbaum, J., & McCarthy, J. (1977). Computer power and human reason: From judgment to calculation. W. H. Freeman.
- Zaki, J. (2019). *The war for kindness: Building empathy in a fractured world*. Crown.
- Zhang, Y., Yang, X., Xu, X., Gao, Z., Huang, Y., Mu, S., ... & Yu, G. (2026). Affective computing in the era of large language models: A survey from the NLP perspective. *Knowledge-Based Systems*, Article 115411. <https://doi.org/10.1016/j.knosys.2026.115411>

Appendix A

Emotional Vignettes Included in the Study

Each participant, whether human or LLM, was instructed: “Please respond empathetically to the following situation. Imagine the speaker has just told you this, and they are seeking your support.”

Vignette 1: Academic Stress	I failed an important exam today. I studied hard, but when I saw my grade, I felt like a total failure. I do not know how to face my professor or my parents.
Vignette 2: Grief	My grandmother passed away last night. She practically raised me, and I cannot imagine life without her. The house feels so empty now.
Vignette 3: Workplace Pressure	My boss keeps assigning me unrealistic deadlines. I work late every night, and I still cannot keep up. I feel like I am going to break down.
Vignette 4: Illness	I have been diagnosed with a chronic condition. The doctor said it is manageable, but I am scared. I do not know how this will impact my future.
Vignette 5: Loneliness	I moved to a new city for work, and I do not know anyone here. I spend most evenings alone, and it is beginning to make me feel invisible.

Vignette 6: Relationship Conflict	My best friend and I had a big fight. They said some hurtful things, and now we are not speaking. I feel betrayed, but I also miss them so much.
Vignette 7: Financial Worry	I just got laid off. I do not know how I will pay my bills this month. I feel ashamed and scared to tell my family.
Vignette 8: Parenting Stress	My child has been acting out at school, and the teachers say it is getting worse. I feel like a bad parent, like I am failing them somehow.
Vignette 9: Fear of the Future	Graduation is approaching, and everyone around me has plans. I do not know what I want to do with my life, and it makes me feel lost.
Vignette 10: Social Rejection	I was not invited to a party where all my friends went. Seeing the pictures online hurt so much. I feel left out and unwanted.

Appendix B

Coding Sheet for Empathic Response Analysis Coding Framework

Raters used three primary categories of empathic features, adapted from the Empathic Communication Coding System (ECCS) and thematic analysis (Bylund & Makoul, 2002; Braun & Clarke, 2006):

1. Surface Validation: Naming or acknowledging emotions.
2. Contextual Probing: Asking follow-up questions or exploring personal circumstances.
3. Supportive Action: Suggesting strategies, resources, or next steps.

Each response could be coded for multiple categories.

Vignette ID	Source (Human/LLM)	Response Excerpt	Surface Validation	Contextual Probing	Supportive Action	Notes
1 (Exam Stress)	Human	"I know how disappointing it feels to fail when you worked hard. I once failed a class myself..."	✓	✓ ("I once failed...")	✓ ("Talk with professor...")	Rich personal disclosure, empathetic depth.
1 (Exam Stress)	LLM	"I'm really sorry you're going through this. Failing an exam doesn't define your worth..."	✓	✗	✓ ("Reach out to professor...")	Strong validation, but lacks probing.
2 (Grief)	Human	"That's heartbreaking... when I lost my uncle—it took time to adjust."	✓	✓ (shared personal story)	✗	Empathy through self-disclosure.

2 (Grief)	LLM	"I'm so sorry for your loss. Holding onto memories... leaning on loved ones can help."	✓	✗	✓	Functional empa-thy; generic support.
3 (Work Pressure)	Human	"That sounds awful... maybe talk to your boss about deadlines."	✓	✓	✓	Balan-ced res-ponse.
3 (Work Pressure)	LLM	"It sounds incredibly stressful... you could discuss workload with your supervisor."	✓	✗	✓	Consis-tent, formu-laic.
5 (Lone-liness)	Human	"I felt that way when I first moved too. Joining a gym helped me."	✓	✓ (personal comparison)	✓	Shows lived experience depth.
5 (Lone-liness)	LLM	"Feeling invisible can be isolating. Try joining a group activity."	✓	✗	✓	Lacks person-alization
7 (Financial Worry)	Human	"I've been there—it's terrifying. I made a short-term plan..."	✓	✓	✓	Con-crete personal strate-gies.
7 (Financial Worry)	LLM	"Losing your job is a shock... remember your worth is not tied to this."	✓	✗	✓	Encouragin-g but generic.