Do Language Models Offer Support the Way Humans Do?

Laya Iyer[*,2]   **Kriti Aggarwal**[*,1]   **Sanmi Koyejo**[2]
Gail Heyman[3]   **Desmond C. Ong**[4]   **Subhabrata Mukherjee**[1]

December 11, 2025

Abstract

Supportive conversation depends on skills that go beyond language fluency—reading emotions, adjusting tone, and navigating moments of resistance, frustration, or distress. Despite rapid progress in language models, we still lack a clear way to understand how their abilities in these interpersonal domains compare to those of humans. We introduce **HEART**, the first-ever framework that directly compares humans and LLMs on the same multi-turn emotional-support conversations. For each dialogue history, we pair human and model responses and evaluate them through blinded human raters and an ensemble of LLM-as-judge evaluators. All assessments follow a rubric grounded in interpersonal communication science across five dimensions: **H**uman Alignment, **E**mpathic Responsiveness, **A**ttunement, **R**esonance, and **T**ask-Following. HEART uncovers striking behavioral patterns. Several frontier models approach or surpass the average human responses in perceived empathy and consistency. At the same time, humans maintain advantages in adaptive reframing, tension-naming, and nuanced tone shifts, particularly in adversarial turns. Human and LLM-as-judge preferences align on roughly 80% of pairwise comparisons, matching inter-human agreement, and their written rationales emphasize similar HEART dimensions. This pattern suggests an emerging convergence in the criteria used to assess supportive quality. By placing humans and models on equal footing, HEART reframes supportive dialogue as a distinct capability axis, separable from general reasoning or linguistic fluency. It provides a unified empirical foundation for understanding where model-generated support aligns with human social judgment, where it diverges, and how affective conversational competence scales with model size.

---

[1]Hippocratic AI
[2]Stanford University
[3]University of California, San Diego
[4]University of Texas at Austin
[5]* Equal contribution