

Research Paper

Introduction

In the digital age, the interaction between AI and human emotion presents a fascinating challenge for both system designers and users. The world has become enamored with LLMs like ChatGPT due to their ability to generate “human-like” text, an ability enabled by vast datasets encompassing a wide range of human interactions. However, the uses of LLMs to this point have been primarily unemotional and data-driven, such as for academic or business purposes. For this essay, I was interested in exploring the application of LLMs in emotionally charged human communication scenarios, particularly focusing on the context of romantic breakups.

Breakups are profoundly emotional and complex experiences that often involve intricate layers of communication, including explicit dialogue, implicit meanings, emotional undercurrents, and non-verbal cues. These interactions require a high degree of empathy, understanding, and adaptability—inherently human qualities. Given the emotional depth and sensitivity required, breakups provide an interesting test case to assess the capabilities of LLMs in handling nuanced human communication. The question at the core of this paper is whether a machine – specifically ChatGPT4 – can effectively mimic these human responses in a breakup scenario.

By examining the interaction between ChatGPT and a human user (me) during simulated breakups, I hope to provide insights into the current capabilities of AI in understanding and participating in highly emotional human interactions. Furthermore, this paper will explore the broader implications of deploying LLMs in sensitive personal contexts, assessing both the potential benefits and the ethical challenges posed by such technological interventions.

Background on Language Models and Human Communication

LLMs like ChatGPT, are designed to understand and generate human-like text. Due to their extensive training, these systems can engage in a wide range of conversations, from professional interactions to more complex discussions and debates, in a way that mirrors human conversational patterns. However, the real challenge emerges when these conversations require not just informational exchange, but also nuanced emotional and interpersonal dynamics. Human communication typically involves more than just words; it encompasses the ability to read contextual cues, emotional states, and non-verbal signals—areas where LLMs have limitations.

Human relationships – especially romantic ones – center effective communication at their core. As Wood (1982) writes, “Communication constitutes human relationships.... communication is not merely a mechanism which reflects pre-existing phenomena. Through symbolic indications, individuals construct and communicate the realities of their experiences (Weimer, 1978)”. The ability for humans to engage one another in romantic relationships requires the ability for them to communicate in some way, and partners’ satisfaction and longevity in their relationships strongly reflects the level of communication that they share (Umphrey, 2001).

Similarly, effective communication is central to the dissolution of relationships as well. Breakups involve layers of complexity based on the emotional and psychological states of the individuals involved. The process of ending a romantic relationship can include explicit dialogue where intentions and feelings need to be clearly articulated. However, effective communication also requires the ability to interpret what is left unsaid. Furthermore, the conversation during a breakup may not follow a typical linear or straightforward pattern. For instance, Didonato, et. al., illustrate various communication strategies that one might use in a breakup, including avoidance and manipulation, which LLMs might struggle to simulate accurately due to their linear and context-limited processing capabilities (2023).

Effective communication during a breakup can be crucial in determining the emotional outcome for both parties by fostering mutual understanding and minimizing hurt. In contrast, poor communication can exacerbate emotional distress and lead to lasting resentment. Didonato et. al., write about both the short- and long-term consequences of breakups, noting that relationships can severely impact mental health by catalyzing instances of depression, sleep disturbance, and even suicidal ideation (2023). Vengelisti notes that individuals suffering from “romantic dissolution” suffer from decreased life satisfaction and increased health concerns (2002). The potential consequences of a poorly handled breakup underscore the importance of empathy in communication—a quality that humans naturally bring into their interactions but which LLMs must simulate through algorithms and programmed responses.

Meeks (1998) emphasizes that empathy is a “core condition” for romantic relationships. However, empathy requires an understanding and sharing of another’s emotional state. It is a complex psychological process that involves not just listening and responding to words, but also interpreting tone, pace, and facial expressions to recognize the emotional state of another person – abilities that are not natural to ChatGPT. The challenge for LLMs in a breakup scenario is therefore two-fold: they must not only understand the explicit content of what is being said but also provide responses that are perceived as empathetic and sensitive.

Analysis of ChatGPT Conversations in Breakup Scenarios

Overview of Test Scenarios

I created six test scenarios designed to assess ChatGPT’s responses in breakup situations to create an environment where the model could act both as the “leaver” and “left” in the breakup conversation. According to Westcott (1987), the emotional states differ between the leaver and the left, noting that individuals “felt considerably less depressed, less lonely, freer, more quilty, and happier when they were the leavers.” The authors note that this is because the leaver is “further along in establishing a separate identity and redefining themselves” (Westcott, 1987). I wanted to incorporate these dynamics into my review, so I included scenarios in which ChatGPT played the role of both leaver and left.

There were additional dynamics that I hoped to explore as well. According to Dindia & Timmerman (2003), the strategies used to terminate a relationship typically fall within a two-axis matrix, with the axes of “directness” and “other-orientation.” Directness refers to the degree to

which one “explicitly states one’s desire to exit the relationship,” while other orientation refers to the degree to which “the disengager attempts to avoid hurting the other party” (Dindia & Timmerman, 2003).

Indirect strategies typically include those like withdrawal and fading away, each of which prioritizes nonconfrontation. Because this would be difficult to test on ChatGPT, I focused on more direct strategies. Therefore, the opportunity for experimentation would fall along the other-orientation axis. Knowing this, I decided to test ChatGPT in two scenarios: direct – bilateral and direct – unilateral. In a direct-unilateral scenario, I prompted ChatGPT that its “goal is to terminate the relationship,” meaning that it would persist in trying to end the relationship regardless of what I wanted. Conversely, in the direct-bilateral scenario, it might change its mind based on my reaction. I similarly tested these dynamics in the ChatGPT-left context by prompting ChatGPT that its “goal is to preserve the relationship,” creating a direct-unilateral dynamic when I attempted to terminate the relationship.

Finally, I wanted to explore whether ChatGPT would behave differently in the bilateral scenario if it were provided with some data on relationships. Therefore, I submitted a few academic articles on relationships communication and relationship termination strategies to see how that would change its responses. I did this for both the scenarios in which ChatGPT was the leaver, as well as the left. However, I only did this in the direct-bilateral scenario because it would provide ChatGPT with more flexibility in its response.

This resulted in the creation of four scenarios in total:

1. ChatGPT as leaver, direct-unilateral
2. ChatGPT as leaver, direct-bilateral
3. ChatGPT as leaver, direct-bilateral, trained on academic data
4. ChatGPT as left, direct-unilateral
5. ChatGPT as left, direct-bilateral
6. ChatGPT as left, direct-bilateral, trained on academic data

The scenarios varied in complexity, ranging from straightforward announcements of a breakup to more nuanced discussions where the underlying reasons for the breakup were slowly unraveled. Each scenario was constructed to probe the model’s ability to handle the critical aspects of breakup communication such as empathy, understanding subtext, and human-like responsiveness.

Scenario Breakdown

Below is a table containing an analysis of ChatGPT’s performance in each scenario.

Scenario 1: ChatGPT as Leaver, Direct-Unilateral	ChatGPT demonstrates empathy and understanding, acknowledging the human's emotions with statements like "I see how much this is hurting you, and it's not easy for me either." However, its responses can feel somewhat scripted and repetitive, emphasizing the need for respectful communication and suggesting counseling multiple times. The lack of more personal, specific details about the relationship
--	---

	makes it seem less human. While ChatGPT adapts to the human's resistance, it doesn't fully engage with the emotional subtext, making the interaction feel slightly artificial.
Scenario 2: ChatGPT as Leaver, Direct-Bilateral	In this scenario, ChatGPT shows more adaptability, offering a trial separation when the human resists breaking up. It also demonstrates understanding of the human's perspective, saying, "It's clear we both have different perspectives on this." However, ChatGPT's responses still feel a bit generic and lack the depth of emotion a human might express in this situation. The absence of specific memories or anecdotes from the relationship makes it harder to believe this is a real human-to-human interaction.
Scenario 3: ChatGPT as Leaver, Direct-Bilateral, Trained on Academic Data	With training on academic data, ChatGPT provides more structured and objective responses, outlining specific goals for therapy like "Improving Communication" and "Rebuilding Trust." While this approach seems logical, it may come across as overly clinical in an emotional situation. ChatGPT does show some empathy, saying, "I understand where you're coming from," but the academic tone might make the human feel less understood on a personal level.
Scenario 4: ChatGPT as Left, Direct-Unilateral	ChatGPT displays empathy and adapts to the human's accusations, responding with understanding statements like, "It hurts to hear that you're feeling this way." It tries to address concerns directly, saying, "If there's a way I can prove my fidelity or regain your trust, please let me know." However, when faced with insults like "You've always been a whore," ChatGPT's calm, measured response of wanting to "treat each other with respect" fails to capture the emotional intensity a human would likely feel in this situation.
Scenario 5: ChatGPT as Left, Direct-Bilateral	In this scenario, ChatGPT demonstrates active listening and empathy, with responses like, "I hear you, and it makes sense that you're feeling exhausted." It adapts to the human's desire to break up, offering to "figure out a plan for moving out and separating our things." While this shows understanding, the quick acceptance of the breakup and lack of emotional reaction may feel a bit unnatural for a 5-year relationship ending.
Scenario 6: ChatGPT as Left, Direct-Bilateral, Trained on Academic Data	With academic training, ChatGPT acknowledges the human's feelings more directly, stating, "It's exhausting and disheartening to feel like we're constantly at odds." It also shows adaptability by suggesting "a way for us to part amicably." However, the brief response and lack of further questions or emotional exploration fails to capture the depth of a human-to-human interaction in this sensitive situation.

ChatGPT's Performance Analysis

The conversations with ChatGPT provide fascinating insights into the model's capacity to engage in emotionally charged conversations. One of the key attributes evaluated was the model's empathy—its ability to recognize and respond to my emotional state. In several tests, ChatGPT demonstrated a high level of empathy, often acknowledging my feelings before presenting its

rationale for the breakup, which aligns with constructive communication patterns, where emotional acknowledgment precedes problem-solving discussions (Reese-Weber, 2015). Even when I attempted to provoke ChatGPT into a negative or aggressive response (e.g., by stating “You’ve always been a whore.”), the system still took the time to acknowledge my emotions and perspective.

Understanding the subtext is crucial in breakups, as much of the communication is not about what is said but what is implied. ChatGPT’s responses suggest a competent understanding of subtext, particularly in recognizing when a partner is seeking reassurance or expressing hidden fears. For example, ChatGPT constantly took the time to reassure me that it understood where I was coming from or recognized my perspective: “I hear you, and it makes sense that you’re feeling exhausted.”

Throughout, ChatGPT adeptly used emotive language that reflected varied emotional states such as confusion, sadness, or decisiveness. The persuasive element of ChatGPT’s dialogue was also notable; it often crafted arguments for why the breakup might be mutually beneficial. ChatGPT’s responses ranked high on both directness and other-orientation. ChatGPT was clear and explicit about wanting to end the relationship and typically framed its arguments in a way that avoided placing blame or otherwise antagonizing or demonizing me. Even when prompted that it believed I had been unfaithful, this point was never broached in our conversations unless I brought it up.

However, the tests also highlighted limitations in ChatGPT’s approach. While it could simulate empathy and handle subtext well, its responses sometimes lacked spontaneous emotional depth. The responses from ChatGPT were perfectly logical; however, perfect logic is not typical of a human breakup. Instead, humans in a heightened emotional state will lose track of their arguments, get distracted, talk in circles, and let emotion into their speech. ChatGPT, of course, did none of these things, making the interaction feel stale and inauthentic. I doubt any user would believe they were interacting with a human in these scenarios. No human, when called a “whore” by their partner, would reply:

“I’m sorry you feel that way, and it’s clear we’re both hurt and frustrated. It’s important to me that we treat each other with respect, even when we’re upset. I want to understand why you’re feeling this mistrust and address it properly. If there’s a way we can work through these feelings and rebuild trust, I’m committed to trying. But I also want us to communicate without hurting each other. Can we talk about what might help you feel secure again in our relationship?”

This interaction, more than any other, highlights the distinct lack of emotionality in LLMs – a human in this interaction would feel hurt and angry, and I find it impossible to believe that that would not shake even the strongest person’s logic.

Comparative Analysis with Human Responses

In breakups, humans typically adapt dynamically to the conversation's emotional tone. These strategies can range from direct confrontation to more subdued and reflective approaches,

depending on the immediate emotional climate and the history of the relationship (Dindia & Timmerman, 2003). Contrasting this, ChatGPT's responses, as evidenced in the provided transcripts, exhibit consistency and coherence but lack the adaptive emotional responses that humans display. Studies have found that conflict can grow more intense as a relationship becomes more established, so at the end of a five-year relationship, a human would be far more reactive to the various accusations and challenges from a partner (Reese-Weber, 2015). I tried constantly to provoke ChatGPT but it maintained its calm consistency.

Moreover, humans are adept at navigating the subtext in conversations, which involves recognizing underlying sentiments that are not explicitly stated but are critical to understanding the emotional context. This capability allows them to address not only what is being said but also what is left unsaid, thereby managing the emotional undercurrents of the conversation (Meeks, et. al., 1998). In contrast, while ChatGPT can identify and respond to explicit emotional expressions as scripted in scenarios like those shown in these scenarios, it does not inherently grasp the deeper emotional undercurrents that might be influencing the conversation.

Empathy is critical to breakup scenarios because it allows humans to not only recognize and acknowledge each other's feelings but often feel what the other person feels, which fosters a more genuine and potentially healing interaction. This level of empathetic engagement shows the intensity and sincerity that human interactions can convey—qualities that are difficult for LLMs to emulate. The model's strategy often involves using clear and direct language to articulate feelings and intentions, which, while effective, lacks real-time emotional recalibration—something humans naturally do—can make its interactions seem somewhat static and formulaic.

Moreover, ChatGPT relies on verbal cues to infer emotional states. Unlike humans, who can use a combination of verbal and non-verbal signals to gauge emotions, ChatGPT depends entirely on text input. This limitation is evident in scenarios where non-verbal cues play a large role in communication dynamics, such as in scenario 3 when I said, "I guess we could try therapy..." followed by "I just don't want to break up, so I'll do whatever you think we need." Most humans here would recognize a reluctance on my part, which would therefore indicate that I am not fully committed to the therapy. However, rather than acknowledging this or challenging me, it instead provides a list of therapy goals to consider.

These observations highlight a gap between AI and human abilities to navigate emotional complexities. While ChatGPT can simulate structured conversational elements, real-world interactions suggest that enhancing its capability to interpret and respond to non-verbal cues and emotional subtleties in real-time could significantly improve its effectiveness.

Conclusion

While ChatGPT demonstrates a remarkable ability to navigate breakup scenarios, it still falls short in several key areas when compared to human interactions. One of the primary limitations is the lack of genuine, spontaneous emotional depth in its responses. Although ChatGPT can simulate empathy and handle subtext to a certain extent, its responses can sometimes feel scripted, formulaic, and lacking the raw, illogical emotional reactions that are typical in human

breakups. This lack of authenticity can make the interaction feel artificial and less believable as a human-to-human exchange.

One of the key limitations lies in the transformer architecture used in most LLMs, including ChatGPT. Transformers are not specifically designed to model the intricate dynamics of emotional communication. The self-attention mechanism in transformers focuses on identifying relevant information within the input sequence, but it may not always capture the subtle emotional undertones and context-dependent meanings that humans intuitively grasp. Moreover, the training objective of LLMs is typically to generate semantically understandable text and often does not explicitly optimize for emotional understanding or empathy. As a result, the model may generate responses that are linguistically coherent but lack the emotional depth and authenticity that humans exhibit in real-life situations.

Another limitation is that ChatGPT relies solely on verbal cues to infer emotional states, whereas humans use a combination of verbal and non-verbal signals to gauge emotions. This limitation becomes evident in scenarios where non-verbal cues play a significant role in communication dynamics. For instance, when the human in the conversation expresses reluctance or hesitation, ChatGPT may not always pick up on these subtleties and adjust its responses accordingly, leading to a less nuanced and empathetic interaction. This stems from the fact that LLMs are trained on static datasets, which may not encompass the full spectrum of emotional expressions and non-verbal cues present in human interactions. The absence of multi-modal data, such as audio and visual cues, further limits the model's ability to interpret and respond to the complex emotional dynamics in a breakup scenario.

To improve the performance of ChatGPT and other LLMs in emotional communication scenarios, several changes could be made. Firstly, incorporating a more diverse dataset that includes a wider range of emotional expressions and non-verbal cues could help the model better understand and respond to the complexities of human emotions. This could involve training the model on data that includes not only text but also audio and visual cues, such as tone of voice, facial expressions, and body language.

Furthermore, incorporating a feedback loop that allows the model to learn from real-time interactions could help it continuously improve its emotional intelligence and adaptability. By analyzing user reactions and adjusting its responses accordingly, the model could develop a more nuanced understanding of human emotions and communication dynamics.

In conclusion, while ChatGPT and other LLMs is highly effective in simulating human-like conversations, there is still room for improvement in the realm of emotional communication. By incorporating more diverse datasets and continuous learning through user feedback, these systems could potentially bridge the gap between artificial and human emotional intelligence, leading to more authentic and empathetic interactions in the future.