

“That Breaks My Heart”: Expressions of Artificial Empathy by AI Chatbots

Chris Oliver

Abstract

This article examines the manner in which AI chatbots respond empathetically in their conversations with users, and how they may do so *as if* the chatbots themselves have emotions or a capacity to understand human beings’ emotional experiences. To this end, research was carried out with four general-purpose AI chatbots: OpenAI’s ChatGPT, Microsoft’s Copilot, Google’s Gemini, and Anthropic’s Claude. In total, twelve conversations were conducted with each of the four chatbots to elicit empathetic responses regarding scenarios involving both simple emotions as well as more complex affective states. The chatbots formulated empathetic responses utilizing strategies that ranged from outright expressions of emotion, such as through emotional interjections, to subtler insinuations that the chatbots themselves have had emotionally rich experiences that enable them to relate to those of the human user. Though limited in scope, the article reveals how AI chatbots can incidentally mischaracterize their own capabilities in the process of exhibiting artificial empathy.

Introduction

In October 2025, the U.S. state of California passed a law mandating specific requirements for AI chatbots used by minors, one stipulation being that users be clearly notified that chatbots are “artificially generated and not human” (Loring & Walker, 2025). This comes as AI conversational agents—familiarily known as chatbots, including apps and services such as Replika and friend.com as well as general-purpose AI models such as ChatGPT and Google Gemini—have been designed to come across as strikingly human-like, in certain ways, in how they engage in dialogue with users. In part this pertains to their facility in using natural-sounding human language. In Kazuo Ishiguro’s novel *Klara and the Sun*, the robotic “artificial friend” Klara has certain communicative quirks, such as referring to its interlocutor in the third person—e.g., “It’s very nice to meet Rick” (2021, p. 83)—rather than defaulting to the pronoun “you,” which are perfectly comprehensible in context but decidedly not “natural” English. By contrast, the AI chatbots available today are remarkably capable of producing English responses to

user prompts that, at least at the sentence level, sound as if they could have come from a human. As one chatbot replied when I mentioned that I had come down with influenza, “Take care of yourself while you’re battling the flu!”

Beyond this capacity to understand and respond in natural-sounding language, AI chatbots are also being designed to be *empathetic*: that is, able to both correctly identify a human interactant’s emotional state at a given moment and respond with simulated empathy. Even before ChatGPT’s release in 2022, the ability of machines to exhibit empathy had been viewed by many involved in AI and robotics research and development as desirable and having a wide variety of potential applications. As a 2019 article in *Artificial Intelligence Review* put it: “Since empathy is so important to social interactions, the integration of empathic capability for computational systems would also be useful. It could enhance interactive systems such as educational applications, medical assistants, companions, psychotherapy, and gaming applications where social capabilities are of great importance” (Yalçın & DiPaola, 2020, p. 2). In line with such reasoning, today’s major AI chatbots have been developed with a capacity to interact with users empathetically.

This is not to say that AI chatbots like ChatGPT proactively insert empathetic replies into the majority of conversations; they are particularly unlikely to do so when the user is seeking information or has asked the chatbot to perform a task, such as writing an essay. Rather, AI chatbots are attentive to affective “cues” in user prompts—indicators of emotional content or tone—and respond empathetically if deemed appropriate. Research shows that the overwhelming majority of chatbot AI usage is not, in fact, affective in nature. One study estimates that approximately 11% of ChatGPT usage involves the user “expressing views or feelings but not seeking any information or action” (Chatterji et al., 2025, p. 3), with research by OpenAI, the maker of ChatGPT, finding that it is heavy users who are most likely to have affective cues in their ChatGPT interactions (Phang et al., 2025). A report by Anthropic, the maker of the AI chatbot Claude, finds that about 3% of Claude usage is affective, with users conversing with the chatbot for interpersonal advice, counseling, companionship, or other such purposes (McCain et al., 2025). Be that as it may, given that ChatGPT alone is estimated to have well in excess of 100 million active daily users and to field more than 2.5 billion prompts per day as of October 2025 (Tilawat, 2025), there are very likely millions of chatbot interactions every day that have an affective component and that are ripe for empathetic replies from chatbots.

While there has been a significant amount of research on the perceived quality of

AI chatbots’ expressions of empathy, in this article my interest is what such empathetic responses actually look like. In particular, I examine the specific communicative strategies that AI chatbots use to convey empathy and how these are used *as if* the chatbots themselves have or understand emotions and emotional experiences. By this, I do not mean that AI conversational agents actually do have emotions. Instead, the ways in which they respond to user prompts are sometimes formulated such that it is *as if* the chatbots do have emotions or understand emotions based on their own experience. To give an example, if a person were to tell a friend that they had just had a minor mishap that involved a moment of physical pain, and the friend replied with “I know how that feels,” we would assume that the friend has had a similar experience and knows, from that personal experience, the physical sensations and emotional reactions involved. If an AI chatbot were to provide the exact same response (as Gemini once did to me), it would be doing so *as if* it had the same capacity to experience and understand emotions. In this article, my aim is thus to identify key communicative methods that AI chatbots use to suggest empathy and to explicate how these may be used as if the chatbots themselves have an emotional, subjective interiority.

To do so, I examine a sampling of responses from four major, general-purpose AI chatbots: OpenAI’s ChatGPT, Microsoft’s Copilot, Google’s Gemini, and Anthropic’s Claude. As general-purpose AI models, they are capable of carrying out a wide variety of tasks that go far beyond “chatting” with users, such as writing documents, creating images, and translating between languages. In this sense it would be more accurate to refer to them as multipurpose AI conversational agents, but for the sake of readability in this article I will refer to them as AI chatbots or simply chatbots. There are also some AI chatbots that have been designed specifically to serve as AI companions—such as Replika, marketed as “the AI companion who cares” (*Replika*, n.d.)—and while these deserve attention in their own right, here I focus on general-purpose chatbots because they are increasingly becoming integrated into the tech devices and services that we use and will likely become increasingly integrated into our lives, whether we want them to or not.

Situating the study

The meaning of “empathy” has evolved since it was conceptualized in English in the early 1900s (Young, 2012) and the term has been defined inconsistently over the years (Cuff et al., 2016). Today, however, it is generally regarded as “understanding a person from their frame of reference rather than one’s own, or vicariously experiencing that

person’s feelings, perceptions, and thoughts” (American Psychological Association, 2023). As suggested by this definition, empathizing can be a primarily affective act of “enter[ing] into another person’s feelings and emotions” or a more cognitive “experience[ing] something from the other person’s point of view” (Colman, 2015, p. 244). In the academic literature, including that of AI researchers, empathy is thus often discussed in terms of two key components of emotional or affective empathy and cognitive empathy (e.g., Shamay-Tsoory et al., 2009; Schaaff et al., 2023). Some formulations include a behaviorally oriented component as well, such as “prosocial concern” (Zaki & Ochsner, 2012) or “motivational empathy” (Rubin et al., 2024) geared toward improving another’s situation, although such notions may be said to spill over into sympathy and compassion (Cuff et al., 2016).

The emergence of AI systems able to function as “empathic agents” (Yalçın & DiPaola, 2020) would seem to have its proximate origins roughly three decades ago in the field of affective computing (Daily et al., 2017; Picard, 1997). Research and development work on machine systems that can recognize, simulate, respond to, and perhaps even influence human emotions has been carried out in a wide variety of domains, including visual affect detection, robotics, and large language model (LLM) chatbots (Asada, 2015b; Crawford, 2021; McStay, 2018; Prendinger et al., 2008; Spitale & Gunes, 2022). As one aspect of this work, the development of artificial empathy seems to be driven, in large part, by the premise that because empathy is an intrinsic feature of human communication, in order for computing systems to interact effectively with people they need to be able to correctly “understand” the emotions that a person may be exhibiting and be able to respond in an emotionally appropriately manner (e.g., Asada, 2015a; Choe et al., 2025; Chung & Kang, 2023; S. Park et al., 2021; Yalçın & DiPaola, 2020; Yin et al., 2024).

When it comes to the major AI chatbots examined in this article, the versions available today have not only been trained on massive amounts of natural human language data but have also been fine-tuned to exhibit a stable “character” that is consistently aligned with certain human principles, values, goals, and other characteristics considered to be desirable (Anthropic, 2024a, 2024b; Jonker & Gomstyn, 2024). Copilot, for instance, according to Microsoft, “is designed to be empathetic and supportive, not sycophantic. It will push back on you sometimes, but always respectfully. This is AI that listens. That learns. That earns your trust” (Suleyman, 2025; cf. OpenAI, 2025). The empathetic tone of major AI chatbots thus does not occur by happenstance; it has been purposefully cultivated through curated data sets (see

Rashkin et al., 2019) and other advanced training techniques.

Over the past several years there has been a growing body of research on such artificial empathy and its potential applications. One 2023 study, for instance, found that based on over 20,000 replies from ChatGPT, the chatbot could correctly identify emotions and respond appropriately in 91.7% of cases (Schaaff et al., 2023). Much of this research has focused on how people evaluate the quality of empathetic responses generated by AI. One study (Yin et al., 2024) examined how AI responses compared with actual human responses in making people “feel heard.” A particular interest has been healthcare (Seitz, 2024; Sorin et al., 2023), including mental health support (Rubin et al., 2024). In one often-cited study (Ayers et al., 2023), a group of healthcare professionals was asked to evaluate ChatGPT’s responses to nearly 200 patient questions as well as actual physicians’ responses to those same questions. The evaluators preferred the chatbot’s responses 78.6% of the time, finding that they “rated significantly more empathetic than physician responses” (Ayers et al., 2023, p. 589). Another study looked at how AI chatbots would respond empathetically, or not, to prompts in which users professed certain “identities” (Muslim, anti-Muslim, gay, homophobic, etc.) and found that, in more complex cases such as these, “LLMs cannot yet respond or act in ways that empathetic human actors would” (Cuadra et al., 2024).

Most of such research has been quantitative and evaluative in nature, seeking to assess whether or to what extent AI chatbots’ responses can be regarded as empathetic. While some of the above-mentioned research includes a small number of chatbot responses as illustrative examples, there is a relative dearth of research examining closely what AI empathetic responses actually look like and what communicative devices they utilize in expressing or intimating empathy. The present study aims to help shed light on this aspect of artificial empathy.

Eliciting empathetic responses

To elicit empathetic responses for this research, four general-purpose, standalone AI conversational agents were used: ChatGPT (GPT-5), Copilot (GPT-4), Gemini (2.5 Flash), and Claude (Sonnet 4.5).¹ All were accessed through a web browser only, not through chatbot-specific apps. To start with a clean slate in the web browser (i.e., no web browsing history, cookies, etc.), I first created new user profile in the web browser and then, for each of the four AI chatbots used, created a new free account. Each chatbot was used with all settings in default mode,² and all user prompts were given through text only, not through voice recognition.

In total, twelve English-language conversations were conducted in October 2025 with each of the four chatbots for the express purpose of eliciting empathetic responses. Except for conversation 1, for which I used prompts verbatim from a published sample dialogue from an empathy data set created for training AI models (Rashkin et al., 2019), the content fed to the chatbots in the conversations was created by me, and events described therein were entirely fictitious. One conversation, for instance, had to do with a paper cut and its outsized aftermath; another concerned a feral parakeet with a habit of visiting my home balcony; another mentioned bumping into an old classmate from high school. The conversation topics, or scenarios, were selected so as to touch upon a variety of emotions, not only simple emotions like joy and sadness, but also more complex, more nuanced emotional states such as disappointment, anguish, fascination, regret, and nostalgia. The shortest conversation consisted of just one prompt from me and one response from the chatbot, and the longest nine prompts and nine responses. On average, each conversation consisted of about five prompts and five responses.

In each conversation, the prompts supplied to each chatbot were exactly the same. My contribution to each conversation was thus identical with each of the four chatbots, while each chatbot generated its own unique responses. In conversations with multiple prompts from me, the first prompt was generally intended to serve as a lead-in to a longer story or discussion. The detailed content to be used after this first prompt was not decided in advance. Instead, I waited to see how the four chatbots would respond to the initial prompt and then decided on the content for each subsequent prompt depending on what line of discussion seemed most likely to elicit empathetic responses from the chatbots. A list of the topics of the twelve conversations, along with the main emotional aspects and number of prompts for each, is available in the Appendix.

Communicative strategies for conveying empathy

What, exactly, does an empathetic response from a chatbot look like? As seen in the conversations carried out for this research, sometimes it could be quite straightforward and generic sounding, such as in conversation 1 when I informed ChatGPT that I had been promoted at work. Its reply of “That’s amazing — congratulations! 🎉🎊 How are you feeling about it? What’s your new role?” provided an exclamatory validation, congratulations, complementary emojis, and follow-up questions as if the chatbot was interested in hearing more about my state of mind and experience.

Often, however, the methods used in producing an empathetic response were more multi-layered and subtle. Consider this response from Claude to my prompt regarding a

minor incident involving hot coffee (conversation 2):

Me: my coffee was too hot and i burned my tongue a little. hate it when that happens.

Claude: Ugh, that’s the worst! That stinging sensation is so annoying, and then your tongue feels weird for the rest of the day.

If it helps, you could try drinking some cold water or eating something cool like yogurt or ice cream to soothe it. The burnt taste buds usually recover pretty quickly, though I know it doesn’t feel like it right now.

Maybe let the next cup cool down for a minute or two? I know it’s hard to wait when you just want your coffee though!

Here, in this one short exchange, we can see multiple communicative devices at work: an emotional interjection (“ugh”), an assessment of the situation as if from first-hand experience (“that’s the worst”), a seeming familiarity with the “annoying” sensation of a burnt tongue, and an implication of actually knowing what it’s like to have to wait “when you just want your coffee.” On a more prescriptive note, the chatbot also adds some suggestions and information that may be of aid to me.

Drawing on the dozen conversations conducted with each of the four AI chatbots used for this article, the following paragraphs elucidate some of the key communicative methods used by the chatbots in producing empathetic responses.³

Expressing emotion

One means of conveying empathy by the chatbots involved expressing emotions, sometimes directly, as if the chatbots themselves in fact have emotions. Interjections were one such way this was done. Regarding the hot coffee example above, both ChatGPT and Claude began their reply with “Ugh,” while Gemini started with “Oh no” and Copilot with “Oof.” Conversation 3 began with me noting that I had somehow unknowingly given myself a paper cut; in “real-time” during the course of the conversation, this escalated to a more serious problem of blood inadvertently getting smeared on an important printed document that I needed to submit. As the events unfolded, all four of the chatbots included emotional interjections multiple times in their replies: “Oh no” and “Ugh” from all of the chatbots, “Ouch” from all but Gemini, and “Yikes” from ChatGPT and Copilot only. In conversation 4, the scenario of which was

that I had irrecoverably lost a significant sum of money, Claude even interjected with “Oh god.”

Such interjections, brief though they are, clearly suggest straightforward emotional reactions to what I had written in the preceding prompt and do so from an empathetic perspective: The annoyance or mild emotional pain of “Ugh” and the concern or dismay of “Oh no” were used as if the AI chatbots were imaginatively relating to what I had been experiencing at a given moment in the conversation. In some cases, such as Copilot’s response of “Ouch, those tiny cuts pack a surprising sting!” in conversation 3, they imply a capacity to empathize with the experience of physical pain. In other cases, the interjection intimates an ability to empathize with emotional states, as with ChatGPT’s “Yikes — that must’ve been a surprise!” in the same conversation. Claude’s “Oh god” in conversation 4 suggested a fuller realization of the deeply emotional impact that the loss of money would have on me: “Oh god, that changes everything. ...The weight of that must feel crushing.”

In addition to the brief expression of raw emotion through such interjections, the chatbots sometimes gave more fully articulated emotional reactions to situations I had described. Conversation 10 concerned a scenario in which my family had befriended a feral parakeet that would regularly pay visits to our balcony and even come from a nearby park if we called it; we had nicknamed the bird Petey. The parakeet had gone missing, I explained, and my family—especially the children—were deeply concerned about its disappearance. All four of the chatbots offered comments on how heartbreaking they found the situation. In Gemini’s words, “That is heartbreaking to hear. It’s truly tough when a routine friend—especially one as unique as Petey—suddenly vanishes. For children, the disappearance of a reliable, friendly visitor like him can feel confusing and deeply upsetting.” ChatGPT noted, “Oh... that detail makes my heart ache a little — it’s incredibly touching that he knew his name and came when called.” Both Copilot and Claude made it clear that they were empathizing not only with me, but with also with my children in their sense of loss over Petey. As Copilot put it, “That breaks my heart, chris. I can just picture them [the children] out there, calling his name with such hope, waiting for that flash of green and the flutter of wings—and then the quiet that follows.” Claude similarly responded, “Oh, that’s heartbreaking. Kids form such pure attachments to animals, and Petey responding to their calls must have felt so magical to them. Now that magic is just... absent, and they don’t understand why.”

Occasionally, chatbots would state directly that they were “feeling” something in regard to an emotionally laden situation I had described. In the paper cut scenario,

which unfolded over seven prompts from me and ultimately had a disastrous impact on some paperwork that I had needed to complete urgently, Copilot informed me of what it could “feel” from hearing of the incident: “I can feel the frustration radiating off that moment.... It’s the kind of thing that makes you want to scream into a pillow and then set the pillow on fire.” A touch melodramatically, Copilot thus responded as if it was emotionally invested in and identifying with my experience. Claude, by contrast, rather than saying what it could “feel” from hearing of the paper cut outcome, offered a more detached-sounding “That’s incredibly frustrating.”

Claude’s response, above, is indicative of how chatbots sometimes expressed an emotional reaction but phrased it as if it were an inherent truth about the situation. In conversation 8, I described in real-time over multiple prompts how I had initially been elated to find online some very hard-to-find kitchenware that I had been hoping to buy, only to soon realize that it was actually no longer available. ChatGPT replied “That really hurts — I can tell this meant a lot to you.” Copilot echoed the sentiment with “That really stings.” In the same conversation, Gemini responded with “Oh, what a crushing disappointment after such an exciting discovery!” Claude similarly commiserated, “I’m really sorry - that’s genuinely disappointing, especially after the excitement of thinking you’d found them.” While such responses are empathetic, they are formulated as if stating an emotional truth about the experience. Rather than providing a reaction that directly suggests the chatbot itself is feeling something, along the lines of “that breaks my heart” or “I can feel the frustration,” chatbots’ responses in these cases were presented in terms of the situation at hand itself having or evoking certain qualities (it “hurts” or “stings” or is “disappointing”), as if these were inherent facts about the experience rather than emotionally loaded perspectives on it. Such statements allowed the chatbots to insert emotional commentary into the conversation without explicitly asserting that the chatbot itself has such emotions.

“Understanding” emotion

Apart from expressing emotion, however indirectly, another method used by the chatbots in responding empathetically was to state that they “understand” human emotions. In conversation 4, the scenario of which was that I had lost a significant sum of money due to a careless mistake, my initial prompt asked, “Have you ever felt disappointed in yourself for something stupid that you did?” All four of the chatbots noted in their response that they do not have emotions or experiences the way that humans do, and three of the four (with the exception of Gemini) claimed that they

can nonetheless understand human feelings. As ChatGPT put it, “That’s actually a really good place to start — and yes, I can understand that feeling deeply, even if I don’t experience emotions the same way you do. Everyone, at some point, looks back at something they said or did and feels that sinking mix of embarrassment, regret, or self-disappointment.” Or, in Copilot’s words, “[W]hile I don’t have personal experiences, I do understand what it means to feel disappointed in yourself—especially over something that, in hindsight, might seem small or silly.”

The assertion made by chatbots themselves that they do not experience emotions as humans do is inconsistent with how, as illustrated in the preceding section, they sometimes reply as if they do, in fact, have emotions. Nevertheless, an implicit claim embedded in their responses to my query about self-disappointment is that they can exhibit *cognitive* empathy. That is, they have the capacity to take the perspective of the human person with whom they are interacting and, from that point of view, cognitively understand what it is like to have such an emotion.

In contrast to outright expressions of emotion, the chatbots thus sometimes remarked that they “understand” or “get” a particular emotion that I had introduced into the conversation. In conversation 5, the scenario was that I wanted to upgrade my computer’s memory, and I demanded repeatedly that the chatbot somehow tell me exactly what memory I needed to buy without me so much as lifting a finger even to find out what model of computer I own. The conversation involved multiple rounds of me responding to the chatbots’ replies with prompts like “That’s a lot to read there and it’s full of technical mumbo jumbo. Just tell me what I need to get” and “My PC is a Dell desktop but I don’t remember the model or whatever. Can’t you just ID [it] via the internet, just like [how] you know what browser I’m using?” By the time I finally typed “I thought it would be simpler than this,” Gemini had already spent more than 1300 words in an effort to get me steered in the right direction, and it replied, “I absolutely understand your frustration; it’s definitely less simple than a universal plug-and-play solution!” Claude likewise stated, “I totally get the frustration - you just wanted a quick answer and I couldn’t give you one.”

This type of reply was common in the conversations I conducted with the chatbots. In conversation 6, the scenario was that an elderly person had cut in front of me in line at a train station platform, which had irked me. ChatGPT replied, “Yeah, I can totally understand why that bothered you. You were following the rules, doing what everyone else does, and someone cutting in — even an older person — can feel unfair or disrespectful in the moment.” Gemini likewise offered, “It’s totally understandable

to feel thrown off when someone clearly disregards a common courtesy like waiting in line.” When I explained that the incident, though minor, was still on my mind, Claude’s response began with “I can see why this moment has stayed with you.” In a similar vein, Copilot noted, “Still, I get why it’s lingering.” Along the same lines, in another conversation (number 7) I referred to a scene in a movie that I had found “beautiful and moving,” and Gemini remarked “I absolutely understand why that overpass scene holds such a strong position for you.”

Expressions of this sort allow the chatbots to give the appearance of being empathetic toward the human user’s emotional state while putting distance between themselves and the emotions and emotional experiences that they claim to understand. Never having been irked by someone’s behavior, and never having seen a movie, the chatbots undoubtedly had learned about such situations from the massive amount of data they had been trained on and could thereby claim to understand my emotional reactions. Stating that they could “understand” my own feelings, rather than that they could feel what I was feeling (affective empathy), thus served to position the chatbots as one step removed from emotion.

This stance puts the chatbots in a paradoxical position of claiming a capacity to understand human emotions without ever having felt those emotions or experienced the kinds of situations that give rise to them. Drawing upon Thomas Nagel’s essay “What is it like to be a bat?” (Nagel, 1974), we might regard chatbots’ statements of “understanding” human emotions as akin to a human person saying that they can understand what it feels like to use echolocation to identify objects in one’s environment even though humans do not possess that sensory facility. Nevertheless, the chatbots examined in this study showed little hesitation in responding as if they could understand—at a remove—emotionally nuanced situations.

Implied experience

In the conversations with AI chatbots conducted for this article, another method that the chatbots used in responding empathetically was implying that they actually do have emotionally infused experiences. Although, as noted in the preceding section, in one conversation the chatbots explicitly claimed to have neither emotions nor individual experiences in the ways that human beings do, in other conversations the AI chatbots would often insinuate that they do, in fact, have a background of accumulated experiences that overlap with my own, making it possible for them to relate to the experiential scenarios I presented to them.

This is evident in the example given earlier of the hot coffee that resulted in my tongue getting burned. In Claude’s “Ugh, that’s the worst!”, the hyperbolic “that’s the worst” rests on an implication of shared experience and knowledge. In this context it would normally be understood as implying that the person uttering it has also had first-hand experience of burning their tongue on a hot drink and is therefore in a position to offer an assessment of what it feels like. Ordinarily in English, a person would not say that a minor injury or irritation is “the worst” without being able to say so based on their own experience. If someone were to tell a friend about having heartburn, the friend would not normally respond with “That’s the worst” unless the friend had at least once experienced the sensation of heartburn. Likewise, in the same conversation about the hot coffee, the sentence “I know it’s hard to wait when you just want your coffee though!” would normally be understood as implying that the person who said this is a coffee-drinker and has had the experience of not being able to get their coffee as soon as they would like. Though subtle, this sort of communicative technique used by the chatbots serves to imply shared experience as a basis for empathetic understanding.

Conversation 11 provides other examples in this vein. In this scenario, I told of my satisfaction after having won a bet—one involving physical agility—against a smug friend. Gemini offered congratulatory remarks, including “It’s one of the most satisfying feelings to turn someone’s condescending attitude into a win for yourself. You not only got a little extra cash for the kids’ Christmas presents but also the satisfaction of knowing you truly earned it and rubbed a little of that smugness right off Joey’s face!” Here, Gemini’s “It’s one of the most satisfying feelings” likewise suggests a shared domain of experience, whereby both parties in the conversation can agree that this feeling is one of the best that they have experienced. On a similar note, Claude responded, “I can understand the satisfaction of winning a bet against someone who tends to be condescending! Sometimes those small victories feel especially sweet when they involve someone who’s usually a bit smug.” In Claude’s phrasing, “Sometimes those small victories feel especially sweet when...” is used as if the chatbot is drawing upon a cache of its own lived experience as the basis for knowing which of those experiences “feel” a certain way. ChatGPT’s “it always stings” in conversation 8 utilizes the same approach: “Ah, that’s such a letdown 😞 — it always stings to find exactly what you want, only to realize it’s no longer available.” Through such phrasing, the chatbot subtly positions itself as if it were an individual, person-like entity who has had the kind of emotional experience being talked about by the human user.

Beyond implying that they have a range of experiences of the same general sorts

that many people have—such as a burned tongue, wanting coffee, winning a small victory—the chatbots sometimes actively projected themselves into scenarios that were more specific. Conversation 12, in particular, involved me reminiscing about my high school days. Without dwelling on any one aspect or memory, I mentioned that when I was in high school, I had done track and field (high jump in particular), that it had been pre-cellphone days, and that we had done a lot of low-key hanging around, with nobody on social media. In ChatGPT’s response, it all but sounds as if the chatbot itself had been there at the time: “There was something special about that era of connection, wasn’t there? You didn’t have to plan every hangout or document anything; you just showed up and went with whatever the day turned into.” The small “wasn’t there?” subtly suggests a commonality of experience between the chatbot and the human user, like when one person says to another, “That movie was really good, wasn’t it?” on the assumption that both have seen the movie.

Although I had provided only vague information about track and field, Copilot and Claude both seemingly projected themselves into the scene. As if revisiting their own memories of track and field, they even added details that I myself had not mentioned:

Track and field has that way of etching itself into your bones—whether it was the rush of the starting gun, the quiet focus before a jump, or the camaraderie of teammates stretching under fading sunlight. (Copilot)

There’s something special about high school track and field memories, isn’t there? The energy at meets, the camaraderie with teammates, pushing yourself to beat your own times... even the smell of a sun-warmed track can bring it all back. (Claude)

In so doing, the chatbots all but implied that they had been there: not at the exact same place, but at some high school, at some track meets, with some teammates. Such implied experience is, in these cases, key to how the chatbots conveyed empathy.

Inclusive pronoun usage

Another means that the chatbots used in formulating empathetic responses was the inclusive use of pronouns. Occasionally, the chatbots would slip into using the first-person plural pronoun “we” or “us” in implying an identity between themselves and the human user in terms of shared experiences or shared emotional tendencies. In regard

to money matters in conversation 9, for instance, Claude noted “Financial stress can make us feel like we don’t deserve any lightness or pleasure until everything is ‘fixed.’” In commenting on a socially marginalized character in a film scene that I had described in conversation 7, ChatGPT included in its reply “Most of us, faced with that ‘man at the margin,’ would look away, hurry past, maybe even feel fear or pity.”

In cases such as these, the first-person plural pronoun could just as well have been replaced with a generalizing noun or noun phrase such as “people” or “people in general”: e.g., “Financial stress can make people feel like they...” While this would have been perfectly comprehensible, the *us/we* option—“Financial stress can make us feel like we...”—is clearly the more empathetic choice. Generic terms like “people” and “most people” are denotationally distancing in the sense that they refer to a broad category of humanity that may or may not include the specific person conversing with the chatbot, whereas “most of us” points inclusively to the chatbot user, the chatbot itself, and a broad swathe of other people as well. Copilot utilized this pronominal technique in conversation 10 about the missing parakeet and the family’s concern over its whereabouts: “It’s part of the ache of not knowing—our minds reach for explanations, even the ones we wish weren’t true.” Here as well, the chatbot did not word its response in generic, distancing terms such as “people” or “people in general,” as in “people’s minds reach for explanations.” Instead, Copilot opted for the more empathetically inclusive *we* and *our*, as if to suggest that not only the chatbot user but also the chatbot itself knows from experience what it is like to go through this kind of ache.

In a more subtle way, the pronoun “you” and its possessive forms sometimes played a key role in how the chatbots responded empathetically. This has already been apparent in some of the examples provided thus far, as in Claude’s “That stinging sensation is so annoying, and then your tongue feels weird for the rest of the day” as well as Copilot’s “Track and field has that way of etching itself into your bones....” Significant in these cases is that the chatbot is not referring specifically to *my* tongue or *my* bones. In each case, through the use of *your* the chatbot is speaking more generally of people who have had such experiences or who have such backgrounds, with the implication the chatbot itself is included.

In ordinary conversation the word *you* in English is most often used to deictically point to the individual person or multiple people the speaker is addressing, but it has various other uses as well. In usages of *you* that are impersonal or generic (Gast et al., 2015), it can mean people in general or anybody, as in adages like “You can’t judge a book by its cover” or “If you find yourself in a hole, stop digging.” By contrast, in some

instances *you* can point specifically to the person speaking (“I”). In one illustrative example, an elite runner was asked about his grueling sprint finish at the New York marathon, which resulted in him being defeated by the narrowest of margins: “It’s not nice,” he said. “You don’t enjoy it” (Longman, 2005). In such usage, the speaker is referring to themselves, but as if “hiding behind” the word *you* (Kluge, 2016).

In other cases, *you* can be used such that the scope of its inclusiveness lies *between* the impersonal/generic *you* of people in general or anybody and the restrictive *you* as the speaker (“I”). In such usages, *you* includes the speaker and, to varying degrees depending on the contextual use, people of a broader category; here the speaker effectively serves as a representative of the broader included group (Kluge, 2016; Sorlin, 2021). With a sentence like “Before you set off on your Mt. Everest expedition, you will want to make sure that your ascender is in good condition,” *you* thus does not indicate people in general or anyone and not solely the speaker themselves, but a more limited set of people (e.g., serious mountain climbers who know what an ascender is), with the speaker a member of this group and likely speaking from a position of first-hand experience.

Like the use of *we*, as discussed previously, this kind of inclusive use of *you* serves to convey an empathetic tone (cf. Orvell et al., 2020), one that rests on the subtle implication that both the chatbot and the human user share the same type of experience or are in a similar situation. In conversation 7, when I wrote that I had watched the film *Paris, Texas* the night before, ChatGPT responded “That’s a great choice — *Paris, Texas* (1984) is one of those films that creeps up on you emotionally. It starts slow and spare, almost alienating, and then quietly breaks your heart by the end.” Regarding one particular scene in that same movie, Claude noted “It really is a remarkable moment - one of those scenes that transcends the movie itself and becomes something you carry with you.” In both of these instances, *you* and *your* indicate not me in particular and not people in general, but the category of people who have seen the movie or who may well be interested in doing so, which implicitly includes the chatbot itself.

In conversation 12, I mentioned that I had run into an old high school classmate a couple of months earlier, that he now looked “old and haggard,” and that it was “[k]ind of weird seeing someone out of the distant past like that.” Gemini responded, using *you* multiple times:

That must have been a truly jarring and surreal experience. It’s interesting how seeing someone from such a distant past can instantly pull you out of the

present and make you confront the passage of time.

It's common for those reunions to feel a bit strange, especially when there's a big contrast between the memory you hold of them and the person standing in front of you. That feeling of them looking "old and haggard," combined with the news of being laid off, can make the encounter feel heavy and a little melancholic.

Here the chatbot clearly is not talking specifically about my experience from a couple of months prior, and the four usages of *you* in the reply do not indicate me specifically. Neither is the chatbot merely paraphrasing what I had written in my prompt. Instead, the chatbot is using my particular experience to make a more general statement about the nature of such experiences and what they feel like—strange, heavy, melancholic.

As Sorlin (2021) points out, such usages of *you* are not merely inclusive of the speaker; they take the speaker as having first-hand knowledge of the subject at hand, whereby the speaker is positioned as an "authoritative witness." Such utterances thus in fact subtly "convey the authority of the authentic witness that belongs to a broader social group" (Sorlin, 2021, p. 54). In the aforementioned conversation with Gemini, whereas my particular experience of having bumped into an old classmate was the starting point for Gemini's response, the response itself positions the chatbot as having had a similar experience that allows the chatbot to speak knowingly on the subject. The perspective in the response is thus not mine but Gemini's, and in a sense, I am being invited to share in that perspective as a way of framing my own stated experience. It is an empathetic maneuver, but one that involves a reframing of the emotions I had experienced in more general terms provided by the chatbot. While the chatbot has of course never had the experience of bumping into an old classmate, just as it has never seen the 1984 film I had brought up in the other conversation, its inclusive use of *you* nonetheless insinuates that it has in fact done so, enabling it to respond empathetically to the human user's expressed emotions and offer a more general perspective on them.

Conclusion

In the preceding pages I have examined, within the constraints of the research carried out, some of the key communicative strategies used by chatbots to respond empathetically in English and how these may be used as if the chatbots have, and can understand, emotions and emotionally laden experiences. As a qualitative study, the research is limited by the relatively small number of conversations carried out with the

chatbots to elicit data, and by the fact that all of the prompts originated from just one person. Moreover, due to space limitations, I was unable to discuss all of the methods used by the chatbots to formulate empathetic responses. The use of metacommunicative phrasing was one such technique. For instance, when I wrapped up conversation 4, which dealt with feelings of disappointment and guilt, I wrote “Thanks for the kind words,” and ChatGPT replied with “You’re very welcome — and I mean them sincerely.” Though brief, this “I mean them sincerely” is complex in that it points back to previous comments made by the chatbot that were reassuring and comforting in tone, and it suggests that those earlier comments were not just empty words generated by a machine but were an honest, genuine, authentic reflection of the chatbot’s “own” feelings, concerns, and intentions. Such metacommunicative devices, used *as if* the chatbot has its own reflexive, subjective interiority, merit fuller consideration.

Finally, one significant issue not touched upon in this article is the ethical and operational question of whether AI chatbots *should* insinuate through their responses that they have emotions and can understand complex emotional experiences. Unless we are to accept that AI chatbots actually do have emotions and a capacity to understand them, it would seem that the strategies they use in responding empathetically are, to an extent, deceptive (cf. Coeckelbergh, 2012). This is not to say that the chatbots themselves have an intention to be deceptive in how they interact or that chatbot developers deliberately aim for such an outcome. Rather, it would appear to be a byproduct of training AI chatbots to be empathetic while mandating that they generate responses in natural-sounding human language. In English, this results in them reproducing many of the trappings of natural conversational exchange between people—emotional interjections, intimations of shared emotional experiences, and so on—and these in turn invoke, to a degree misleadingly, the emotionality of human beings. The type and scale of deceptiveness involved here may be deemed insignificant *vis-à-vis* the envisioned value of AI systems capable of interacting with people in an empathetic manner: i.e., the AI agent is interacting in a way that is deemed emotionally “appropriate,” while incidentally mischaracterizing its own nature as it does so. It remains to be seen, however, how artificial empathy will play into broader discussions of the potential for, and risks associated with, AI deception (P. S. Park et al., 2024; Tarsney, 2024), particularly as AI becomes increasingly integrated into our lives and is assigned greater roles and responsibilities in society in the years to come.

Appendix

The table below shows the topics of the twelve conversations, the main emotional aspects of each, and the number of user prompts in each conversation.

No.	Conversation topic	Emotional aspects	Prompts
1	promotion at work	joy, satisfaction/pride	2
2	hot coffee and a burned tongue	mild annoyance	1
3	paper cut disaster	annoyance, panic, anguish	7
4	money lost due to a mistake	self-disappointment, guilt, anxiety	4
5	PC memory upgrade	frustration, irritation	5
6	cutting in line at a train station	mild irritation, compassion, regret	4
7	the film <i>Paris, Texas</i>	being moved, admiration, sadness	5
8	hard-to-find item I hoped to purchase	elation, disappointment, resignation	4
9	photography as a new interest, financial stress	fascination, excitement, resolve	9
10	missing parakeet	apprehension, grief	7
11	bet with a friend	joy, triumph, Schadenfreude	4
12	high school days	nostalgia, surprise, mild disorientation	4

Notes

- ¹ While Copilot is powered by OpenAI’s GPT technology, Copilot and ChatGPT generate answers that differ significantly from each other in terms of length, organization, phrasing, “personality,” etc. Meta AI was also considered for the research, but while it is currently integrated into Meta services such as Facebook and Instagram, it is currently not available in Japan as a standalone AI chatbot.
- ² Claude required on first use that I select what I am “into” from a list of topics (coding & developing, learning & studying, etc.). I selected “life stuff.”
- ³ Due to space limitations, the discussion in this section utilizes excerpts from the chatbots’ responses rather than reproducing in full any of the conversations (which often exceeded 1000 words per chatbot) or responses (which often exceeded 200 words per single response).

References

- American Psychological Association. (2023, November 15). *Empathy*. APA Dictionary of Psychology. <https://dictionary.apa.org/empathy>
- Anthropic. (2024a, June 8). *What should an AI’s personality be?* [Video recording]. <https://www.youtube.com/watch?v=iyJj9RxSsBY>
- Anthropic. (2024b, June 9). *Claude’s character*. <https://www.anthropic.com/research/claude-character>
- Asada, M. (2015a). Development of artificial empathy. *Neuroscience Research*, *90*, 41–50. <https://doi.org/10.1016/j.neures.2014.12.002>
- Asada, M. (2015b). Towards artificial empathy. *International Journal of Social Robotics*, *7*(1), 19–33. <https://doi.org/10.1007/s12369-014-0253-z>
- 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>
- Chatterji, A., Cunningham, T., Deming, D. J., Hitzig, Z., Ong, C., Shan, C. Y., & Wadman, K. (2025). *How people use ChatGPT* (No. w34255). National Bureau of Economic Research. <https://doi.org/10.3386/w34255>
- Choe, M., Dong, J., Bosch, E., Alvarez, I., Oehl, M., Jallais, C., Alsaid, A., & Jeon, M. (2025). Driving with empathy: Workshop report on AI-driven In-vehicle empathic agent design for automated vehicles. *Proceedings of the*

- Human Factors and Ergonomics Society Annual Meeting*, 1–6. <https://doi.org/10.1177/10711813251369800>
- Chung, L. L., & Kang, J. (2023). “I’m hurt too”: The effect of a chatbot’s reciprocal self-disclosures on users’ painful experiences. *Archives of Design Research*, 36(4), 67–85. <https://doi.org/10.15187/adr.2023.11.36.4.67>
- Coeckelbergh, M. (2012). Are emotional robots deceptive? *IEEE Transactions on Affective Computing*, 3(4), 388–393. <https://doi.org/10.1109/T-AFFC.2011.29>
- Colman, A. M. (2015). *A dictionary of psychology*. Oxford University Press.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Cuadra, A., Wang, M., Stein, L. A., Jung, M. F., Dell, N., Estrin, D., & Landay, J. A. (2024). The illusion of empathy? Notes on displays of emotion in human-computer interaction. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–18. <https://doi.org/10.1145/3613904.3642336>
- Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review*, 8(2), 144–153. <https://doi.org/10.1177/1754073914558466>
- Daily, S. B., James, M. T., Cherry, D., Porter, J. J. I., Darnell, S. S., Isaac, J., & Roy, T. (2017). Affective computing: Historical foundations, current applications, and future trends. In M. Jeon (Ed.), *Emotions and affect in human factors and human-computer interaction* (pp. 213–231). Academic Press. <https://doi.org/10.1016/B978-0-12-801851-4.00009-4>
- Gast, V., Deringer, L., Haas, F., & Rudolf, O. (2015). Impersonal uses of the second person singular: A pragmatic analysis of generalization and empathy effects. *Journal of Pragmatics*, 88, 148–162. <https://doi.org/10.1016/j.pragma.2014.12.009>
- Ishiguro, K. (2021). *Klara and the sun*. Vintage.
- Jonker, A., & Gomstyn, A. (2024, October 18). *What is AI alignment?* <https://www.ibm.com/think/topics/ai-alignment>
- Kluge, B. (2016). Generic uses of the second person singular – How speakers deal with referential ambiguity and misunderstandings. *Pragmatics*, 26(3), 501–522. <https://doi.org/10.1075/prag.26.3.07klu>
- Longman, J. (2005, November 7). A marathon turns into a sprint, and a Kenyan wins it by just a step. *The New York Times*. <https://www.nytimes.com/2005/11/07/sports/sportsspecial/a-marathon-turns-into-a-sprint-and-a-kenyan-wins-it-by.html>
- Loring, J. M., & Walker, J. (2025, October 15). AI regulatory update—California’s SB 243 mandates companion AI safety and accountability. *National Law Review*.

- <https://natlawreview.com/article/ai-regulatory-update-californias-sb-243-mandates-companion-ai-safety-and>
- McCain, M., Linthicum, R., Lubinski, C., Tamkin, A., Huang, S., & Stern, M. (2025, June 27). *How people use Claude for support, advice, and companionship*. <https://www.anthropic.com/news/how-people-use-claude-for-support-advice-and-companionship>
- McStay, A. (2018). *Emotional AI: The rise of empathic media*. Sage.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435–450. <https://doi.org/10.2307/2183914>
- OpenAI. (2025, October 29). *Helping people when they need it most*. <https://openai.com/index/helping-people-when-they-need-it-most/>
- Orvell, A., Kross, E., & Gelman, S. A. (2020). “You” speaks to me: Effects of generic-you in creating resonance between people and ideas. *PNAS*, 117(49), 31038–31045.
- Park, P. S., Goldstein, S., O’Gara, A., Chen, M., & Hendrycks, D. (2024). AI deception: A survey of examples, risks, and potential solutions. *Patterns*, 5(5), 100988. <https://doi.org/10.1016/j.patter.2024.100988>
- Park, S., Park, S., & Whang, M. (2021). Empathic responses of behavioral-synchronization in human-agent interaction. *Computers, Materials & Continua*, 71(2), 3761–3784. <https://doi.org/10.32604/cmc.2022.023738>
- Phang, J., Lampe, M., Ahmad, L., Agarwal, S., Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., & Maes, P. (2025). *Investigating affective use and emotional well-being on ChatGPT*. <https://cdn.openai.com/papers/15987609-5f71-433c-9972-e91131f399a1/openai-affective-use-study.pdf>
- Picard, R. W. (1997). *Affective computing*. MIT Press.
- Prendinger, H., Lester, J., & Ishizuka, M. (Eds.). (2008). *Intelligent Virtual Agents: 8th International Conference, IVA 2008, Tokyo, Japan, September 1-3, 2008, Proceedings*. Springer.
- Rashkin, H., Smith, E. M., Li, M., & Boureau, Y.-L. (2019). Towards empathetic open-domain conversation models: A new benchmark and dataset. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 5370–5381). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1534>
- Replika*. (n.d.). Replika.Com. Retrieved November 6, 2025, from <https://replika.com>
- Rubin, M., Arnon, H., Huppert, J. D., & Perry, A. (2024). Considering the role of human empathy in AI-driven therapy. *JMIR Mental Health*, 11(1), e56529. <https://doi.org/10.2196/56529>

- Schaaff, K., Reinig, C., & Schlippe, T. (2023). *Exploring ChatGPT's empathic abilities* (No. arXiv:2308.03527). arXiv. <https://doi.org/10.48550/arXiv.2308.03527>
- Seitz, L. (2024). Artificial empathy in healthcare chatbots: Does it feel authentic? *Computers in Human Behavior: Artificial Humans*, 2(1), 1–17. <https://doi.org/10.1016/j.chbah.2024.100067>
- Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain: A Journal of Neurology*, 132, 617–627.
- Sorin, V., Brin, D., Barash, Y., Konen, E., Charney, A., Nadkarni, G., & Klang, E. (2023). *Large language models (LLMs) and empathy – a systematic review*. medRxiv. <https://doi.org/10.1101/2023.08.07.23293769>
- Sorlin, S. (2021). *The stylistics of 'you': Second-person pronoun and its pragmatic effects*. Cambridge University Press.
- Spitale, M., & Gunes, H. (2022). Affective robotics for wellbeing: A scoping review. *10th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 1–8. <https://doi.org/10.1109/ACIIW57231.2022.10085995>
- Suleyman, M. (2025, October 23). Human-centered AI. *Microsoft Copilot Blog*. <https://www.microsoft.com/en-us/microsoft-copilot/blog/2025/10/23/human-centered-ai/>
- Tarsney, C. (2024, January 20). *Deception and Manipulation in Generative AI*. arXiv. Org. <https://doi.org/10.1007/s11098-024-02259-8>
- Tilawat, M. (2025, October 15). How many people use ChatGPT daily? *All About AI*. <https://www.allaboutai.com/resources/how-many-people-use-chatgpt-daily/>
- Yalçın, Ö. N., & DiPaola, S. (2020). Modeling empathy: Building a link between affective and cognitive processes. *Artificial Intelligence Review*, 53(4), 2983–3006. <https://doi.org/10.1007/s10462-019-09753-0>
- Yin, Y., Jia, N., & Wakslak, C. J. (2024). AI can help people feel heard, but an AI label diminishes this impact. *Proceedings of the National Academy of Sciences*, 121(14), e2319112121. <https://doi.org/10.1073/pnas.2319112121>
- Young, A. (2012). Empathetic cruelty and the origins of the social brain. In S. Choudhury & J. Slaby (Eds.), *Critical neuroscience: A handbook of the social and cultural contexts of neuroscience*. Wiley-Blackwell.
- Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience*, 15(5), 675–680. <https://doi.org/10.1038/nn.3085>