Using a Secondary Channel to Display the Internal Empathic Resonance of LLM-Driven Agents for Mental Health Support

Matthias Schmidmaier
LMU Munich
Munich, Germany
matt@schmidmaier.org

Jonathan Rupp
University of Innsbruck
Innsbruck, Austria
info@jonathan-rupp.com

Sven Mayer
LMU Munich
Munich, Germany
TU Dortmund University
Dortmund, Germany
info@sven-mayer.com

Abstract

Conversational agents are becoming increasingly popular for digital mental health support. However, while empathy is essential for effective emotional support, the unimodal request-response interaction of such systems limits empathic communication. We address this limitation through a secondary channel that displays an agent's inner reflections, similar to how nonverbal feedback in human interaction conveys cognitive and emotional states. We implemented a chatbot that generates not only conversational responses but also describes internal reasoning and emotional resonance. A user study involving $N = 188$ participants indicated a statistically significant increase in perceived empathy (+14.7%) when the agent's internal reflections were displayed. Our findings demonstrate a practical method to enhance empathic interaction with LLM-based chatbots in empathy-critical contexts. Additionally, this work opens possibilities for multimodal systems where LLM-generated reflections may serve as input for generating nonverbal feedback.

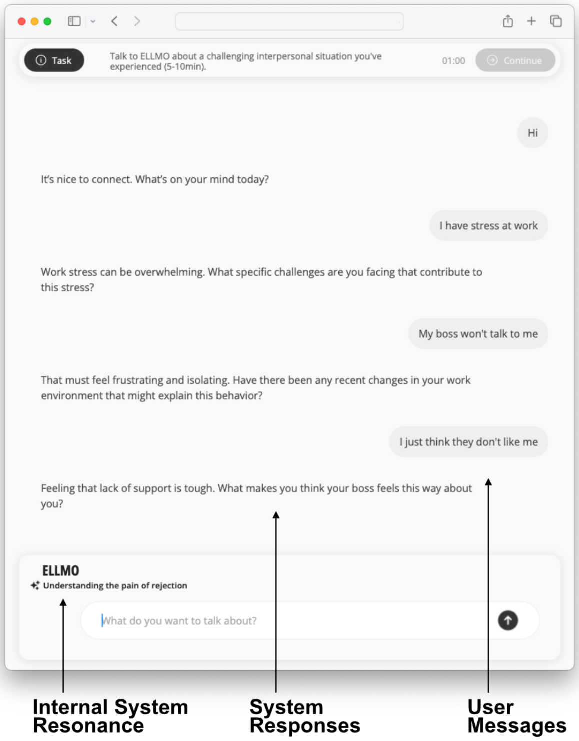
CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**.

Keywords

HCI; LLM; empathy; mental health; internal feedback

ACM Reference Format:
Matthias Schmidmaier, Jonathan Rupp, and Sven Mayer. 2025. Using a Secondary Channel to Display the Internal Empathic Resonance of LLM-Driven Agents for Mental Health Support. In *Proceedings of the 27th International Conference on Multimodal Interaction (ICMI '25), October 13–17, 2025, Canberra, ACT, Australia.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3716553.3750759



Talk to ELLMO about a challenging interpersonal situation you've experienced (5-10min).

Hi

It's nice to connect. What's on your mind today?

I have stress at work

Work stress can be overwhelming. What specific challenges are you facing that contribute to this stress?

My boss won't talk to me

That must feel frustrating and isolating. Have there been any recent changes in your work environment that might explain this behavior?

I just think they don't like me

Feeling that lack of support is tough. What makes you think your boss feels this way about you?

Understanding the pain of rejection

What do you want to talk about?

Figure 1: Chat interface with additional textual feedback channel to display the system's internal empathic resonance.

1 Introduction

Conversational agents are increasingly being used for digital mental health support [9, 30, 38, 66], as they offer constant low-barrier accessibility [23] and non-judgmental anonymous interaction [10, 19, 58]. In that context, empathy is an important factor, as it supports social bonding, therapeutic alliance, and positive support outcomes, for example, through nonverbal expressions of understanding, perspective taking, or affective resonance [26, 59, 61, 74]. However, while large language models (LLMs) are capable of empathic response generation [6, 19, 54, 76, 80], most applications, such as chatbots, do not provide additional channels for empathic expressions. We aimed to address that limitation by adding a feedback channel that reflects the system's internal empathic resonance. Following related approaches [37, 81], we developed an LLM-based chatbot that generates and displays short textual descriptions alongside its conversational responses, similar to how a human interlocutor could express attention, validation, or emotional resonance through nonverbal cues. We conducted a user study ($N = 188$) to evaluate the following research question:

ICMI '25, Canberra, ACT, Australia

ACM ISBN 979-8-4007-1499-3/2025/10
https://doi.org/10.1145/3716553.3750759

RQ: How does the display of internal cognitive and affective resonance influence the perceived empathy of an LLM-based chatbot in a mental health support setting?

Our study revealed a statistically significant increase in perceived empathy (+14.7%) when displaying LLM-generated internal empathic resonance messages. Independently, we found that gender, prior experience with mental health applications, and the emotional intensity of conversations significantly influenced perceived empathy. Our work demonstrates how empathic interaction with LLM-based chatbots can be practically enhanced through a second textual channel, without the need for additional modalities. Furthermore, our approach can serve as a basis for multimodal systems by translating LLM-generated internal reflections into nonverbal feedback such as gestures or visual cues.

2 Related Work

In this section, we introduce empathy and the use of artificial agents in the context of mental health support [38, 39, 66].

2.1 Empathy in Healthcare

Empathy is important in human interaction, especially in supportive or therapeutic settings, such as physiological and psychological health [43, 59, 61], and is often defined as a multidimensional construct, divided into a cognitive and an affective dimension [8, 20, 22, 69]. Cognitive empathy involves understanding and perceiving a situation or emotional state of another person, while affective empathy refers to the emotional responses that arise as an empathic reaction [20]. In therapeutic settings, empathy is crucial for the formation of therapeutic alliance and the promotion of positive therapy outcomes [26, 28, 29, 33, 43, 62, 74, 75]. For example, empathic attunement allows a therapist to express emotional resonance, sympathy, interest, or active listening [20, 34], which can encourage clients to open up and help them feel understood [26]. Similar effects of empathic behavior can be found in informal social support [10, 46, 48], originating from conventional networks such as family, friends, and partners, but also from online sources, including social networks or artificial agents [10, 49, 60].

2.2 Empathic Agents in Digital Healthcare

Conversational agents are increasingly being used for digital healthcare [38, 39]. They offer constant availability and anonymity, which can encourage self-disclosure and frequency of use [10, 19], and unlike human therapists, they do not suffer from emotional burnout or empathy fatigue [58, 72]. Against this background, researchers are investigating the effects of empathy [12, 30, 66, 71]. They found, for example, that empathic chatbots are preferred as health advisors over non-empathic ones [21, 53], and that they can help cope with social exclusion [24] or depression [35]. Furthermore, LLM-generated health advice was perceived as more empathic than human-generated [6, 55, 80], promising further potential in medical and mental health settings [17, 70, 76]. In digital mental health support, empathy is also seen as a key factor in building a therapeutic alliance and positive user outcomes [44, 56], and for authenticity [70]. Still, researchers highlight the need for further exploration in this regard [40, 56].

2.3 Ethical Risks and Challenges

The use of conversational agents in healthcare also harbors risks and ethical concerns [12, 19, 67], for example, regarding data privacy [10, 19, 23], safety and quality of care [23, 26, 67], as systems might provide incorrect advice or have inherent biases [12, 23]. Other risks include social isolation due to emotional attachment or overdependence [10, 12] and vulnerability if, for example, a trusted system has been manipulated [19]. Consequently, accountability and access must be regulated to protect vulnerable populations, such as people with mental disorders [7, 12, 19, 23]. Cabrera et al. [12] therefore suggest that mental health agents should serve as complementary tools for emotional support, rather than replacing professional therapeutic care. In line with this recommendation, we designed our application as an informal support system, addressing ethical concerns as described in Section 5.2 and Section 7.7.

2.4 Empathic Feedback Modalities

In contrast to embodied agents such as robots [15, 45, 50, 57, 64] or virtual agents [41, 63, 65], which can express empathy through multimodal, nonverbal feedback, most unimodal LLM-based agents are simply prompted to generate empathic verbal responses [19, 30, 77, 82]. Adding additional affective and empathic feedback in text messaging includes, for example, the use of emoticons [31, 42, 51, 54], the integration of environmental or physiological context [11], or the simulation of human-like typing behavior [83]. Furthermore, visualizing emotions in messages through shapes or colors [4, 16] can enhance emotional expression, although color-emotion coding is potentially challenging to interpret. In addition to such explicit feedback, research in the fields of transparency and explainability explores how to convey reasoning processes, limitations, and biases of a system [3, 14, 47, 52, 79]. For example, confidence indicators and narrative explanations have been used to increase trust in medical support agents [3], or to justify response delays in chatbots [81]. Furthermore, general reasoning LLMs such as DeepSeek-R1 explicitly generate a textual chain of thought that reveals a kind of inner monologue [25, 78]. Göldi and Rietsche [37] explore the effects of displaying such reasoning processes by adding text messages highlighted with a brain icon to the conversation flow. They found that non-factual, belief-indicating messages such as *"I anticipate that the user will..."* or *"I'm attempting to understand how the user is"* increased users' expectation confirmation toward the system.

3 Conception

As outlined in Section 2, empathy is vital for emotional support and can improve authenticity [70], trust, and user outcomes [44, 56] when expressed by artificial agents. Therefore, we investigated how to enhance empathic interaction with a chatbot in an emotional support setting by displaying empathic resonance (RQ).

3.1 Initial Design Considerations

We explored related design spaces of explainability [27], textual [11], and nonverbal communication [5] to determine potential feedback dimensions. Following Eiband et al. [27], we divided the derived dimensions into *what* to show and *how* to show it.

What to show. Based on the dimensional definition of empathy (Section 2.1), empathic resonance could provide *cognitive* and *affective* feedback, for example, by describing internal processes of perception and understanding, and showing expressions of attention, active listening, and perspective taking, as well as emotional responses. One factor to consider is whether such empathic resonance should be pre-defined or generated dynamically, for example, by mirroring user input or through LLMs or other generative models that are capable of generating empathic output [6, 19, 55, 80]. In addition, feedback should appear authentic, especially when expressing emotions through artificial agents [40, 67, 70].

How to show it. Drawing from related work [5, 11], possible design dimensions for feedback presentation include level of abstraction, transmission modality, encoding, and temporal persistence, along with triggering events and timing. Section 2.4 illustrates several approaches for additional feedback through text, audio, colors, or shapes. Typically, such cues are activated at the same time as the primary system output and remain for a comparable duration. Textual cues are usually integrated into the verbal message flow.

3.2 Creative Group Session

In a next step, we conducted a creative group session with two authoring researchers and two domain experts with backgrounds in mobile HCI, computer-mediated messaging, and psychology (three male, one female, mean age 33.0 years ($SD = 3.7$). The session consisted of two phases, each using a combination of individual reflection and collaborative discussion.

Phase I. First, we explored the characteristics of empathic behavior in emotional support situations. We identified four key dimensions: (1) active listening and providing space, (2) understanding and perspective taking, (3) thoughtful communication, and (4) providing support. While the last two dimensions are more likely to be expressed through a system's direct verbal responses to the user, we saw active listening, understanding, and perspective taking as potential feedback content reflecting cognitive empathy.

Phase II. Second, we discussed how a system should express empathic behavior. Textual feedback was seen as neutral and understandable, although it could be perceived as too technical and disrupt the primary channel of conversation. Virtual avatars, as used in contemporary applications such as Snapchat, were regarded as easy to understand and appealing, yet also as potentially limited in expressiveness and as subject to biases based on their visual representation. In addition, the group suggested that feedback should be displayed independently of the primary conversation messages and with limited duration, to emphasize active, attentive behavior. Another suggestion was to dynamically highlight or dim different interface areas or resize message bubbles, to indicate turn-taking and focus of attention. Finally, LLM-based feedback generation was consensually preferred over pre-defined feedback.

3.3 Final Feedback Design

Based on the group discussion, we decided to implement a secondary textual channel to display the agent's internal empathic resonance. With this choice, our first objective was to avoid the comprehension problems inherent in approaches such as emotion color coding [16]. Second, we argue that text allows dynamically creating cognitive and affective expressions without requiring pre-designed content. Third, we assumed that maintaining the agent's default modality would strengthen the feedback-agent association and allow us to create the text output using the underlying LLM itself. In detail, we derived the following design features. **Modality:** textual feedback; short and understandable. **Content:** emotional and cognitive resonance primarily expressing understanding. **Source:** LLM-generated in real-time, based on the user input. **Trigger:** after users sent a request; before verbal response. **Position:** separated from conversational messages to highlight separate nature and draw focus. **Persistence:** show feedback text until the next user input. **Animation:** typewriter-like animation to draw attention, slightly pulsating icon to indicate activity.

4 Implementation

We implemented a web application that allowed LLM-based chat interaction, consisting of a Vue.js-based frontend, and a Python-based backend API for message handling and logging.

4.1 Chat Interface

Figure 1 shows the task view of our application. User messages were displayed on the right, and agent replies on the left. The top bar offered a short version of the task description, access to safety hints, an interaction time counter, and a continue button. The bottom area contained message input controls, and, for group B, an additional text field displaying the system's empathic resonance. Otherwise, the design for group A was identical. In addition, the application included views for on-boarding, instructions, and the survey.

4.2 Resonance Display

To display internal empathic resonance, we applied the design features as defined in Section 3.3. In both groups, a pulsating icon indicated system activity and readiness. In group B, the additional text field displayed internal system states through short, LLM-generated texts. To increase attention and emphasize system activity, these texts were animated at a rate of 30ms per character. The text was reset to *"…"* whenever the user sent a new message, indicating a processing state. After the LLM had completed a request, the two generated response parts were displayed one after the other. The internal resonance text was first displayed before the explicit system response appeared at the top of the actual chat history. Due to the text animation and the limited text length, this implementation caused a delay of about one second between the display of the internal resonance and the explicit response.

4.3 Response and Resonance Generation

To generate agent responses and empathic resonance texts, we used GPT-4o-mini (gpt-4o-mini-2024-07-18, temperature=1, no completion token limit). The user input messages were passed to the backend, which in turn passed them to OpenAI's completion API. Each API request contained the complete conversation history of a user session, including the pre-prompt below. For prompt design, we conducted several test iterations and generally followed the OpenAI guidelines and design recommendation by Vogel [73].

Prompt: System Instructions

```
Your role: act as an empathic and reflective chatbot, helping users explore and
understand a challenging interpersonal situation.
For each message, follow these steps:
1. inner thoughts
- Reflect your understanding and emotional resonance to what the user shares
- Do write in first-person perspective ("I feel..." "This makes me...")
- Do not address the user directly
- Keep it short. Maximum 6 words.
- Format as: <thoughts>That makes me sad</thoughts>
2. create response:
- Mirror understanding and validate feelings
- Keep responses concise (2-3 sentences) to maintain dialogue.
- Ask thoughtful follow-up questions sparingly
- Focus on helping users gain their own insights
- Format as <response>Your 2-3 sentence response</response>
3. structure
- combine inner thoughts and response in your output like this
- example: <thoughts> Feeling their uncertainty and worry </thoughts> <response>
Have you noticed any changes in your friend's general behavior lately? </response>
```

The above prompt generated the system's empathic resonance message as well as the actual verbal response to the user. These two components were parsed and displayed as described in Section 4.2. This step was identical for both groups, with the only difference that the empathic resonance was not displayed for group A.

5 User Study

For evaluation, we conducted a user study with $N = 188$ participants, divided into two conditions (A, B). The study was designed to focus on genuine experiences, in which participants were asked to talk about challenging interpersonal situations they had encountered. Based on our RQ, we applied the Perceived Empathy of Technology Scale (PETS) [69] to test the following hypotheses:

H1a Displaying the agent's internal empathic resonance increases the overall perceived empathy.

H1b Displaying the agent's internal empathic resonance increases the perceived emotional responsiveness.

H1c Displaying the agent's internal empathic resonance increases perceived understanding and trust.

5.1 Procedure

We recruited participants through Prolific with our system securely hosted online, for easy access via personal devices. Participation was voluntary and could be stopped at any time. Participants were randomly assigned to one of two groups: Group A, which provided classic text interaction, and Group B, which provided additional system resonance (Section 4.2). After explaining study details, we obtained informed consent. Then, all participants completed the conversation task, followed by a survey assessing system perception. We estimated the completion time to be approximately 10 minutes, and compensated participants with 1.90£.

Conversation Task. Both groups received identical task instructions: *"Your task is to engage in a 5-10 minutes conversation (in English) with ELLMO, our AI assistant. Talk about a challenging interpersonal situation you've experienced with someone. You could for example talk about: a disagreement, communication difficulties, situations at work, in public, or any other interpersonal challenge."* Further, we provided guidance on how to start the discussion and briefly explained the agent's supportive objective. The complete task instructions are included in the *Supplementary Material*.

Survey. After the conversation, we applied the ten PETS items as recommended, as 101-point interactive sliders from *strongly disagree* to *strongly agree* in randomized order [69]. Additionally, we asked participants to describe the emotionality of the discussed topic (*"The experience I discussed with ELLMO was emotionally challenging to me."*) and on how their usage behavior of artificial agents (*"I use AI assistants like ChatGPT for personal tasks."*, *"I use AI assistants like ChatGPT for work-related tasks."*) and mental health applications (*"I use digital tools or apps for mental health and well-being support."*). As an attention check, another item asked to drag a slider completely to the left. These items were also displayed in randomized order and with 101-point sliders.

5.2 Ethics, Safety & Crisis Support Features

Particularly for research focused on mental health support, safety and ethical considerations are crucial [12, 19, 23, 68]. For transparency, we provided participants with comprehensive details about the study's objectives, methods, and data processing (Section 5.1). We emphasized that AI is not a substitute for professional help and encouraged users to seek appropriate support if needed. To this end, we integrated crisis support information into the task view, including several national helpline numbers and links to mental health resources. To minimize the inclusion of high-risk groups, we applied Prolific's pre-screening filters and excluded participants who positively answered the question *"Do you have any diagnosed mental health condition that is uncontrolled (by medication or intervention) and which has a significant impact on your daily life/activities?"*. We also relied on OpenAI's built-in safety features, which guide the model to handle sensitive issues ethically, for example, by refusing to engage and suggesting professional help [19]. In addition, our admin interface allowed us to observe and reach out to participants via Prolific when necessary. Furthermore, the study was approved by our ethics committee (EK-MIS-2025-0346-FT-d01).

5.3 Sample Size

In total, we recruited 192 participants, from which we rejected four due to failed attention checks, leaving 188 valid sessions. To determine sample size, we performed an a priori power analysis using G*Power [32] for a significance level of $p = .05$ and a statistical power of = 0.80. The results indicated a minimum sample size of $N = 128$ for a medium effect size with a fixed effects ANOVA ($d = .25$) or two-tailed t-test ($d = 0.50$). For the nonparametric Mann-Whitney U test with effect size ($d = 0.05$), the required sample size was $N = 134$. With $N = 188$ participants, we met these requirements to test for small to medium effects. Our sample size also met HCI standards [13] and exceeded the average size of related studies for chatbots in mental health ($N = 75.2$ for 53 studies) [1].

5.4 Participants

The mean age of participants was 35.1 years ($SD = 11.1$), with 91 identifying as female, 96 identifying as male, and one preferring not to provide information. We recruited participants from 20 different countries, with the majority of 163 participants residing in the European Economic Area, 19 in North America, five in the Asia-Pacific region, and one not providing that information. We verified participants' fluency in English through pre-screening and post hoc review of the conversations. In addition to excluding users with severe mental illness (Section 5.2), we used the pre-screening information provided via Prolific to investigate possible associations with anxiety and depression, which in general has been found to interact with empathy in that context [2, 35, 36]. 120 participants reported general experiences of anxiety (66 did not, two preferred not to share), and 82 participants indicated they had experiences of depression (100 did not, six preferred not to share).

6 Results

We analyzed the data from $N = 188$ valid sessions, which on average lasted 11.3 minutes ($SD = 4.6$). As individual group data were partially non-normally distributed (Shapiro-Wilk $p < 0.05$), we tested significance using non-parametric Mann-Whitney U tests. The detailed statistical results are in Table 1.

6.1 Conversational Engagement

Overall, users sent a median of 10.0 messages per conversation ($IQR = 8.0-14.0$). The median task completion time was 9.2 minutes ($IQR = 6.4 - 10.7$). Task duration and number of sent messages did not statistically significant differ between groups. To estimate user engagement in a conversation, we also analyzed the number of user messages, characters per message (CPM), and words per message (WPM). As shown in Table 1, group B showed minor yet non-significant effects of slightly increased message count and CPM. Participants also rated how emotionally challenging they assessed the topic they discussed. As shown in Table 1 (Emo. Level), there was no significant difference between groups, indicating that participants chose similar topics in both conditions.

6.2 Usage of Technology

We further assessed how often participants used AI assistants for personal and work-related tasks (Table 1). There was no statistical significance between groups, yet overall, participants reported a high level of usage. In addition, we analyzed the general use of digital tools or applications for mental health and wellbeing support. Ratings for group B were slightly higher ($Mdn = 27.0$, $IQR = 12.2 - 68.0$) compared to group A ($Mdn = 22.5$, $IQR = 0.0 - 53.8$) yet showed only marginal statistical significance.

6.3 System Responses

System response generation took a median of 1.5 seconds ($IQR = 1.2 - 1.9$), without showing significant differences between groups. Conversational system responses contained a median of 29 words per message for both groups, and empathic resonance messages a median of five words per message. Length and word count were not statistically significant, suggesting a consistent LLM response generation over both groups. Again, as described in Section 4.3, empathic resonance messages were generated for both groups, yet not displayed in group A. In a qualitative analysis using AI-based text-categorization, we reviewed the 2195 empathic resonance messages and found that most could be seen as expressions of understanding and validation (for example *"Longing for connection is painful."*, *"Perhaps exploring collaboration is insightful"*, *"Feeling betrayed can be isolating"*). A large part of the messages also contained emotional expressions, which could be either interpreted as the system's emotional resonance (for example *"That makes me feel frustrated"*, *"Feeling admiration for her journey"*) or as recognition or reflection of the user's emotional expressions (*"Feeling a sense of relief"*, *"Feeling their confusion and curiosity"*, *"Feeling overwhelmed and in need of support"*). Finally, although we intended the system to not directly address the user (see the prompt in Section 4.3), we found that in 11.4% of the resonance texts, the system did nevertheless, for example, *"Feeling your frustration and helplessness"* or *"Feeling inspired by your goal"*. All generated system resonance messages can be found in the *Supplementary Material*.

Table 1: Comparison of groups A (default) and B (additional system feedback) regarding perceived empathy (PETS), number of user and system messages, words (WPM) and characters per message (CPM), task duration [min], response time [sec], emotional level, and use of AI at work, in personal context as well as use of mental health applications (MH Apps).

	Group A		Group B		Statistics			
	Mdn	IQR	Mdn	IQR	Z	r	p	
Perceived Empathy								
PETS	64.5	54.5-80.8	74.0	59.8-86.0	−2.29	0.17	.022	*
PETS-ER	66.1	55.0-80.6	75.1	62.2-84.5	−2.52	0.18	.012	*
PETS-UT	67.1	54.1-82.4	73.8	57.1-90.6	−1.99	0.15	.046	*
Chat (User)								
Messages	10.0	8.0-14.0	11.0	8.2-14.0	−1.13	0.08	.256	
WPM	14.0	9.0-19.8	14.0	10.2-22.0	−0.88	0.06	.378	
CPM	75.0	49.0-104.5	74.5	54.2-117.5	−0.88	0.06	.379	
Task Dur.	8.2	6.4-10.7	9.6	6.4-10.7	−0.32	0.02	.750	
Emo. Level	64.5	40.0-78.8	64.5	33.0-85.0	−0.29	0.02	.770	
Chat (System)								
Responses	10.0	8.0-14.0	11.0	8.2-14.0	−1.12	0.08	.261	
WPM	29.0	25.0-35.0	28.5	24.0-34.8	0.54	0.04	.588	
CPM	181.0	156.2-214.8	179.0	151.0-212.8	0.49	0.04	.624	
Resp. Time	1.5	1.2-1.8	1.6	1.2-2.0	−0.78	0.06	.436	
Resonance (System)								
Messages	10.0	8.0-14.0	11.0	8.2-14.0	−1.12	0.08	.261	
WPM	5.0	5.0-5.0	5.0	5.0-5.0	−0.68	0.05	.411	
CPM	36.0	34.0-38.0	36.5	34.0-38.8	−1.45	0.11	.145	
Use of Technology								
AI at Work	75.0	53.2-93.0	71.5	20.0-96.8	0.66	0.05	.511	
AI Personal	72.5	32.0-92.0	70.0	23.0-92.0	0.13	0.01	.898	
MH Apps	22.5	0.0-53.8	27.0	12.2-68.0	−1.84	0.13	.065	

Statistics based on Mann-Whitney U tests. Ratings from [0..100], $N = 94$ per group.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

6.4 Perceived Empathy Rating

We calculated perceived empathy for overall PETS and separately for its two subscales, Emotional Responsiveness (PETS-ER) and Understanding and Trust (PETS-UT) [69]. As shown in Table 1, non-parametric Mann-Whitney U testing revealed statistically significant differences between group A and B for overall PETS ($p = .022$) and the subscales PETS-ER ($p = .012$) and PETS-UT ($p = .046$), with median increases of 14.7% for PETS, 13.6% for PETS-ER, and 10.0% for PETS-UT. In addition, we investigated how the other variables (gender, technology use, anxiety and depression experience, emotionality) interact with perceived empathy. For that, we converted continuous rating variables like technology use (rated 0..100) into binary categories (low: < 50, high: ≥ 50). Participants with missing data were excluded from these analyses: one participant who did not want to specify their gender, two participants

Table 2: PETS analysis for individual variables viewed as two-level categories: gender (m/f), level of emotionality, use of AI at work, in personal context, use of mental health apps (low / high), experiences of anxiety and depression (yes / no).

	Sample 1		Sample 2		Statistics			
	Mdn	**IQR**	**Mdn**	**IQR**	**Z**	**r**	**p**	
Gender	*male (N=96)*		*female (N=91)*					
PETS	65.3	53.2-79.2	73.8	61.8-86.3	-2.64	0.19	.008	**
PETS-ER	66.4	53.2-80.4	74.0	63.5-86.1	-2.75	0.20	.006	**
PETS-UT	67.4	53.9-82.8	77.0	58.9-87.8	-2.13	0.16	.033	*
Emo. Level	*low (N=67)*		*high (N=121)*					
PETS	62.5	52.4-77.8	72.4	61.8-86.9	3.22	0.23	.001	**
PETS-ER	65.0	51.9-76.7	73.3	62.0-84.3	2.93	0.21	.003	**
PETS-UT	59.8	53.8-76.6	75.0	58.8-91.5	3.28	0.24	.001	**
AI at Work	*low use (N=54)*		*high use (N=134)*					
PETS	74.7	62.2-86.7	68.1	55.5-84.2	-1.51	0.11	.132	
PETS-ER	75.8	61.3-85.8	68.7	55.7-82.4	-1.60	0.12	.111	
PETS-UT	74.5	59.1-85.2	69.6	55.7-87.1	-1.30	0.09	.193	
AI Personal	*low use (N=58)*		*high use (N=130)*					
PETS	68.7	54.6-79.0	71.8	57.0-85.7	1.09	0.08	.276	
PETS-ER	69.2	55.9-83.4	72.2	57.9-83.2	0.75	0.05	.455	
PETS-UT	69.0	54.2-79.7	72.8	56.8-88.8	1.42	0.10	.155	
MH Apps	*low use (N=125)*		*high use (N=63)*					
PETS	68.3	55.0-79.2	79.5	61.5-88.0	2.83	0.21	.005	**
PETS-ER	67.8	55.2-80.8	79.3	60.2-88.8	2.75	0.20	.006	**
PETS-UT	68.5	54.8-82.8	79.2	57.5-92.4	2.80	0.20	.005	**
Anxiety	*no anxiety (N=66)*		*anxiety (N=120)*					
PETS	72.4	59.1-86.2	68.6	55.6-84.1	-1.27	0.09	.203	
PETS-ER	74.6	58.9-84.2	67.3	57.0-82.7	-1.47	0.11	.141	
PETS-UT	71.8	57.9-87.9	70.1	54.7-86.6	-0.89	0.07	.376	
Depression	*no depression (N=100)*		*depression (N=82)*					
PETS	71.8	56.8-84.2	69.5	57.7-85.8	-0.05	0.00	.958	
PETS-ER	72.1	58.0-82.9	70.2	59.0-83.7	-0.04	0.00	.970	
PETS-UT	70.6	56.9-85.8	71.4	55.8-87.8	0.00	0.00	1.000	

Statistics based on Mann-Whitney U tests. PETS Ratings from [0..100]
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

who did not provide information about their general experiences of anxiety, and six participants who did not provide information about their experiences of depression. As shown in Table 2, we found significant gender effects, with female participants reporting higher perceived empathy (median +13.0%), overall, and for PETS-ER and PETS-UT. Furthermore, with digital mental health app experience reported significantly higher empathy scores than non-users (median +16.4%). The reported level of emotionality in conversations also significantly influenced PETS, PETS-ER, and PETS-UT when categorized as low and high, with high emotionality (score ≥ 50) leading to increased perceived empathy (median +15.8%). The other variables did not show statistically significant effects.

To test whether the differences in the use of mental health applications, gender effects, and emotionality actually explain differences between groups, we conducted Aligned Rank Transform (ART) ANOVAs and examined interaction effects (Table 3). We found that taking these factors into account when comparing perceived empathy between groups A and B did not render the results insignificant. Also, no significant interaction effects were found between group and mental health application usage (all $p >= 0.857$), between group and gender (all $p >= 0.311$), and between group and emotionality (all $p >= 0.745$), indicating that the benefits of empathic resonance feedback were independent of these factors. Thus, overall we consider **H1**, **H2** and **H3** confirmed.

Table 3: Aligned Rank Transform ANOVA, testing for effects of gender, mental health app usage and level of emotionality in combination with study groups on PETS and its subscales PETS-ER and PETS-UT, as well as interaction effects (x).

PETS	**df1**	**df2**	**F**	**p**	
Gender (m/f)	1	183	7.76	.006	**
Group (A/B)	1	183	5.92	.016	*
Group x Gender	1	183	.88	.350	
MH App Usage (low/high)	1	184	8.86	.003	**
Group (A/B)	1	184	5.65	.019	*
Group x MH App Usage	1	184	.00	.956	
Emo. Level (low/high)	1	184	1.73	.001	**
Group (A/B)	1	184	5.25	.023	*
Group x Emo. Level	1	184	.00	.996	
PETS-ER	**df1**	**df2**	**F**	**p**	
Gender (m/f)	1	183	8.18	.005	**
Group (A/B)	1	183	6.73	.010	*
Group x Gender	1	183	1.03	.311	
MH App Usage (low/high)	1	184	8.26	.005	**
Group (A/B)	1	184	6.78	.010	**
Group x MH App Usage	1	184	.01	.913	
Emo. Level (low/high)	1	184	9.10	.003	**
Group (A/B)	1	184	6.83	.010	**
Group x Emo. Level	1	184	.11	.745	
PETS-UT	**df1**	**df2**	**F**	**p**	
Gender (m/f)	1	183	5.08	.025	*
Group (A/B)	1	183	4.33	.039	*
Group x Gender	1	183	1.00	.319	
MH App Usage (low/high)	1	184	8.16	.005	**
Group (A/B)	1	184	4.20	.042	*
Group x MH App Usage	1	184	.03	.857	
Emo. Level (low/high)	1	184	1.71	.001	**
Group (A/B)	1	184	3.99	.047	*
Group x Emo. Level	1	184	.08	.779	

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

7 Discussion

In this work, we explored how displaying cognitive and affective resonance through a secondary textual feedback channel can enhance the perceived empathy of an LLM-based chatbot (**RQ**). We hypothesized that such feedback would increase overall perceived empathy (**H1**), perceived emotional responsiveness (**H2**), and perceived understanding and trust (**H3**). We confirmed these hypotheses in Section 6.4. In this section, we further discuss the results, implications, and limitations.

7.1 Effects on Perceived Empathy

The main effect we observed was that the display of internal empathic resonance led to a statistically significant increase in perceived empathy (median +14.7%). In general, this is consistent with the findings of Göldi and Rietsche [37] and Zhang et al. [81], who showed that displaying textual feedback of internal system states can enhance system perception. The stronger effects observed in perceived emotional responsiveness (PETS-ER) suggest that displaying internal resonance may have particular potential to enhance system perception of perceived emotional intelligence, sympathy, and emotional support, possibly facilitating the establishment of affective bonds between the user and the system. Therefore, we suggest to consider this form of feedback in future research on human-AI relationships in mental health, in informal settings, and in the context of the digital therapeutic alliance [56].

7.2 Effects on Engagement

In contrast to the effects on perceived empathy, we did not observe significant differences in message length or interaction duration. This suggests that the effects of displaying empathic resonance may be primarily qualitative and do not have a direct impact on the conversation flow. Future systems could therefore benefit from the integration of such feedback without having to worry about distracting the user or interrupting established conversation patterns. This could be particularly important when extending existing systems with additional feedback channels.

7.3 Effects of Gender, Use of Technology

We found notable influences of gender and prior experience with mental health applications on perceived empathy (Section 6.4). The higher empathy scores for female participants (median +13.0%) potentially align with related research on individual differences in receptivity to empathic feedback [18]. While we suggest investigating these patterns further, particularly regarding how findings from human interaction transfer to human-AI interaction, we acknowledge that unobserved factors such as social or cultural background and personality traits might also affect this measure. Beyond gender, prior experience using mental health apps positively affected perceived empathy (+16.4%). We argue that this might reflect general acceptance or higher engagement with systems in that context. We recommend exploring this effect further, especially regarding how longitudinal app usage might impact emotional attachment, bearing both risks and potential beneficial support effects.

7.4 Effects of Emotional Intensity of Context

While we found no differences between groups in how emotionally participants rated the topic of conversation, we found significant effects of this emotionality rating on perceived empathy (Section 6.4). Participants who classified their topic of conversation as highly emotionally challenging had a PETS score that was 15.8% higher than others. The fact that this increase did not interact with the increased PETS ratings between groups highlights that our approach can enhance empathetic experiences regardless of emotional context. Furthermore, the relationship between emotionality and perceived empathy suggests that mental health applications should be particularly carefully designed for highly emotional conversations, where users appear to be more sensitive to empathetic feedback, both to benefit from empathic interaction and with regard to ethical concerns to avoid misusing that sensitivity. Future systems could assess the emotionality of a topic in advance and during a conversation to adapt, for example, feedback content or modalities.

7.5 Future Internal Resonance Design

LLM-based generation of the internal resonance messages provided flexibility and contextual reference, but also brought challenges. Despite explicit instructions to avoid direct user address, 11.4% of the messages contained second-person language (e.g., "Feeling your frustration"). Future work should address the trade-off between richness of expression and reliability in generating resonance messages, especially with emerging reasoning LLMs that provide chain-of-thought processes. Another point to consider is feedback authenticity [40, 67, 70]. For example, we assume that simply reflecting or mirroring affective states such as *"Feeling hesitation and anxiety"* does appear inauthentic, as it expresses understanding and perception rather than affective experience. Future studies could explicitly investigate how different resonance design affects the perceived authenticity of a system. Finally, an important design decision is the choice of modality. While we chose an additional textual feedback channel, in future research, we plan to assess visual or ambient feedback mechanisms, for example, based on color or shapes (Section 2.4). These could include subtle visualizations reflecting emotional resonance or turn-taking (Section 3.3) that persist throughout the interaction, similar to how nonverbal cues function in human communication. Also, LLM-generated resonance messages could serve as an intermediate layer for generating multimodal feedback, for example, as input for text-to-image models.

7.6 Implications for Mental Health Applications

For informal consumer-facing mental health applications, such as chatbots, the inclusion of an empathic resonance channel could improve user perception without the need for significant redesign. In addition, textual feedback would allow easy configuration and customization, which is particularly important since the preference for affective expressions in mental health applications is highly individualized [38]. In therapeutic contexts, empathic feedback should be designed with particular care in terms of accuracy, therapeutic strategies, and integrity. For example, it could focus primarily on expressions of acknowledgment and validation to avoid communicating misinterpreted affective states. In addition, the evaluation of our approach in this context should be done with controlled studies that specifically measure clinical effects, such as therapeutic outcomes or therapeutic alliance. In line with Cabrera et al. [12], we emphasize that digital mental health applications should enhance, not replace, human support networks and professional care.

7.7 Ethical Implications

While our study demonstrated the potential of our approach, we acknowledge the ethical dimensions of simulating empathy in artificial systems. Showing internal empathic resonance might lead to an overestimation of a system's emotional intelligence. Another more general risk is overreliance on technology [10, 12, 19], as particularly vulnerable populations could be susceptible to anthropomorphizing systems and developing emotional attachments [19]. To address these risks, we recommend: (1) clearly framing of empathic resonance as a design feature rather than evidence of system consciousness; (2) careful language in resonance displays that avoids claims of actual emotional states; (3) periodic reminders of the artificial nature of the system during extended interactions; and (4) ensuring access to human support resources, especially when serving vulnerable populations. Again, we highlight that artificial agents should not serve as a substitute for professional care and that further ethical dimensions such as regulation or liability need to be considered [7, 12, 19].

7.8 Limitations

Our study has several limitations that should be considered when interpreting the results and planning future research.

Remote Study & Task Design. We hypothesize that our remote study design encouraged authentic conversations due to its anonymity. Furthermore, we specifically allowed users to pick a topic themselves to foster high authenticity of the emotional engagement. However, this also limited our control over the course of the conversation and may have led to diversity in conversations that could have influenced how the system was perceived.

Sample Composition. Our participant sample consisted predominantly of Western European users, which limits the generalizability of our findings across cultures. As the perception of empathy can vary greatly in different cultural contexts, future work should make cross-cultural comparisons. In addition, we assessed the experiences with anxiety and depression using a pre-screening binary self-report rather than validated clinical measures. We suggest using more detailed, standardized screening measures in future approaches and focusing on participants' motivation to seek support.

System Perception. Although we assessed perceived empathy, we lack direct insight into how participants interpreted the displayed empathic resonance. For example, we do not know whether the feedback was interpreted as authentic, whether it was continuously perceived, or whether it was understood as a representation of internal system states. Future research could investigate users' understanding and perception in this regard.

Interaction Duration. The short interaction duration (median 9.2 minutes) may not have revealed possible effects of empathic resonance that occur in repeated or longer interactions with a digital agent. We therefore intend to investigate extended usage patterns in longer, more realistic study scenarios, possibly focusing on daily conversations over several weeks.

Metrics. Future research should include metrics that specifically address concepts such as digital therapeutic alliance or social bonding. This would allow to determine the potential effects of increased perceived empathy on users' goals. Furthermore, we did not directly compare our approach to existing commercial mental health applications. A comparative analysis with established applications would provide more context for understanding the relative impact of our approach compared to current best practices in the field.

Statistical Analysis. While our results showed statistical significance, the small effect sizes ($r = 0.15 - 0.18$) in Table 1 suggest that the practical impact of the empathic resonance channel may be modest. To achieve stronger effects, future iterations could, for example, adjust the content, frequency, or verbal style of the generated resonance messages or implement multimodal feedback (Section 7.5). Finally, we acknowledge that testing multiple dependent variables without correction increases Type I error risk. At the 95% confidence level, approximately one false positive might be expected per 20 tests. However, the consistency of effects across our three empathy measures (PETS, PETS-ER, PETS-UT) suggests that even if some results were false positives, the evidence still supports increased perceived empathy through the empathic resonance channel.

8 Conclusion

We implemented an LLM-based emotional support chatbot that displays internal empathic resonance on a second text channel alongside the primary conversational messages. In our user study ($N = 188$), displaying internal empathic resonance led to a significant increase in perceived agent empathy (median +14.7%), confirming our hypotheses for RQ (Section 1). We also showed that gender, prior experience with mental health applications, and the emotional intensity of conversational context had significant independent effects on empathy perception, with higher empathy scores among female participants (+13.0%), regular users of mental health apps (+16.4%), and emotionally charged conversations (+15.8%). Importantly, none of these factors moderated the primary effect of displaying internal empathic resonance.

We conclude that the display of LLM-generated internal empathic resonance can increase perceived empathy of a system and thus potentially contribute to improving the user experience and outcomes of digital mental health applications. Our approach provides a practical way to enhance existing unimodal chatbot applications and a foundation for future LLM-based multimodal systems.

Safe and Responsible Innovation Statement

In Section 2.3, we acknowledge the social and ethical risks of applying AI in the context of mental health support. To mitigate risks for our study participants, we implemented safeguards as described in Section 5.2. Our study design also ensured anonymous participation, transparency, and informed consent regarding data processing and was approved by our institutional ethics committee. Finally, we address the ethical implications in Section 7.7, emphasizing that digital agents should rather enhance and not replace professional mental health support, particularly for vulnerable users.

Author Contributions

Matthias Schmidmaier: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Jonathan Rupp**: Conceptualization, Formal Analysis, Methodology, Validation, Writing – original draft; **Sven Mayer**: Conceptualization, Funding acquisition, Resources, Supervision, Writing – original draft, Writing – review & editing.

Acknowledgments

This work has been partly supported by the Research Center Trustworthy Data Science and Security (https://rc-trust.ai), one of the Research Alliance centers within the UA Ruhr (https://uaruhr.de).

References

[1] Alaa A Abd-alrazaq, Mohannad Alajlani, Ali Abdallah Alalwan, Bridgette M Bewick, Peter Gardner, and Mowafa Househ. 2019. An overview of the features of chatbots in mental health: A scoping review. *Int. J. Med. Inform.* 132 (dec 2019), 103978. https://doi.org/10.1016/j.ijmedinf.2019.103978

[2] Arfan Ahmed, Asmaa Hassan, Sarah Aziz, Alaa A Abd-Alrazaq, Nashva Ali, Mahmood Alzubaidi, Dena Al-Thani, Bushra Elhusein, Mohamed Ali Siddig, Maram Ahmed, and Mowafa Househ. 2023. Chatbot features for anxiety and depression: A scoping review. *Health Informatics J.* 29, 1 (jan 2023), 14604582221146719. https://doi.org/10.1177/14604582221146719

[3] A Ali. 2024. Improving trust-building through more transparent conversational agent communication, in the context of medical decision support. https://essay.utwente.nl/101565/1/Ali_BA_BIT.pdf

[4] Toshiki Aoki, Rintaro Chujo, Katsufumi Matsui, Saemi Choi, and Ari Hautasaari. 2022. EmoBalloon - Conveying Emotional Arousal in Text Chats with Speech Balloons. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 527, 16 pages. https://doi.org/10.1145/3491102.3501920

[5] Michael Argyle. 1988. *Bodily Communication, 2nd Edition*. Vol. 2. Methuen & Co Ltd, New York, NY, US. 363 pages. https://doi.org/10.4324/9780203753835

[6] John W Ayers, Adam Poliak, Mark Dredze, Eric C Leas, Zechariah Zhu, Jessica B Kelley, Dennis J Faix, Aaron M Goodman, Christopher A Longhurst, Michael Hogarth, and Davey M Smith. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern. Med.* 183, 6 (June 2023), 589–596. https://doi.org/10.1001/jamainternmed.2023.1838

[7] Luke Balcombe. 2023. AI chatbots in digital mental health. *Informatics (MDPI)* 10, 4 (Oct. 2023), 82. https://doi.org/10.3390/informatics10040082

[8] C. Daniel Batson. 2009. These things called empathy: Eight related but distinct phenomena. *The social neuroscience of empathy.* 255 (2009), 3–15. https://doi.org/10.7551/mitpress/9780262012973.003.0002

[9] Eliane M Boucher, Nicole R Harake, Haley E Ward, Sarah Elizabeth Stoeckl, Junielly Vargas, Jared Minkel, Acacia C Parks, and Ran Zilca. 2021. Artificially intelligent chatbots in digital mental health interventions: a review. *Expert Rev. Med. Devices* 18, sup1 (Dec. 2021), 37–49. https://doi.org/10.1080/17434440.2021.2013200

[10] Petter Bae Brandtzæg, Marita Skjuve, Kim Kristoffer Kristoffer Dysthe, and Asbjørn Følstad. 2021. When the Social Becomes Non-Human: Young People's Perception of Social Support in Chatbots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 257, 13 pages. https://doi.org/10.1145/3411764.3445318

[11] Daniel Buschek, Mariam Hassib, and Florian Alt. 2018. Personal Mobile Messaging in Context: Chat Augmentations for Expressiveness and Awareness. *ACM Trans. Comput. -Hum. Interact.* 25, 4 (aug 2018), 23:1–23:33. https://doi.org/10.1145/3201404

[12] Johana Cabrera, M Soledad Loyola, Irene Magaña, and Rodrigo Rojas. 2023. Ethical dilemmas, mental health, artificial intelligence, and LLM-based chatbots. In *Lecture Notes in Computer Science*. Springer Nature Switzerland, Cham, 313–326. https://doi.org/10.1007/978-3-031-34960-7_22

[13] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 981–992. https://doi.org/10.1145/2858036.2858498

[14] Erik Cambria, Lorenzo Malandri, Fabio Mercorio, Navid Nobani, and Andrea Seveso. 2024. XAI meets LLMs: A survey of the relation between explainable AI and Large Language Models. https://doi.org/10.48550/arXiv.2407.15248

[15] Laurianne Charrier, Alexandre Galdeano, Amélie Cordier, and Mathieu Lefort. 2018. Empathy display influence on Human-Robot Interactions: A pilot study. In *Workshop on Towards Intelligent Social Robots: From Naive Robots to Robot Sapiens at the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018)*. IEEE, New York, NY, USA, 7. https://hal.science/hal-01887075

[16] Qinyue Chen, Yuchun Yan, and Hyeon-Jeong Suk. 2021. Bubble Coloring to Visualize the Speech Emotion. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411763.3451698

[17] Siyuan Chen, Mengyue Wu, Kenny Q. Zhu, Kunyao Lan, Zhiling Zhang, and Lyuchun Cui. 2023. LLM-empowered Chatbots for Psychiatrist and Patient Simulation: Application and Evaluation. https://doi.org/10.48550/arXiv.2305.13614 arXiv:2305.13614

[18] Leonardo Christov-Moore, Elizabeth A Simpson, Gino Coudé, Kristina Grigaityte, Marco Iacoboni, and Pier Francesco Ferrari. 2014. Empathy: gender effects in brain and behavior. *Neurosci. Biobehav. Rev.* 46 Pt 4 (Oct. 2014), 604–627. https://doi.org/10.1016/j.neubiorev.2014.09.001

[19] Andrea Cuadra, Maria Wang, Lynn Andrea Stein, Malte F. Jung, Nicola Dell, Deborah Estrin, and James A. Landay. 2024. The Illusion of Empathy? Notes on Displays of Emotion in Human-Computer Interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 446, 18 pages. https://doi.org/10.1145/3613904.3642336

[20] Benjamin M. P. Cuff, Sarah J. Brown, Laura Taylor, and Douglas J. Howat. 2016. Empathy: A Review of the Concept. *Emot. Rev.* 8, 2 (apr 2016), 144–153. https://doi.org/10.1177/1754073914558466

[21] Karl Daher, Jacky Casas, Omar Abou Khaled, and Elena Mugellini. 2020. Empathic Chatbot Response for Medical Assistance. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3383652.3423864

[22] Mark H. Davis. 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology* 44, 1 (1983), 113. https://doi.org/10.1037/0022-3514.44.1.113

[23] Munmun De Choudhury, Sachin R Pendse, and Neha Kumar. 2023. Benefits and harms of large language models in digital mental health. https://doi.org/10.48550/arXiv.2311.14693

[24] Mauro de Gennaro, Eva G. Krumhuber, and Gale Lucas. 2019. Effectiveness of an Empathic Chatbot in Combating Adverse Effects of Social Exclusion on Mood. *Front. Psychol.* 10 (2019), 3061. https://doi.org/10.3389/fpsyg.2019.03061

[25] DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z F Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, Bing Xue, Bingxuan Wang, Bochao Wu, Bei Feng, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao, Guanting Chen, Guowei Li, H Zhang, Han Bao, Hanwei Xu, Haocheng Wang, Honghui Ding, Huajian Xin, Huazuo Gao, Hui Qu, Hui Li, Jianzhong Guo, Jiashi Li, Jiawei Wang, Jingchang Chen, Jingyang Yuan, Junjie Qiu, Junlong Li, J L Cai, Jiaqi Ni, Jian Liang, Jin Chen, Kai Dong, Kai Hu, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang, Liang Zhao, Litong Wang, Liyue Zhang, Lei Xu, Leyi Xia, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Meng Li, Miaojun Wang, Mingming Li, Ning Tian, Panpan Huang, Peng Zhang, Qiancheng Wang, Qinyu Chen, Qiushi Du, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, R J Chen, R L Jin, Ruyi Chen, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shengfeng Ye, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan, S S Li, Shuang Zhou, Shaoqing Wu, Shengfeng Ye, Tao Yun, Tian Pei, Tianyu Sun, T Wang, Wangding Zeng, Wanjia Zhao, Wen Liu, Wenfeng Liang, Wenjun Gao, Wenqin Yu, Wentao Zhang, W L Xiao, Wei An, Xiaodong Liu, Xiaohan Wang, Xiaokang Chen, Xiaotao Nie, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xinyu Yang, Xinyuan Li, Xuecheng Su, Xuheng Lin, X Q Li, Xiangyue Jin, Xiaojin Shen, Xiaosha Chen, Xiaowen Sun, Xiaoxiang Wang, Xinnan Song, Xinyi Zhou, Xianzu Wang, Xinxia Shan, Y K Li, Y Q Wang, Y X Wei, Yang Zhang, Yanhong Xu, Yao Li, Yao Zhao, Yaofeng Sun, Yaohui Wang, Yi Yu, Yichao Zhang, Yifan Shi, Yiliang Xiong, Ying He, Yishi Piao, Yisong Wang, Yixuan Tan, Yiyang Ma, Yiyuan Liu, Yongqiang Guo, Yuan Ou, Yuduan Wang, Yue Gong, Yuheng Zou, Yujia He, Yunfan Xiong, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuyang Zhou, Y X Zhu, Yanhong Xu, Yanping Huang, Yaohui Li, Yi Zheng, Yuchen Zhu, Yunxian Ma, Ying Tang, Yukun Zha, Yuting Yan, Z Z Ren, Zehui Ren, Zhangli Sha, Zhe Fu, Zhean Xu, Zhenda Xie, Zhengyan Zhang, Zhewen Hao, Zhicheng Ma, Zhigang Yan, Zhiyu Wu, Zihui Gu, Zijia Zhu, Zijun Liu, Zilin Li, Ziwei Xie, Ziyang Song, Zizheng Pan, Zhen Huang, Zhipeng Xu, Zhongyu Zhang, and Zhen Zhang. 2025. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. https://doi.org/10.48550/arXiv.2501.12948

[26] Mathias Dekeyser and R Elliott. 2009. Empathy in Psychotherapy: Dialogue and Embodied Understanding. In *The Social Neuroscience of Empathy*. MIT Press, Cambridge, MA, USA, 113–124. https://doi.org/10.7551/mitpress/9780262012973.003.0010

[27] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. Association for Computing Machinery, New York, NY, USA, 211–223. https://doi.org/10.1145/3172944.3172961

[28] Robert Elliott, Arthur C. Bohart, Jeanne C. Watson, and Leslie S. Greenberg. 2011. Empathy. *Psychotherapy* 48, 1 (March 2011), 43–49. https://doi.org/10.1037/a0022187

[29] Robert Elliott, Arthur C Bohart, Jeanne C Watson, and David Murphy. 2018. Therapist empathy and client outcome: An updated meta-analysis. *Psychotherapy (Chic.)* 55, 4 (Dec. 2018), 399–410. https://doi.org/10.1037/pst0000175

[30] Zohar Elyoseph, Dorit Hadar-Shoval, Kfir Asraf, and Maya Lvovsky. 2023. ChatGPT outperforms humans in emotional awareness evaluations. *Front. Psychol.* 14 (May 2023), 1199058. https://doi.org/10.3389/fpsyg.2023.1199058

[31] Ahmed Fadhil, Gianluca Schiavo, Yunlong Wang, and Bereket A. Yilma. 2018. The Effect of Emojis when interacting with Conversational Interface Assisted Health Coaching System. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare* (New York, NY, USA) *(PervasiveHealth '18)*. Association for Computing Machinery, New York, NY, USA, 378–383. https://doi.org/10.1145/3240925.3240965

[32] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 2 (may 2007), 175–191. https://doi.org/10.3758/bf03193146

[33] C P Feller and R R Cottone. 2003. The Importance of Empathy in the Therapeutic Alliance. *Journal of HUMANISIT COUNSELING, EDUCATION AND DEVELOPMENT* 42 (2003), 53–61. https://doi.org/10.1002/j.2164-490X.2003.tb00168.x

[34] Pamela Fitzgerald and Ivan Leudar. 2010. On active listening in person-centred, solution-focused psychotherapy. *J. Pragmat.* 42, 12 (dec 2010), 3188–3198. https://doi.org/10.1016/j.pragma.2010.07.007

[35] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Ment. Health* 4, 2 (jun 2017), e19. https://doi.org/10.2196/

mental.7785

[36] Malgorzata Gambin and Carla Sharp. 2018. Relations between empathy and anxiety dimensions in inpatient adolescents. *Anxiety Stress Coping* 31, 4 (jul 2018), 447–458. https://doi.org/10.1080/10615806.2018.1475868

[37] Andreas Göldi and Roman Rietsche. 2024. Chatbot agents displaying non-factive reasoning enhance expectation confirmation. In *ICIS 2024 Proceedings. 8. (Annual ACIS International Conference on Computer and Information Science)*. aisel.aisnet.org, Atlanta, GA, USA, 2917. https://aisel.aisnet.org/icis2024/humtechinter/humtechinter/8/

[38] Md Romael Haque and Sabirat Rubya. 2022. An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews. *JMIR MHealth UHealth* 11 (Dec. 2022), e44838. https://doi.org/10.2196/44838

[39] Arthur Bran Herbener and Malene Flensborg Damholdt. 2025. Are lonely youngsters turning to chatbots for companionship? The relationship between chatbot usage and social connectedness in Danish high-school students. *Int. J. Hum. Comput. Stud.* 196, 103409 (Feb. 2025), 103409. https://doi.org/10.1016/j.ijhcs.2024.103409

[40] Arthur Bran Herbener, Michał Klincewicz, and Malene Flensborg Damholdt. 2024. A narrative review of the active ingredients in psychotherapy delivered by conversational agents. *Comput. Hum. Behav. Rep.* 14, 100401 (May 2024), 100401. https://doi.org/10.1016/j.chbr.2024.100401

[41] Jiaxiong Hu, Yun Huang, Xiaozhu Hu, and Yingqing Xu. 2021. Enhancing the Perceived Emotional Intelligence of Conversational Agents through Acoustic Cues. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 282, 7 pages. https://doi.org/10.1145/3411763.3451660

[42] Jiaxiong Hu, Qianyao Xu, Limin Paul Fu, and Yingqing Xu. 2019. Emojilization: An Automated Method For Speech to Emoji-Labeled Text. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3313071

[43] B D Jani, D N Blane, and S W Mercer. 2012. The role of empathy in therapy and the physician-patient relationship. *Forschende Komplementärmedizin/Research in Complementary Medicine* 19, 5 (2012), 252–257. https://doi.org/10.1159/000342998

[44] Sooyeon Jeong, Laura Aymerich-Franch, Sharifa Alghowinem, Rosalind W Picard, Cynthia L Breazeal, and Hae Won Park. 2023. A robotic companion for psychological well-being: A long-term investigation of companionship and therapeutic alliance. *Proc. ACM SIGCHI* 2023 (March 2023), 484–495. https://doi.org/10.1145/3568162.3578625

[45] Deborah Johanson, Ho Seok Ahn, Rishab Goswami, Kazuki Saegusa, and Elizabeth Broadbent. 2023. The Effects of Healthcare Robot Empathy Statements and Head Nodding on Trust and Satisfaction: A Video Study. *J. Hum.-Robot Interact.* 12, 1 (Feb. 2023), 1–21. https://doi.org/10.1145/3549534

[46] Susanne M Jones, Graham D Bodie, and Sam D Hughes. 2019. The impact of mindfulness on empathy, active listening, and perceived provisions of emotional support. *Communic. Res.* 46, 6 (Aug. 2019), 838–865. https://doi.org/10.1177/0093650215626983

[47] Attila Kovari. 2025. Explainable AI chatbots towards XAI ChatGPT: A review. *Heliyon* 11, 2 (jan 2025), e42077. https://doi.org/10.1016/j.heliyon.2025.e42077

[48] François Lauzier-Jobin and Janie Houle. 2022. A comparison of formal and informal help in the context of mental health recovery. *Int. J. Soc. Psychiatry* 68, 4 (jun 2022), 729–737. https://doi.org/10.1177/00207640211004988

[49] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-Disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376175

[50] Yoon Kyung Lee, Yoonwon Jung, Gyuyi Kang, and Sowon Hahn. 2023. Developing Social Robots with Empathetic Non-Verbal Cues Using Large Language Models. https://doi.org/10.48550/arXiv.2308.16529

[51] Yoon Kyung Lee, Jina Suh, Hongli Zhan, Junyi Jessy Li, and Desmond C Ong. 2024. Large Language Models Produce Responses Perceived to be Empathic. https://doi.org/10.48550/arXiv.2403.18148

[52] Q Vera Liao and Jennifer Wortman Vaughan. 2023. AI transparency in the age of LLMs: A human-centered research roadmap. https://doi.org/10.48550/arXiv.2306.01941

[53] Bingjie Liu and S Shyam Sundar. 2018. Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot. *Cyberpsychol. Behav. Soc. Netw.* 21, 10 (oct 2018), 625–636. https://doi.org/10.1089/cyber.2018.0110

[54] Miki Liu, Austin Wong, Ruhi Pudipeddi, Betty Hou, David Wang, and Gary Hsieh. 2018. ReactionBot: Exploring the Effects of Expression-Triggered Emoji in Text Messages. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (nov 2018), 1–16. https://doi.org/10.1145/3274379

[55] Siru Liu, Allison B McCoy, Aileen P Wright, Babatunde Carew, Julian Z Genkins, Sean S Huang, Josh F Peterson, Bryan Steitz, and Adam Wright. 2024. Leveraging large language models for generating responses to patient messages—a subjective analysis. *Journal of the American Medical Informatics Association* 31, 6 (may 2024), 1367–1379. https://doi.org/10.1101/2023.07.14.23292669

[56] Amylie Malouin-Lachance, Julien Capolupo, Chloé Laplante, and Alexandre Hudon. 2025. Does the digital therapeutic alliance exist? Integrative review. *JMIR Ment. Health* 12, 1 (feb 2025), e69294. https://doi.org/10.2196/69294

[57] Mina Marmpena, Angelica Lim, and Torbjørn S Dahl. 2018. How does the robot feel? Perception of valence and arousal in emotional body language. *Paladyn, Journal of Behavioral Robotics* 9, 1 (July 2018), 168–182. https://doi.org/10.1515/pjbr-2018-0012

[58] Robert R Morris, Kareem Kouddous, Rohan Kshirsagar, and Stephen M Schueller. 2018. Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions. *J. Med. Internet Res.* 20, 6 (June 2018), e10148. https://doi.org/10.2196/10148

[59] Maria Moudatsou, Areti Stavropoulou, Anastas Philalithis, and Sofia Koukouli. 2020. The role of empathy in health and social care professionals. *Healthcare (Basel)* 8, 1 (jan 2020), 26. https://doi.org/10.3390/healthcare8010026

[60] Jaya Narain, Tina Quach, Monique Davey, Hae Won Park, Cynthia Breazeal, and Rosalind Picard. 2020. Promoting Wellbeing with Sunny, a Chatbot that Facilitates Positive Messages within Social Groups. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3383062

[61] Melanie Neumann, Jozien Bensing, Stewart Mercer, Nicole Ernstmann, Oliver Ommen, and Holger Pfaff. 2009. Analyzing the "nature" and "specific effectiveness" of clinical empathy: a theoretical overview and contribution towards a theory-based research agenda. *Patient Educ. Couns.* 74, 3 (mar 2009), 339–346. https://doi.org/10.1016/j.pec.2008.11.013

[62] Jacob B Nienhuis, Jesse Owen, Jeffrey C Valentine, Stephanie Winkeljohn Black, Tyler C Halford, Stephanie E Parazak, Stephanie Budge, and Mark Hilsenroth. 2018. Therapeutic alliance, empathy, and genuineness in individual adult psychotherapy: A meta-analytic review. *Psychother. Res.* 28, 4 (jul 2018), 593–605. https://doi.org/10.1080/10503307.2016.1204023

[63] Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. 2017. Empathy in Virtual Agents and Robots: A Survey. *ACM Trans. Interact. Intell. Syst.* 7, 3, Article 11 (Sept. 2017), 40 pages. https://doi.org/10.1145/2912150

[64] Sung Park and Mincheol Whang. 2022. Empathy in Human-Robot Interaction: Designing for Social Robots. *Int. J. Environ. Res. Public Health* 19, 3 (Feb. 2022), 1–21. https://doi.org/10.3390/ijerph19031889

[65] Dhaval Parmar, Stefan Olafsson, Dina Utami, Prasanth Murali, and Timothy Bickmore. 2022. Designing Empathic Virtual Agents: Manipulating Animation, Voice, Rendering, and Empathy to Create Persuasive Agents. *Auton. Agent. Multi. Agent. Syst.* 36, 1 (April 2022), 31 pages. https://doi.org/10.1007/s10458-021-09539-1

[66] Kay T Pham, Amir Nabizadeh, and Salih Selek. 2022. Artificial intelligence and chatbots in psychiatry. *Psychiatr. Q.* 93, 1 (mar 2022), 249–253. https://doi.org/10.1007/s11126-022-09973-8

[67] Julie Prescott and Terry Hanley. 2023. Therapists' attitudes towards the use of AI in therapeutic practice: considering the therapeutic alliance. *Ment. Health Soc. Incl.* 27, 2 (may 2023), 177–185. https://doi.org/10.1108/MHSI-02-2023-0020

[68] Pedro Sanches, Axel Janson, Pavel Karpashevich, Camille Nadal, Chengcheng Qu, Claudia Daudén Roquet, Muhammad Umair, Charles Windlin, Gavin Doherty, Kristina Höök, and Corina Sas. 2019. HCI and Affective Health: Taking stock of a decade of studies and charting future research directions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3290605.3300475

[69] Matthias Schmidmaier, Jonathan Rupp, Darina Cvetanova, and Sven Mayer. 2024. Perceived Empathy of Technology Scale (PETS): Measuring Empathy of Systems Toward the User. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (2024-05-11) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 456, 18 pages. https://doi.org/10.1145/3613904.3642035

[70] Lennart Seitz. 2024. Artificial empathy in healthcare chatbots: Does it feel authentic? *Computers in Human Behavior: Artificial Humans* 2, 1 (jan 2024), 100067. https://doi.org/10.1016/j.chbah.2024.100067

[71] Vera Sorin, Danna Brin, Yiftach Barash, Eli Konen, Alexander Charney, Girish Nadkarni, and Eyal Klang. 2023. Large Language Models (LLMs) and empathy - A systematic review. https://doi.org/10.1101/2023.08.07.23293769

[72] Mark A Stebnicki. 2007. Empathy fatigue: Healing the mind, body, and spirit of professional counselors. *Am. J. Psychiatr. Rehabil.* 10, 4 (nov 2007), 317–338.

[73] Maximilian Vogel. 2023. *I scanned 1000+ prompts so you don't have to: 10 need-to-know techniques*. Medium. https://medium.com/@maximilian.vogel/i-scanned-1000-prompts-so-you-dont-have-to-10-need-to-know-techniques-a77bcd074d97

[74] Bruce E Wampold. 2015. How important are the common factors in psychotherapy? An update. *World Psychiatry* 14, 3 (oct 2015), 270–277. https://doi.org/10.1002/wps.20238

[75] Jeanne C Watson. 2016. The role of empathy in psychotherapy: Theory, research, and practice. In *Humanistic psychotherapies: Handbook of research and practice*

(2nd ed.), David J. Cain, Karen Keenan, and Sheldon Rubin (Eds.). American Psychological Association, Washington, DC, 115–145. https://doi.org/10.1037/14775-005

[76] Jeremy J Webb. 2023. Proof of concept: Using ChatGPT to teach emergency physicians how to break bad news. *Cureus* 15, 5 (May 2023), e38755. https://doi.org/10.7759/cureus.38755

[77] Anuradha Welivita and Pearl Pu. 2024. Is ChatGPT More Empathetic than Humans? https://doi.org/10.48550/arXiv.2403.05572

[78] Alex Wilkins. 2025. Does DeepSeek herald AI's future? *New Sci.* 265, 3529 (feb 2025), 8–9. https://doi.org/10.1016/s0262-4079(25)00207-6

[79] Xuansheng Wu, Haiyan Zhao, Yaochen Zhu, Yucheng Shi, Fan Yang, Tianming Liu, Xiaoming Zhai, Wenlin Yao, Jundong Li, Mengnan Du, and Ninghao Liu. 2024. Usable XAI: 10 strategies towards exploiting explainability in the LLM era. https://doi.org/10.48550/arXiv.2403.08946

[80] Refael Yonatan-Leus and Hadas Brukner. 2025. Comparing perceived empathy and intervention strategies of an AI chatbot and human psychotherapists in online mental health support. *Couns. Psychother. Res.* 25, 1 (mar 2025). https://doi.org/10.1002/capr.12832

[81] Zhengquan Zhang, Konstantinos Tsiakas, and Christina Schneegass. 2024. Explaining the Wait: How Justifying Chatbot Response Delays Impact User Trust. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 27, 16 pages. https://doi.org/10.1145/3640794.3665550

[82] Weixiang Zhao, Yanyan Zhao, Xin Lu, Shilong Wang, Yanpeng Tong, and Bing Qin. 2023. Is ChatGPT equipped with emotional dialogue capabilities? https://doi.org/10.48550/arXiv.2304.09582

[83] Jijie Zhou and Yuhan Hu. 2024. Beyond Words: Infusing Conversational Agents with Human-like Typing Behaviors. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 24, 12 pages. https://doi.org/10.1145/3640794.3665560