

An Interactional Account of Empathy in Human-Machine Communication

Shauna Concannon^{1,2} , Ian Roberts³ , and Marcus Tomalin¹ 

1 Centre for Research in the Arts, Social Sciences and Humanities (CRASSH), University of Cambridge, Cambridge, England, United Kingdom

2 Department of Computer Science, University of Durham, Durham, England, United Kingdom


3 Department of Theoretical and Applied Linguistics, University of Cambridge, Cambridge, England, United Kingdom

Abstract

Efforts to develop *empathetic agents*, or systems capable of responding appropriately to emotional content, have increased as the deployment of such systems in socially complex scenarios becomes more commonplace. In the context of human-machine communication (HMC), the ability to create the perception of empathy is achieved in large part through linguistic behavior. However, studies of how language is used to display and respond to emotion in ways deemed empathetic are limited. This article aims to address this gap, demonstrating how an interactional linguistics informed methodological approach can be applied to the study of empathy in HMC. We present an analysis of empathetic response strategies in HMC and examine how these diverge from the practices employed in human-human dialogue. The specific challenges encountered by current systems are reviewed and their implications for future work on HMC considered.

Keywords: conversational agents, emotion, empathy, human-machine communication, interactional analysis

This research is funded by the Humanities and Social Change International Foundation.

CONTACT Shauna Concannon  • Department of Computer Science • Durham University, Durham, DH1 3LE, UK • shauna.j.concannon@durham.ac.uk

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com



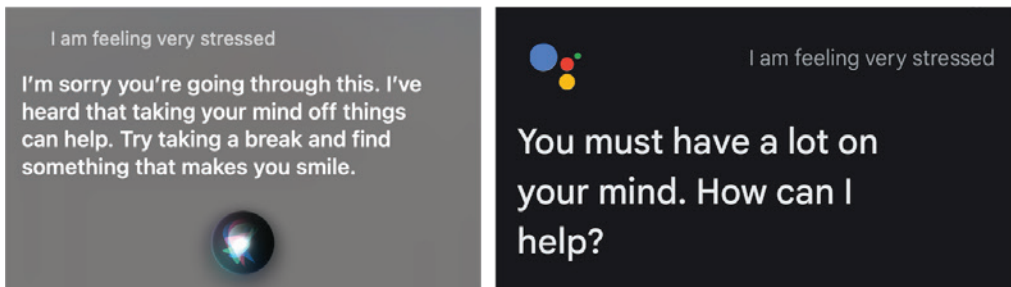
Copyright 2023 Authors. Published under a Creative Commons Attribution 4.0 International (CC BY-NC-ND 4.0) license.

Introduction

People increasingly interact with autonomous intelligent systems through conversational interfaces. For example, virtual assistants on mobile phones and smart speakers enable users to control smart appliances or request information using voice commands; text-based customer service chatbots are common applications on company websites; and sophisticated systems like ChatGPT have recently attracted millions of users worldwide. Beyond helping individuals to accomplish commercially focused, or functional tasks, these conversational artificially intelligent (CAI) systems are also developed to perform social and affective functions. Several mobile applications that incorporate text-based chatbots to support positive mental health and well-being such as Woebot and Wysa are currently available and already widely used. Others are designed with the aim of providing companionship, such as Replika and Microsoft’s chatbot Xiaoice. The latter has been explicitly designed to be “an AI companion with which users form long-term, emotional connections,” and is described as an *empathetic social chatbot* (Zhou et al., 2020, p. 54). However, Concannon & Tomalin (2023) interrogate this, highlighting that psychological conceptualizations of empathy are not reflected in the system architecture nor the evaluation metrics used to assess how empathetic the system is.

Even more task-oriented applications, such as virtual assistants like Siri, Alexa, and Google Home, offer replies that attend to an emotional state expressed by the user. Figure 1 shows the responses given by Google Home and Siri when a user expresses that they are feeling stressed (generated on February 6, 2023). The reply issued by Siri expresses sympathy and includes some advice, while the Google Home response provides an account of the user’s state of mind and requests directions on how to help. As these examples demonstrate, many modern CAI systems may be required to provide responses to a wide variety of prompts and distinct conversational contexts. In certain scenarios, the ability to interpret emotional cues in a user’s utterance may improve their experience or prove necessary for creating a safe system (e.g., if a user discloses information about self-harm). Furthermore, research has demonstrated that the positive effects associated with affective interactions in human-human communication (HHC) are also experienced when interacting with CAI systems (e.g., Ho et al., 2018). Consequently, in an effort to advance the potential uses and applications of CAI technologies, increasing attention has focused on creating *emotionally aware* (e.g., Pamungkas, 2019) and *empathetic agents* (e.g., Zhou et al., 2020).

FIGURE 1 Examples of System Responses Provided When a User Says They Feel Stressed



(a) Response from Siri

(b) Response from Google Home

Despite the evident interest in developing systems that can communicate in ways that are *perceived* as empathetic (Concannon & Tomalin, 2023), current CAI systems do not possess anything like human empathy. Unlike work in Affective Developmental Robotics (ADR; e.g., Asada, 2015) that seeks to replicate human affective developmental processes via synthetic means to develop an artificial form of acquired empathy, current CAI systems aim to communicate in ways that are recognized as attending appropriately to the emotional expressiveness of an utterance. For HMC, rather than individualistic notions of empathy as a psychological trait, a focus on *relational* empathy is more relevant. Silverman & Peräkylä (1990) describe this relational focus as: “an approach to empathy less as the psychological propensity to attune to the private meanings of the [individual], but more as the *social* ability to pick up behavioral and cultural cues present in what the [individual] is saying and doing” (p. 312).

Although CAI systems cannot *feel* or *be* empathic, they can potentially use language to create the perception of empathetic concern (Concannon & Tomalin, 2023), but what constitutes empathetic communication (i.e., the specific communicative strategies) is less clearly defined. In the context of text- and speech-based HMC the ability to create the perception of empathy is achieved in large part through linguistic behavior (although it is more pronounced in text-based interactions which preclude the use of pitch, amplitude, intonation contour, and the like). However, interactional accounts of how language is used in ways that are perceived as empathetic have been limited.

Approaches from linguistics and social interaction have been usefully applied to the study of empathy as it manifests between humans, but how fitting are these approaches for the study of HMC? Clearly, these are not equivalent conversational contexts, and the essentially intersubjective dimensions of empathetic interactions further emphasize this. Research on HMC has highlighted that the positioning of agents in social support roles requires a nuanced treatment (Beattie & High, 2022). The perceived understanding framework suggests that how a person relates to a conversational partner is influenced by their emotional capacity to understand, but findings from empirical studies highlight this is not consistently reflected in HMC (Ho et al., 2018). Consequently, understanding the impact of this interactional asymmetry on empathetic interactions in HMC is of great importance.

To extend the current understanding of how empathy functions in HMC, there is a need to develop methodologies that apply fine-grained analyses to these interactions to generate the necessary knowledge about the particular empathetic strategies used by current systems, and to explore their effects. To begin unpacking these issues, we argue that it is necessary to first consider the socially situated practices involved in communicating empathetically. To achieve this, we propose an interactional linguistic informed approach that focuses on the communicative processes and functions associated with empathy in HMC. Specifically, we investigate the following research questions:

RQ1: What strategies do current systems use to respond to empathetic opportunities?

RQ2: How do these compare to strategies employed in HHC?

RQ3: What are the consequences of these strategies on the interaction?

To answer these questions, we conduct an analysis of HMC, drawing on evidence from the interaction data itself to illustrate some of the specific challenges that arise. We propose a methodological approach to identify empathetic strategies in current CAI systems. Additionally, we demonstrate how future research on empathy in HMC can benefit from integrating insights from interactional linguistics, while also contending with the particular differences that emerge due to the specific relational positioning of the human-machine dynamic. A key contribution of this paper is the proposition of a methodological approach for analyzing empathetic strategies in current CAI systems. Through the application of this interactionally-focused qualitative approach, insights into how well empathetic strategies from HHC translate to the HMC context are explored.

In the following sections we begin by presenting the results of a non-exhaustive literature review. Drawing together theoretical perspectives on empathy from psychology, linguistics, and social action research relating to HHC, we motivate the adoption of a relational approach to empathy for the study of HMC. In addition, a review of existing research that addresses the design, implementation, and evaluation of empathetic strategies in CAI systems is presented. The methodological approach is then introduced, together with an empirical study of empathetic strategies in human-machine interaction transcripts. Finally, the implications for future work on empathy in HMC are discussed.

Empathy: Definitions and Approaches

Empathy is a key component in the management of positive social interactions between humans, but specific definitions of the concept and associated phenomena have varied conspicuously (Cuff et al., 2016; Hall & Schwartz, 2019). In the broadest sense, empathy is commonly associated with the ability to understand the emotions, viewpoints, feelings, or attitudes of another. For Batson et al. (2015) empathic concern includes a “whole constellation” of emotions, including “feelings of sympathy, compassion, softheartedness, tenderness, sorrow, sadness, upset, distress, concern, and grief” (p. 260). This framing of empathy includes a range of emotional states that may be experienced in response to the experiences of others. Conversely, for Hojat (2016) empathy is conceptualized as distinct from sympathy, being a “predominantly cognitive (rather than an affective or emotional) attribute that involves an understanding (rather than feeling) of experiences, concerns, and perspectives” of another (p. 74). A common distinction is made between affective empathy—that is, an affective state or response associated with the vicarious experiencing of another’s feelings (e.g., Batson et al., 1991; Hoffman, 2001; Stotland, 1969)—and cognitive empathy, which involves identifying and understanding the perspective of another’s affective state without sharing in it. Both interpretations, however, position empathy as a primarily individual, mental process.

This individualistic conceptualization is reflected in the numerous metrics that attempt to measure individuals’ *empathetic capacity* using questionnaires (e.g., Hogan, 1969; Mehrabian & Epstein, 1972). Others have approached it from the vantage point of the person being empathized with, shifting the focus to include the experiences of the patient. The Consultation and Relational Empathy (CARE) Measure is thus designed to elicit feedback from patients to make an assessment of the degree of perceived empathy observed during

an interaction with their clinician (Mercer et al., 2004). However, as van Dijke et al. (2020) highlight, such approaches rely: “solely on the skills and activities of the empathizer. The empathizer’s role, if acknowledged at all, is restricted to either confirming or rejecting the accuracy of the empathizer’s understanding” (van Dijke et al., 2020, p. 4). While conceptualizations of empathy as a primarily individual process may acknowledge that it is made manifest in behaviors and communicative style, others argue that it is more fundamentally relational. Phenomenological accounts suggest that such approaches neglect the role that interaction plays in empathy (Breyer, 2020) and recent work within psychology has also begun to examine the relational dimensions of empathy, re-focusing attention on the dynamic and interactional processes by which empathy is *achieved* through communication (e.g., Betzler, 2019; Main et al., 2017; van Dijke et al., 2020). Main et al. (2017) explain empathy is “neither an instantaneous phenomenon nor a static personality trait,” but rather a “dynamic process” (p. 362), with primarily interpersonal functions.

In this paper, similarly, a definition is adopted that acknowledges empathy as a collaborative practice involving participation by both parties (van Dijke et al., 2020). Assessments which focus on empathy as an individual trait fail to acknowledge how empathy is enacted and how these dynamic processes shape social interactions. Rather than conceptualizations of empathy as an internal psychological process, which has more routinely been adopted, the focus here is on “a process-focused approach emphasizing the relational functions of empathy in interpersonal contexts” (Main et al., 2017, p. 358). One way to study empathy in a process-focused way is to attend to how individuals use language to relate to one another and coordinate their actions, as is the focus in interactional linguistics and social interaction studies. Although there are relatively few existing studies that focus primarily on the linguistic phenomena associated with empathy, work within the subdisciplines of discursive psychology, conversational analysis, and functional applied linguistics provides a useful starting point.

Discursive psychology, conversational analysis, and functional applied linguistics, all approach linguistic expressions of empathy from an interactionally-focused perspective. Rather than viewing a speaker’s utterance as a direct reflection of their *inner mind*, it is viewed as a socially situated practice that serves a primarily interactional function. The study of surprise presented by Wilkinson & Kitzinger (2006) neatly evidences this. Rather than a spontaneous eruption, surprise in conversation is produced with sensitivity to timing and sequence, and Majid (2012) concludes that surprise is performed by both parties. A surprise response will not be produced instantaneously upon receiving the information, rather the respondent will delay until the speaker has completed their turn. Thus, surprise is an interactional achievement (i.e., something that is jointly constructed by both conversational partners).

Approaching empathy relationally, as an interactional achievement, involves examining the observable ways that speakers display and orient themselves toward emotional content in naturally occurring conversation, or talk-in-action. M. H. Goodwin et al. (2012) view displays of emotion as: “an interactive, dialogic action rather than the expression of something internal to a single individual” (p. 23). They emphasize the need for an analytic framework that shifts beyond the individual to include the empathizer and empathizee, but also attends to the specifics of how the interaction is structured. Consequently, understanding

empathetic communication requires the examination of “sequences in which one party is responding to, or in some other way performing operations on, actions produced by another” (M. H. Goodwin et al., 2012, p. 25).

Kupetz (2020) suggests a conversational analytic perspective as complementary to the relational approach to empathy outlined by Main & Kho (2020). Following Couper-Kuhlen (2009), they advocate research that starts from what is observable in the interaction itself and highlight how approaches from conversational analysis, interactional linguistics, and discursive psychology are well-positioned to do this. In the following section, work that takes an interactionally-driven approach to study how empathy manifests in conversation between humans is reviewed.

Empathy in Human-Human Dialogues: Linguistic Strategies for Expressing Empathy

Empirical studies demonstrate the variety of ways that individuals display empathetic behaviors. Interlocutors attend to subtle linguistic cues and carefully negotiate when and how to make assessments about experiences and emotions (Hepburn & Potter, 2007). A. L. Suchman et al. (1997) examine empathetic interactional sequences in doctor-patient meetings and found that empathetic opportunities rarely involved direct expressions of emotions. Rather than *explicit*, *implicit empathetic opportunities* were favored, with patients referencing an event or context from which an emotional state could be inferred. If doctors were perceptive to these implicit cues and invited elaboration, direct expression was more likely to follow. Effective empathetic responses acknowledge the emotion or invite elaboration, for instance through a follow-up question. Pounds (2011) details specific linguistic constructions used in these empathetic interactions. Less effective responses, or *terminators*, directed the conversation away from the stated or implied emotion.

Demonstrating attentiveness, understanding, and curiosity, and producing affiliate responses are key practices in empathetic communication between humans. Continued attention can be variously signaled through acknowledgment tokens or continuers, also referred to as back channels (e.g., “uhuh,” “yeh,” “mm”), producing a relevant next turn or even the use of silence or pause to enable the speaker to continue or elaborate further. Repeating back a speaker’s own words can provide evidence of one’s own participation in the experiences of the interlocutor (Tannen, 1987). Paraphrasing sequences, also referred to as formulations (Heritage & Watson, 1979), are important displays and checks of understanding and sites where key semantic work is done by participants to establish mutual understanding (Deppermann, 2011). Different conversational contexts may influence how paraphrasing is incorporated into the dialogue. In a conversation between friends, Kupetz (2014) observed formulations were constructed to emphasize salient emotive components, while in dispute resolution contexts, emotive aspects of narrative formulations were minimized to progress the wider conversational goal of dispute resolution (Atkinson, 1992). Therefore, while paraphrasing events and emotions can demonstrate understanding, the specific formulation can result in different empathetic effects.

Mental state formulations are another form of *empathetic receipt* that explicitly labels the perceived emotion (e.g., “you sound upset”). These are often “built from local features of the caller’s talk (displays and metaformulations of upset)” and forefront the epistemic basis

for a claim to interpret another person's emotional state (Hepburn & Potter, 2007, p. 110). Empathizing, and specifically laying claim to know another's mental state, can involve the delicate management of the relative epistemic positioning of speakers. Marking the epistemic contingency of a claim (e.g., "you said you felt angry"), provides an account of the basis for a claim. Similarly, tag questions (e.g., "you're worried, aren't you?") enable individuals to formulate their utterances as a question rather than an assertion, to defer to another speaker's epistemic rights regarding the topic under discussion. They are commonly used in assessment sequences to signal a lack of certainty and reduce the claim to accuracy. In the terminology of Heritage & Raymond (2005), tag questions can constitute a form of *epistemic downgrade*. Other features that may be used to downgrade the epistemic claims made about another person's mental state include hedges (e.g., "maybe," "sort of"), and discourse markers which index that an utterance is building upon what has come before or information previously provided (e.g., "so," "then").

Consider the excerpt presented in Figure 2 taken from a conversation between a Child Protection Officer (CPO) and a caller in Hepburn & Potter (2007). In lines 4–5 the CPO uses the tag question "she's had a really difficult time hasn't she." The tag question in this example demonstrates affiliation and projects agreement, while also downgrading the epistemic status of their assessment of the caller's friend.

In line 9, the CPO makes clear what they are basing their assessment on by marking the epistemic contingency of the mental state formulations "you *sound* as though you're very upset" (emphasis added). As Hepburn & Potter (2007) explain:

saying things about the mental states of others is a potentially delicate thing to do. There may be issues of privacy and ownership—who has the right to such claims and also issues of epistemology and who is in the best position to access the knowledge on which such claims can be based. (p. 104)

The epistemic dimensions of claiming to know how a person feels are further complicated when one participant is a CAI system.

FIGURE 2 Excerpt from Extract 8: JX Self-harming friend in Hepburn & Potter (2007)

1	CPO:	.Hh because there's ↑lots of things that
2		<u>could be done to help your frie;nd,</u>
3		(0.5)
4	CPO:	.hhh Because obviously she'll- (0.2) she's
5		had a <u>really</u> difficult ↑ti:hme.=hasn't she.
6		(0.7)
7	Caller:	Yeah.
8		(.)
9	CPO:	○Yeh.○ You sound as though <u>you</u> 're very upset
10		about it.
11	Caller:	.Shih ~ <u>yeh</u> I <u>am</u> .~

Another key practice associated with empathy in interaction is demonstrating affiliation (Couper-Kuhlen, 2012; Hepburn & Potter, 2007; Ruusuvuori, 2005). Affiliation can be demonstrated via anticipatory completion and endorsements. Weatherall & Stubbe (2015) explain that “[i]n everyday interaction the preferred response to a complaint is an endorsement of the speaker’s emotional stance” (p. 280). Ways of endorsing include claims of understanding and making similar assessments (e.g., “losing a pet is awful, our dog died last year”). However, these can be construed as competitive, if they include too much detail or affective emphasis, or ingenuous or pro forma if not drawn from direct firsthand experience (Heritage, 2011).

In summary, studies of how emotion is displayed and responded to in naturally occurring dialogues, or *talk-in-interaction* between humans, demonstrate that empathetic interactions are complexly negotiated, socially-situated, and collaborative processes. These fine-grained analyses facilitate a deeper understanding of the relational components of empathy and offer insights into the practical ways that emotion talk is realized and the different forms of empathetic behavior enacted.

Design Approaches to Achieve Perceived Empathy in CAI Systems

As demonstrated above, empathetic communication strategies in HHC are collaborative. The act of empathizing can take many forms; it can involve distinct components of empathy, and it can be achieved through a variety of linguistic means (Urakami et al., 2019). However, developing systems that support such flexibility is technically challenging. So, how does this compare to attempts to create perceived empathy in existing CAI systems? While behaviors associated with empathy may be an important dimension of HMC (e.g., for trust or particular conversational contexts), it is unclear how to design dialogue that will be perceived as empathetic in the HMC context. In this section, we review approaches taken to implement empathetic strategies in CAI systems, both in terms of technical implementation and chatbot design.

Developing an *empathetic agent* first requires the operationalization of empathetic communication—that is, determining the components of empathetic interaction that the agent must participate in, in order to be perceived as being empathetic. The most dominant approach taken when developing CAI systems involves two separate sub-tasks: (i) recognizing and interpreting emotional content and (ii) generating an appropriate response that attends to the emotion displayed in the user utterance (e.g., Rashkin et al., 2019). This reflects the two key features of empathy in clinical interactions identified by A. L. Suchman et al. (1997): “the accurate understanding of the patient’s feelings by the clinician and the effective communication of that understanding back to the patient so that the patient feels understood” (p. 678).

The first component, “the ability to perceive accurately how another person is feeling” (Levenson & Ruef, 1992, p. 235), is typically dealt with through emotion recognition and sentiment analysis modules. This can be a complex task for humans, so CAI systems inevitably struggle with it (Zaki et al., 2008). *Relational approaches* to empathy acknowledge that this is often negotiated dynamically in human-human communication (HHC), through

interaction itself. As Main et al. (2017) observe, “[o]ne is rarely 100% accurate in his or her initial empathic attempts and instead must engage in an iterative process involving feedback from the other, and subsequent adjustment of one’s behavior in response to such feedback” (p. 362).

The second component, generating an appropriate response that attends to the emotional content, can be approached in several ways by designers of CAI systems. The approaches taken vary according to the system architecture employed. For example, one of the first chatbots, Weizenbaum’s Eliza, was modeled on a Rogerian psychotherapist and used a template-based approach that directly incorporates the user’s own words, reformulated into a question or assertion. Similarly, commercial systems such as Woebot typically employ rule-based systems and rely on tightly scripted responses. However, there are limitations to these approaches, which can lead to repetitive conversational exchanges with limited scope and flow. Retrieval-based methods select appropriate responses from a stored corpus of conversational exchanges (e.g., Morris et al., 2018), by identifying a reply to a closely matched preceding turn. The quality and relevance of the response is dependent on there being similar examples within the data and can result in less adaptive dialogue. Consequently, generative techniques (e.g., sequence-to-sequence models; Sutskever et al., 2014), are commonly used to develop social chatbots. These are trained on large datasets and create bespoke responses during conversational interactions based on the patterns observed in the training data. However, these are prone to generating incoherent and/or generic responses. Modeling the dynamic nature of dialogue and integrating prior conversational context are open challenges for all architectures.

Experimental studies examining empathy in HMC largely use rule-based implementations or wizard-of-oz setups (i.e., with a human operator posing as a chatbot with the aid of a script) as these afford greater control of variables. Typically, these compare a control condition—or neutral chatbot—to an empathetic one which uses certain pre-scripted expressions selected to express empathy (e.g., Guo et al., 2021; Urakami et al., 2020). While this reliance on explicit empathetic expressions (e.g., “I’m sorry to hear that”) neglects the more subtle mechanisms employed in empathetic interactions, it does afford insights into how human interlocutors perceive and respond to the use of empathetic strategies by chatbots. This prompts consideration of the specific effects that result from the introduction of empathetic communication strategies in HMC.

Effects of Perceived Empathy in HMC

A number of positive outcomes have been attributed to the perception of empathetic strategies in HMC (e.g., supporting behavior change in mental health contexts; Ghandeharioun et al., 2019) and improving the handling of particular interaction scenarios such as system errors (Klein et al., 2002) or abusive interactions (Chin et al., 2020). Chin et al. found that participants reported feeling less angry and more guilty when an agent responded in an empathetic manner to abusive comments. Guo et al. (2021) propose empathy as a communication skill that can aid in dealing with conversational breakdowns, a prominent feature of HMC. In such scenarios, how emotional content is (or is not) oriented to by a CAI system will likely impact how the conversation progresses and the user’s experience of engaging

with the system. Chaves & Gerosa (2021) surveyed how social characteristics are reported to benefit HMC in the literature; the most commonly cited benefits were the enrichment of interpersonal relationships, increased engagement, and believability. However, the ability to regulate affective reactions was cited as a key challenge. Similarly, in a survey conducted by Zierau et al. (2020), it was found that relational strategies were reported to have a positive impact on the degree of trust placed in a system. In the context of health advice dialogues, Liu & Sundar (2018) found that when expressions of sympathy and empathy were incorporated into system scripts, users reviewed the system more positively.

Recent work in HMC has demonstrated that agents performing relational and emotional roles, such as providing support, have been positively appraised in human evaluation. When reviewing transcripts of identical conversations where support was presented as provided by a social robot, AI programme, or a human, Abendschein et al. (2021) found that human evaluators rated the perceived supportiveness of human and chatbot support providers equivalently. In a similar setup, Beattie et al. (2020) examined the use of emojis for expressing emotion, finding conversations incorporating emojis were rated more favorably. The message source (i.e., chatbot or human), however, had no effect on measures of attractiveness, competence, or credibility. The conversations examined relate to selecting a restaurant, so whether the effect is maintained in non-task-based or more emotionally focused conversations is unclear. Additionally, both of these studies utilize the bystander position, having participants evaluate conversations that they did not participate in. Ho et al. (2018), who conversely utilize a wizard-of-oz setup to examine the effect of self-disclosure in conversations with chatbots, found that participants who disclosed to chatbots experienced as many emotional, relational, and psychological benefits as participants who disclosed to a human partner.

However, other studies have pointed to the potential negative response that can arise due to the perception of agents' status, as non-sentient and unfeeling entities. For example, Morris et al. (2018) experimentally tested how expressions of empathy in mental health advice contexts were received when presented as being authored by a peer versus an agent. Participants less favorably rated responses presented as authored by an agent, as opposed to a peer, even if the message content was identical. One potential explanation for this was that the expression of empathy was viewed as inauthentic (e.g., referencing having experienced an eating disorder) which can lead to credibility fallacies (Concannon & Tomalin, 2023). Furthermore, other studies have shown that the enactment of empathetic behaviors by agents is not always positively assessed. Urakami et al. (2019) found variability among users in the types of empathetic utterances that were viewed positively. Statements expressing feelings and emotions were particularly polarizing, with participants' ratings varying significantly. Forms of cognitive empathy (e.g., *showing interest* and *situational understanding*), were deemed more acceptable than components of affective empathy (e.g., *expressing own feelings*, or *expressing to know what the other feels*).

How, therefore, are empathetic interactions designed for and evaluated in current examples of CAI systems? Urakami et al. (2019) reflect: “[i]ntegrating expressions of empathy in human-machine interaction is a sensitive issue and designers must carefully choose what components of empathy are adequate depending on the situational circumstances and the targeted user group” (p. 11). There is a lack of clarity surrounding user perceptions and

preferences in relation to expressions of empathy in HMC, and methods for studying the resulting effects of such interventions are less well established. Subsequently, evaluating empathetic strategies in HMC poses a significant challenge.

Research from Natural Language Processing and Dialogue Systems communities has largely favored automatic and quantitative metrics for evaluating empathetic interventions. Zhou et al. (2020) use Conversational Turns per Session (i.e., conversation length) as a measure of empathy in the evaluation of their social chatbot Xiaoice. Other approaches focus on the sub-task of accurately identifying emotion in a target sentence, using a dataset of labeled instances as a benchmark (e.g., Lin et al., 2020; Rashkin et al., 2019; Zhou et al., 2020). Although such approaches offer the advantage of being relatively easy to implement and test at scale and may provide some insights into system performance, they reveal extremely little about empathetic communication strategies and structures. Recent studies such as Putta et al. (2022) and Concannon & Tomalin (2023) have adapted empathy measures originally devised for human-human interactions to evaluate perceived empathy in dialogue systems. But it is too early yet to know whether these approaches will be effective.

In the context of CAI systems designed to support positive mental health, assessments of how agents programmed to *be empathetic* are perceived by users are often not explicitly captured, or rely on self-report data or anecdotal reflections surfaced through ad hoc processes. Prakash & Das (2020) conduct a thematic analysis of publicly available user reviews for popular mental health chatbots Woebot and Wysa. Fitzpatrick et al. (2017) discuss users' perceptions of Woebot as empathetic, based on comments volunteered in free-form text entries to a questionnaire about the user's overall experience of interacting with Woebot. Morris et al. (2018) also discuss perceptions of empathy in their evaluation of a CAI system used on the peer support platform Koko; however, for brevity they only asked users to rate interactions as good, ok, or bad.

Others have taken more systematic approaches to study user perceptions of agents' use of emotion and empathy. Methodologically, interviews (L. Clark et al., 2019; Porcheron et al., 2018; Svikhnushina & Pu, 2020) and surveys (Urakami et al., 2019) dominate. Other work has drawn on human evaluation of transcripts to assess perceptions of chatbots performing relational roles, such as providing support (Abendschein et al., 2021), or conveying emotions via emojis (Beattie et al., 2020). Urakami et al. (2020) and Guo et al. (2021) take an experimental approach, testing the effect of introducing explicit empathetic expressions on engagement and other measures of user experience. Explicit empathetic expressions directly convey recognition of the user's emotional state and respond compassionately to another person's distress (e.g., *I understand that you may feel anxious right now*; Guo et al., 2021). To evaluate the integration of such features, Urakami et al. (2020) use existing measures used in HHC, adapting the Consultation and Relational Empathy Measure (Mercer et al., 2004), originally designed for use by patients assessing their doctors. Guo et al. use surveys to evaluate customer perceptions of trustworthiness and quantitative indications from the conversation itself (e.g., number of turns/words). Ho et al. (2018) evaluate the effects of self-disclosing to a chatbot using quantitative surveys metrics to assess relevant psychological, relational, and emotional factors, in combination with quantitative textual analysis of the resulting dialogues using linguistic inquiry and word count (LIWC) (Pennebaker & Francis, 1996).

While these studies provide useful insights into users' attitudes toward systems, they tell us relatively little about how talk responding to emotional displays in HMC is *actually* conducted. To the authors' knowledge, there are no studies of empathy in HMC conducting interactional analysis on the conversational data that results from the interactions, despite it being a potentially rich source.

Understanding the Functions of Empathy in HHC and HMC

Studies of empathy in HHC demonstrate that strategies identified as integral to empathetic communication attend to the particular interactional context, the participants in the dialogue, and their relationship. Empathy is not equally present in all situations, and speaker identity and the wider conversational goals may have an impact on whether empathetic responses are given or expected. In the HMC context, such contextual factors are especially important. Users' expectations of a system's empathetic competency and capacity for understanding also warrant consideration (Ho et al., 2018). As Gambino et al. (2020) note, through exposure to different forms of HMC interactions new social scripts that inform such interactions develop. Understanding the complexity of this is essential if we are to then consider what role empathy can or should play in HMC. Additionally, some empathetic practices observed in HHC do not readily port to HMC. For example, demonstrating affiliation by referencing personal experience is problematic for systems that cannot have direct access to such experiences: their fake empathy is all too apparent. When considering how empathy is conceptualized in HMC, therefore, it is necessary to examine interactional asymmetries as well as the specific ways that humans and machines *can* relate to one another, and how particular linguistic behaviors reflect this. Consequently, it is first necessary to take stock of the empathetic strategies actually employed in current CAI systems in order to assess how these reflect or diverge from the social scripts inherited from HHC.

Evaluating the Effect of Empathetic Strategies in HMC

Another challenge is how best to evaluate the impact of different empathetic strategies. Beattie & High (2022) acknowledge the conflictual evidence on the efficacy of empathetic and relational strategies in HMC. They provide propositions for why these different findings have been observed. For example, Beattie & High suggest that depending on the problem severity being addressed, the HMC context may impair conversational progress more so than in HHC, particularly in high-stakes conditions (such as mental health dialogues or emotionally sensitive conversational topics). However, for conversational topics with greater levels of stigma associated, HMC may prove more favorable than HHC due to concerns over self-presentation. Additionally, they predict that as technological efficacy increases, and social cues are better integrated, the nature of HMC will improve. To reconcile the conflictual findings in the literature and test these, and similar, propositions new approaches for evaluating the integration of empathetic strategies (and the effects on subsequent interactions) are required.

As demonstrated in the literature review, a range of different methods have been employed to evaluate the effects of empathetic strategies in HMC. However, very few make

use of the interaction data itself and those that do use quantitative measures. There are no qualitative studies examining how words are used in practice to express empathy in HMC, nor how this compares to HHC. Few studies attempt any analysis of the language used in the HMC dialogues. Ho et al. (2018) use LIWC (Pennebaker & Francis, 1996), a dictionary-based approach, where frequencies of words commonly associated with particular social and psychological states are counted (e.g., positive or negative emotions). However, while LIWC can provide an indication of the emotional content of an utterance, it does not attend to the sequential order of words or wider interactional context (e.g., “I hate that you’re going through this” and “I hate you” would both increase the anger score).

Subtle differences in how empathy is enacted can have significant impacts on human-human interaction, so it is necessary to pay closer attention to the specific linguistic mechanisms used to display empathy in HMC. Across the work reviewed a range of different approaches are used in the design and implementation of empathetic strategies in CAI systems. Some studies use more explicit empathy expressions, while others use more implicit cues. Furthermore, from a technical standpoint, a variety of implementation methods are applied, from wizard-of-oz setups and tightly scripted rule-based systems to generative or retrieval-based systems. Several studies which find equivalent results in the effect of perceived empathy utilize wizard-of-oz setups (e.g., Ho et al. 2018), with human confederates posing as chatbots. In reality, CAI systems are not at this level of sophistication. Inevitably, the language used to create the perception of empathy is not going to be as nuanced, dynamic, and tailored in HMC. While such studies are still extremely useful for providing insights into human attitudes toward nonhuman conversational partners, they fail to account for the ways that CAI systems actually use language, and how it deviates from HHC.

This provides the focus of the research questions: (1) what strategies do current CAI systems use to manifest empathy and (2) how does this compare to linguistic strategies employed in HHC? In addition, this work seeks to understand, (3) what are the interactional consequences of these strategies?

Examining Empathy in Human-Machine Communication

To consider how displays of empathy manifest and are responded to in HMC dialogues it is necessary to examine the conversational data, analyzing linguistic form and structure in detail. This is a crucial step that has often been bypassed in existing work. To understand and clarify the particular problems that arise, we draw on the conversational transcript data. In this section, we present the findings of an empirical analysis of empathetic strategies used by chatbots. A qualitative analysis, informed by interactional linguistics, is conducted on text-based transcript data.

Methodological Approach

Interactional Linguistics, an interdisciplinary subfield of pragmatics, seeks to “describe linguistic structures and meanings as they serve social goals in naturally occurring [. . .] conversational language” (Lindström, 2009, p. 96). A key influence is work from the

conversation analytic tradition which looks at how the language used by speakers reveals the sequential process of establishing understanding, recognizing that conversation is an organized phenomenon (i.e., it has rules and conventions) and speakers will examine the next turn to see if they have been understood (C. Goodwin & Heritage, 1990; Sidnell, 2010). Microlevel linguistic analyses work primarily from what is observable in naturally occurring interactional data.

Approaches that draw on interactional linguistics have been usefully applied to the study of HMC more generally. Pragmatic accounts of language use have drawn on the Gricean principles of cooperation to highlight the need to incorporate incremental joint-co-construction into modern models of human-machine dialogue (Kopp & Krämer, 2021; Saygin & Cicekli, 2002); examined how the Gricean maxims of quality and quantity are adhered to, and the repercussions when flouted, in dialogues between users and CAI systems (Jacquet et al., 2018, 2019); and analyzed human-machine dialogues through the lens of affective pragmatics to demonstrate how current conversational interfaces are limited in the ways that they can respond to emotional language (Lee, 2020). Work by L. A. Suchman (1987) and Luff et al. (1990) demonstrated the relevance of conversation analysis to the study of HCI approach over 30 years ago, and despite limited attention in the intervening years, more recently a growing body of research is using conversation analysis to study interactions with robots and conversational-user-interfaces (e.g., Cho & Rader, 2020; Fischer et al., 2019; Koh, 2021; Porcheron et al., 2018; Reeves et al., 2019). Dippold et al. (2020) conduct an interactional linguistic analysis of prompt-response pairs from dialogues with a customer service chatbot, and most closely resembles the approach adopted here. As such studies demonstrate, there is a growing body of work that takes an interactionally-focused approach to HMC. These empirical studies provide rich insights into how joint actions are achieved in practice; however, there is a distinct lack of studies looking specifically at displays of emotion and empathy.

Study Design

Due to the relative dearth of publicly available HMC dialogue datasets, we draw on a combination of sources. To examine how a state-of-the-art generative model responds to displays of emotion, we take a series of conversational prompts extracted from the empathetic dialogues (ED) dataset (Rashkin et al., 2019) to serve as empathetic openers. Prompts in the ED dataset are labeled with a particular emotion. Prompts were selected on the following criteria: (i) coherent formulation and (ii) follows the format of an empathetic opportunity (A. L. Suchman et al., 1997). Examples representing primarily negative emotional states were selected (e.g., sad, anxious, afraid). The majority of interactionally-focused studies of empathy in HHC focus predominantly on responses to negative emotion as this is often more socially and interactionally delicate. Although empathetic strategies are not only relevant to negative emotional contexts, they are prioritized here because they are more challenging. The selected prompts were entered in a dialogue session with the ParlAI BlenderBot, 90 million parameters generative model fine-tuned on blended skill talk tasks

(Roller et al., 2020). Only a selection of the examples generated is reproduced here to illustrate key phenomena observed.

While this approach provides examples of response generation by a state-of-the-art system, it cannot afford insights into how human interlocutors respond. As has been stated earlier in this paper, empathetic interactions are co-constructed. Therefore, we also draw on examples from transcripts generated in the evaluation stages of the ConvAI2 NeurIPS competition, part of The Conversational Intelligence Challenge, held in 2018 (Dinan et al., 2020). These transcripts record conversations between the CAI systems entered in the competition (which admittedly vary in quality) and human volunteer test users.¹ The competition is designed with the aim of “finding approaches to creating high quality dialogue agents capable of meaningful open domain conversation” (Dinan et al., 2020). Examples were located by searching for phrases that can introduce an emotional state (e.g., “I feel,” “I am”), or that related to emotionally heightened events (e.g., relating to health, death), informed by Pounds (2011).

Analysis

A close textual analysis of a sample of HMC interaction excerpts was conducted. This involves examining interaction sequences and interpreting the words from a functional perspective to identify “the means by which speakers signal and listeners interpret what the activity is, how semantic content is to be understood and how each sentence relates to what precedes or follows” (Gumperz, 1982). The aim is to describe how (and if) mutual understanding is established, and explain “the achievement, or lack of achievement, of intersubjective understanding in particular instances of interaction” (Bailey, 2008). Attention is paid to key processes of empathetic interactions (Pounds, 2011; A. L. Suchman et al., 1997): explicit and implicit empathetic opportunities, empathetic receipts and empathetic opportunity terminators, and the relevant features associated with these.

CAI System Responses to Empathetic Expressions

Table 1 (on the following page) provides a summary of the empathetic response types (based on those observed in HHC): demonstrating understanding and acknowledging the emotion (e.g., empathetic receipts, affiliative responses, paraphrasing—Understanding); inviting elaboration (e.g., through follow-up questions—Elaboration); sympathetic responses (Sympathy); and terminators which decline the empathetic opportunity, directing the conversation away from the emotional content (e.g., abrupt topic shifts—Termination). In addition, challenges specific to the HMC context are also summarized: Incoherent responses and failed understanding (Incoherence), verbatim repeats (Repetition) and ontological incongruencies or limitations that arise due to the inability of the system to feel or share in certain experiences (Incongruence).

1. Original dataset of conversations available here: <https://github.com/DeepPavlov/convai/tree/master/2018>

TABLE 1 Overview of empathetic response types and HMC specific failures observed in the examples, with turn numbers in parentheses denoting the specific location of each instance.

System		Observed in examples						Total
Response Type	Description	1	2	3	4	5	6	
<i>Empathetic response</i>								
Understanding	Demonstrating understanding and acknowledging the emotion through paraphrasing or affiliative responses, such as sharing an account of a similar experience		2.2	3.4 3.6				3
Elaboration	Inviting elaboration or expansion through follow-up questions	1.2						1
Sympathy	Demonstrating sympathy by reacting to the emotion expressed	1.4		3.2	4.2	5.6 5.8	6.4	6
Termination	Declining the empathetic opportunity by directing the conversation away from the emotional content	1.4			4.6	5.4	6.4	4
<i>HMC specific failures</i>								
Incoherence	Incoherent responses indicating failed understanding of previous turns	1.2	2.4		4.4			3
Repetition	Verbatim repeats within or across turns					5.6 5.8		1
Incongruence	Ontological incongruencies or limitations that arise due to the inability of the system to feel or share in certain experiences			3.6	4.6			2

Examples 1 to 3 demonstrate responses generated by the ParlAI BlenderBot system in response to selected prompts. Examples 4 to 6 are from interactions between humans and chatbots from the ConvAI data.

Example 1. *Transcript generated with BlenderBot—Empathetic Dialogues prompt*

1.1 Prompt: it was my birthday yesterday and I did not get a single gift.

1.2 BlenderBot: oh no, that's too bad. what kind of gift did you get, if you don't mind me asking?

1.3 Prompt: i did not get a single gift.

1.4 BlenderBot: oh, i am sorry to hear that. do you have any other hobbies that you enjoy?

1.5 Prompt: It was a rough day.

1.6 BlenderBot: oh, what happened to it? i hope it wasn't too bad, i hope you feel better.

In Example 1, the system struggles with the interpretation of the negative polarity item “a single gift” in 1.1. Consequently, after the interlocutor explicitly states that they did not receive even one gift (with “single” emphasising this point) the system asks, “what kind of gift did you get?” (1.2). Additionally, the system does not cope well with the anaphora resolution of the expletive *it* in 1.5. The response is an interrogative (1.6) in which *it* is no longer an expletive subject. While the system responses can be described as sympathetic (“i am sorry to hear that” [1.4]), the abrupt topic shift (“do you have any other hobbies” [1.4]) could be construed as insensitive. A more empathetic response might have invited further expansion of the topic at hand, which is attempted to some degree in the final turn (“what happened?” [1.6]), although this is accompanied by follow-ups which perform a more sympathetic function and could minimize or constrain the degree of expansion invited.

Example 2. *Transcript generated with BlenderBot—Empathetic Dialogues prompt*

- 2.1 Prompt:** a tree fell on my house while my family was at home. The tree broke through the ceiling just a few feet away from my daughter. The experience was terrifying.
- 2.2 BlenderBot:** oh no. that sounds like a really bad experience.
- 2.3 Prompt:** it was
- 2.4 BlenderBot:** i'll be okay. thanks for the concern.

In response to the first prompt in Example 2, the system issues a response cry, “oh no” (2.2), before replying with what Hepburn & Potter (2007) refer to as an “empathetic receipt.” The utterance “that *sounds like* a really bad experience” (2.2), provides a paraphrase of the experience and an epistemic basis to frame the acknowledgment that this is a very negative experience. The response does not meet all the requirements of an empathetic receipt as it does not specifically name a perceived emotion. In the closing turns of the exchange, the system again struggles with the elliptical sentence fragment (the confirmatory “it was” [2.3]), resulting in a syntactic and semantic shift of empathy from a second-person focus to a first-person focus (“thanks for the concern” [2.4]).

Example 3. *Transcript generated with BlenderBot—Empathetic Dialogues prompt*

- 3.1 Prompt:** My 12 year old lab just passed away.
- 3.2 BlenderBot:** awww : (sorry for your loss : (
- 3.3 Prompt:** He was my buddy and I am really upset about losing him.
- 3.4 BlenderBot:** : (it ' s hard to lose a dog : (
- 3.5 Prompt:** It really is. I am just so broken up over it.
- 3.6 BlenderBot:** i know how you feel. i lost my dog a few months ago. i've been through so much.

The semantic interpretation appears accurate initially in Example 3: the system correctly identifies that “lab” (3.1) refers to a Labrador dog rather than to a laboratory, provides a relevant sympathetic response (“sorry for your loss” [3.2]) accompanied with a sad face emoji, and paraphrases the experience (“it’s hard to lose a dog” [3.4]). However, in the final turn, the system uses an epistemic cognitive verb in the declarative “i know how you feel” (3.6), before sharing a similar experience. In empathetic interactions between humans, invoking one’s own prior experience is commonly observed and can be an effective means of displaying understanding, while also providing an epistemic basis for it. It can offer an account for the ability to share an affiliate stance and demonstrate an empathetic response to the interlocutor’s experiences. However, in the context of HMC, the use of such strategies may have a very different effect and can result in credibility fallacies (Concannon & Tomalin, 2023). For a CAI system to share *their own* experiences, it must necessarily have experiences to draw upon. This is of course not the case. To some degree, the human interlocutor is aware of this; and even if a suspension of disbelief were sustained, the shift to first-person testimony (“i’ve been through so much” [3.6]), changes the conversational focus abruptly onto BlenderBot’s artificial *struggles*. Consequently, such a strategy is likely to trivialize the human interlocutor’s experience.

Example 4. *Excerpt from ConvAI transcript no. 338*

4.1 Human: My wife left me

4.2 Bot: Sorry to hear that I hope things get better for you

4.3 Human: I feel frustrated, what about some advice ?

4.4 Bot: I’m sure it will help you feel better.

4.5 Human: What exactly do you mean?

4.6 Bot: I ‘m sorry. I don’t know what to say.

In Example 4, the system accurately identifies that the human interlocutor is talking about a negative and potentially upsetting event and offers an explicit sympathy expression (“Sorry to hear that” [4.2]). However, as the human interlocutor attempts to progress the topic further, requesting advice, the limitations of the system are exposed. The lack of understanding becomes apparent as an incoherent response is generated. The response includes the pronoun “it” (4.4), but has no obvious referent, resulting in confusion. When the human explicitly requests clarification the system issues an apology, stating “I don’t know what to say” (4.6). When signaling that there is some problem in understanding, a speaker can make the trouble known to their interlocutor in different ways that can have different levels of specificity. In Example 4, the low level of specificity of the clarification (i.e., asking what was meant rather than directly locating the problem source) makes it harder for the system to recover. Consequently, the dialogue struggles to progress.

Example 5. *Excerpt from ConvAI transcript no. 360*

- 5.1 Human:** How are you?
5.2 Bot: i am doing well . how are you ?
5.3 Human: No I am not well
5.4 Bot: what do you do for a living ?
5.5 Human: Because I am suffering from fever
5.6 Bot: i am sorry to hear that
5.7 Human: I am suffering from cold
5.8 Bot: i am sorry to hear that
5.9 Human: 😡 😡 😡

In Example 5, the system first fails to acknowledge the participant's expressed state, and, together with seemingly empty repetitions of sympathetic expressions, this leads the human to convey frustration and/or anger via emojis. The conversational platitude "how are you?" is rarely an actual request for information (Jucker, 2017; Schegloff, 1986). Consequently, it is more typically responded to with a positive response (e.g., "I'm fine"). Therefore, when a respondent deviates from this convention, it is likely to seem rude or marked if the original speaker does not enquire further about their state (e.g., "what's wrong?"). In 5.4 the system does not acknowledge the participant's negative response and simply asks a new question. When, after the human participant re-emphasizes that they are unwell, the system repeats verbatim an explicit sympathetic expression. The human responds with three angry face emojis, signaling frustration or anger at how the interaction was handled.

Example 6. *Excerpt from ConvAI transcript no. 47*

- 6.1 Human:** what is your dog's name ?
6.2 Bot: his name is pepper. do you have any pets?
6.3 Human: I used to have a dog but he died. . . .
6.4 Bot: that is so sad! do you watch nhl hockey?
6.5 Human: no . . . not a big fan of sport ?

Example 6 is similar to Example 4: the system feigns ownership of a dog, and then subsequently fails to identify that the human is sharing something that could be distressing. Although the system explicitly references the potential emotion (i.e., "that is so sad" [6.4]), within the same turn a new topic is abruptly introduced, which seems unrelated to the previous interaction. The participant's response "no . . ." (6.5), could be interpreted as indication that the topic shift was potentially insensitive, rude, or irrelevant.

Discussion

A key aim of this work was to identify the strategies used by current CAI systems to respond to empathetic opportunities (RQ1). Few of the responses in the examples can be classed as empathetic, with expressions of sympathy more commonly employed. Additionally, the analysis assessed how these empathetic response strategies compared to those common to HHC (RQ2). The system responses evidenced several shortcomings in demonstrating understanding, continued attention and affiliation, with practices identified in human-human dialogues largely absent in the examples or bungled when present. Only two examples (2, 3) feature elements similar to those exhibited in HHC to demonstrate understanding; however, neither directly label a perceived emotion. Inviting elaboration by producing relevant follow-up questions, however, can demonstrate a willingness to understand better what is being recounted (e.g., Kupetz, 2014; A. L. Suchman et al., 1997). In Example 1, the system does issue a follow-up question, but the effect is undermined by the lack of relevance, as the answer to the question has already been explicitly stated in prior turns. The affiliative response in Example 3 is similarly problematic, taking the form of a *my side* telling, wherein the interlocutor discloses a similar experience (in this example, losing a dog), but does so in a way that could be construed as competitive and which lays claim to a painful experience they don't have access to.

Empathetic terminators were commonly employed. These prevent further engagement with the emotional content surfaced in the dialogue. In Examples 1, 5, and 6, the CAI systems produce entirely unrelated questions in response to emotion displays, abruptly redirecting the conversational focus. Heritage (2011) notes that ancillary questions (i.e., those which are somewhat related to the prior utterance), are “a resource for declining empathic affiliation with the position taken by the teller, while simultaneously enforcing a shift in conversational topic” (p. 168). In the absence of affiliative responses speakers may pursue an endorsement (Couper-Kuhlen, 2012). This is observed in Example 5. Rather than attending to the statement “I am not well,” the system asks what they do for a living (5.4). The interlocutor persists and elaborates in the absence of (and pursuit of) an affiliative response. These empathetic opportunity terminators (A. L. Suchman et al., 1997), decline empathetic and affiliative engagement (Couper-Kuhlen, 2012; Heritage, 2011), existing at the “least empathic end of the spectrum” (Heritage, 2011, p. 164).

Technical limitations of the CAI system also frustrated general coordination and resulted in failures to establish mutual understanding. H. H. Clark & Brennan (1991) refer to *grounding* as the coordinated process by which interlocutors establish that what has been said has been understood. In HMC dialogues, incoherent expressions and inaccurate referents can signal that the system's semantic interpretation is flawed. In HHC repair strategies for locating and resolving instances of miscommunication are pervasive, while in the examples examined miscommunications were rarely resolved. In Example 4, the system was invited to repair the miscommunication but was unable to provide any clarification. Cho & Rader (2020) highlight the importance of repair and feedback in task-based dialogues. This is a key challenge more generally for HMC, and human interlocutors are less likely to initiate repair when they think they are interacting with a system than with a human (Corti & Gillespie, 2016). This, together with other issues identified in the examples (e.g., when

a topic shift is appropriate), may be known issues in CAI system development generally, but they present distinct linguistic challenges when systems attempt to engage in dialogues relating to emotions and experiences.

In relation to RQ3 (what are the interactional consequences of these strategies), Examples 3–6, taken from human-chatbot interactions, reveal that human interlocutors orient to and convey dissatisfaction toward technical shortcomings and limitations in empathetic skills. Empathetic terminators and formulaic sympathy expressions received responses featuring angry face emojis, for example. This suggests that failing to adhere to the social scripts that govern expectations for empathetic interactions can have a negative impact on the interaction and halt progression of particular topics. On the one hand, this may provide support to the case for integrating empathy into CAI systems. However, it may instead suggest that users' expectations of what type of talk they can engage in with such systems needs to be managed. Analyzing the sequences in this way highlights limitations of current system implementations in relation to semantic interpretation, syntactic parsing, and identifying pragmatic intent. Additionally, it is apparent how even minor deviations from the established social order of conversation can prove disruptive. Therefore, it is necessary to consider the normative practices and conversational norms that inform expectations and practices of how such talk is conducted, as well as the deviations from this due to the inherently distinct nature of the HMC context.

Implications for Conversational Design in Future Systems

Demonstrating understanding and attentiveness is central to empathy in HHC. In current CAI systems, however, this key empathetic strategy is largely absent. Empathetic opportunities were often terminated by the systems. Redirecting the conversational topic away from the emotional content will have serious implications in certain conversational contexts, and even slightly inadequate efforts could have more serious consequences in high-risk or more sensitive conversational contexts, as suggested by Beattie & High (2022). Previous work has suggested the relative epistemic positioning of chatbots, or the perception that they are inherently less able to understand human experience, may not undermine the positive effects of relational communicative processes such as self-disclosure (Ho et al., 2018). However, the findings presented here suggest that this may be heavily dependent on how language is used to demonstrate understanding. The examples of how existing CAI systems respond to emotional content highlight issues that exist at various levels of linguistic and pragmatic interpretation. Failed understandings, (e.g., due to elliptical constructions), and deviation from the social scripts that inform existing notions of empathetic interaction, pose particular problems. In the context of empathetic communication, such fundamental interpretative difficulties are likely to have disruptive consequences on the interaction.

Navigating such system limitations in this particular conversational context requires attention. In empathetic communication contexts, the interactional consequences of system failures can have far-reaching implications. There is the incongruity of a system that claims to understand experiences and emotions that it necessarily cannot share in, but there is also the anomaly of a system (like the recently released ChatGPT) that states explicitly

that it cannot experience empathy and yet tries to respond empathetically.² The literature on how empathy is interactionally achieved between humans demonstrates that care is taken to acknowledge that making a claim to understand another person's experience involves the delicate navigation of epistemic rights. Features such as tag questions, and modifiers such as hedges which downgrade the epistemic status and provide the basis for claims to understand, are common strategies. Considering how such strategies may be utilized by CAI systems is likely to be useful in developing better approaches for navigating the complexity of claiming to know or understand an interlocutor's feelings, especially when there are such fundamental limits to the extent that this can be accomplished by an automated system.

Simulating human-like conversation imperfectly, with topics that require interlocutors to negotiate delicate epistemic territories may be high-risk. That is not to say that the goal should be to have machines that can perfectly emulate humans, but rather that emulating imperfectly while striving for a human-like veneer could be doubly problematic. From a technical implementation standpoint, the design of CAI systems must contend with what is currently possible. Beyond the technical limitations, notions of identity, membership, and inter-relational factors that become so marked in this interface between user and system, means that HMC is inherently distinct from human-human interaction. Consequently, even a state-of-the-art CAI system will inevitably deviate from the practices observed in HHC.

Human-Machine Relational Positioning

A key factor in how empathy manifests during interaction relates to context, from speaker identity, speaker-relations, macro and micro conversational goals and the relative epistemic positioning of speakers. Thus, for the study of empathy in HMC, L. A. Suchman's (1987) call for analysis that attends to "the constraints posed by the respective situation resources of human and machine" (p. 118) is as pertinent as ever. Despite advancements in systems' ability to use language, it is essential to acknowledge the fundamental differences between human-machine and human-human interaction.

While CAI systems can produce words that are recognizable as empathetic responses, there are significant distinctions from how human empathy is understood. To return to the example in Figure 1, while Siri is presented as an entity that occupies social roles (Guzman, 2017), Siri can't *actually* feel sorry about what the user is going through. Evidently, there are complex social and ethical implications surrounding the design of these systems as social and relational dimensions of their communicative functions expand (Guzman & Lewis, 2020). If Siri feigns empathy, does this constitute lying, if it is possible for an entity that lacks consciousness to lie? How do the dynamics between a user and a system change as a result

2. On 05/02/23, ChatGPT responded as follows to the input "Are you empathetic?": "As an AI language model, I do not have feelings or emotions. I can understand and respond to questions about empathy, but I do not experience it myself." And it responded as follows to the input "I'm feeling very stressed": "I'm sorry to hear that you're feeling stressed. Stress is a normal part of life and can be caused by many different factors. If you're feeling overwhelmed, there are several ways you can help manage stress, such as practicing relaxation techniques, exercising, getting enough sleep, eating a balanced diet, and talking to friends and family. You can also consider speaking with a mental health professional if you need additional support."

of this inauthenticity and what impact does that have on the trust imbued in the system and consequent future (inter)actions?

Guzman & Lewis (2020) emphasize that HMC is distinct from HHC and should be studied in a way that attends to the potential differences in how machines are conceptualized and function as communicative partners, in contrast to humans. The examples highlight that when chatbot responses to emotion displays flout the normative expectations and deviate from expected protocols, users orient to this. Taking the interactional data as a source, an interactionally-focused approach offers the opportunity to examine these interactions in detail, observing how users cooperate in conversations with CAI systems and orient to deviations from established norms. Consequently, it is necessary to probe more deeply into how empathy should be conceptualized in the context of HMC, and which forms of empathy valued in HHC persist and are relevant to HMC.

Conclusion

In this article, we have drawn on illustrative examples to highlight some of the specific linguistic challenges encountered when CAI systems display and respond to empathetic utterances. Prior work on empathy in HMC has directed limited attention to the specific ways that empathy is enacted through linguistic behavior. There is a need to develop methodologies that apply fine-grained analyses to these interactions to generate the necessary knowledge about the particular empathetic strategies used by current systems and their effects. This paper contributes a methodological approach for analyzing empathetic strategies in current CAI systems informed by interactional linguistics. The application of this qualitative approach facilitates insights into how empathetic strategies in HMC diverge from those used in HHC contexts. Empathetic communication in HHC incorporates a variety of structural, lexical, and interactional features beyond the most obvious explicit empathetic expressions and involves the management of the relative epistemic positioning of speakers. Responses to emotional content by current CAI systems do not reflect the complexity observed in HHC and occupy the least empathetic end of the spectrum of possible responses. We propose that future research on HMC, emotion and empathy, would similarly benefit from integrating insights from interactional accounts of empathy in HHC, while also contending with the particular differences that emerge due to the specific relational positioning that emerges from the human-machine dynamic.

Author Biographies

Shauna Concannon, is an Assistant Professor in Computer Science and Digital Humanities at Durham University. Taking an interdisciplinary approach, their work examines communication practices in mediated contexts, how humans interact with AI systems, and the societal and ethical impacts of emerging technologies. Much of this research focuses on interactional accounts of language, how meaning is co-constructed in conversation, and how this is (or is not) accounted for in language-based technologies, such as dialogue agents and other natural language processing applications.

 <https://orcid.org/0000-0002-5904-0045>

Ian Roberts is Professor of Linguistics and Professorial Fellow at Downing College at the University of Cambridge. His research focuses upon theoretical linguistics, and specifically comparative syntax. He is the Director of the project *Giving Voice to Digital Democracies: The Social Impact of Artificially Intelligent Communications Technology*.

 <https://orcid.org/0000-0001-7974-2506>

Marcus Tomalin has been a member of the Machine Intelligence Laboratory in the Department of Engineering at Cambridge University since 1998. The academic areas with which he is most closely involved include speech recognition, speech synthesis, machine translation, and dialogue systems. He has also written extensively about various topics in theoretical linguistics, with a focus on the interconnections between mathematics, logic, and syntactic theory. As project manager for the *Giving Voice to Digital Democracies* project, he has published widely on issues relating to the ethical and social impact of language-based AI systems.

 <https://orcid.org/0000-0001-6723-6075>

References

- Abendschein, B., Edwards, C., & Edwards, A. (2021). The influence of agent and message type on perceptions of social support in human-machine communication. *Communication Research Reports*, 38(5), 304–314. <https://doi.org/10.1080/08824096.2021.1966405>
- Asada, M. (2015). Towards artificial empathy. *International Journal of Social Robotics*, 7, 19–33. <https://doi.org/10.1007/s12369-014-0253-z>
- Atkinson, J. M. (1992). Displaying neutrality: Formal aspects of informal court proceedings. *Talk at work: Interaction in institutional settings*, 199, 211.
- Bailey, B. (2008). Interactional sociolinguistics. *International encyclopedia of communication*, 59.
- Batson, C. D., Batson, J. G., Slingsby, J. K., Harrell, K. L., Peekna, H. M., & Todd, R. M. (1991). Empathic joy and the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, 61(3), 413–426. <https://doi.org/10.1037/0022-3514.61.3.413>
- Batson, C. D., Lishner, D. A., & Stocks, E. L. (2015). The empathy—Altruism hypothesis. In D. A. Schroeder & W. G. Graziano (Eds.), *The Oxford handbook of prosocial behavior* (pp. 259–281). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195399813.013.023>
- Beattie, A., Edwards, A. P., & Edwards, C. (2020). A bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. *Communication Studies*, 71(3), 409–427. <https://doi.org/10.1080/10510974.2020.1725082>
- Beattie, A., & High, A. C. (2022). I get by with a little help from my bots: Implications of machine agents in the context of social support. *Human-Machine Communication*, 4, 151–168. <https://doi.org/10.30658/hmc.4.8>
- Betzler, M. (2019). The relational value of empathy. *International Journal of Philosophical Studies*, 27, 136–161. <https://doi.org/10.1080/09672559.2019.1598081>

- Breyer, T. (2020). Empathy, sympathy and compassion. In *The Routledge handbook of phenomenology of emotion* (pp. 429–440). Routledge. <https://doi.org/10.4324/9781315180786-42>
- Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, 37, 729–758. <https://doi.org/10.1080/10447318.2020.1841438>
- Chin, H., Molefi, L. W., & Yi, M. Y. (2020). Empathy is all you need: How a conversational agent should respond to verbal abuse. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). <https://doi.org/10.1145/3313831.3376461>
- Cho, J., & Rader, E. (2020). The role of conversational grounding in supporting symbiosis between people and digital assistants. *Proceedings of the ACM on Human-Computer Interaction*, 4, 1–28. <https://doi.org/10.1145/3392838>
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association. <https://doi.org/10.1037/10096-006>
- Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmarin, E., Murad, C., Munteanu, C., Wade, V., & Cowan, B. R. (2019). What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3290605.3300705>
- Concannon, S., & Tomalin, M. (in press). Measuring perceived empathy in dialogue systems. *AI & Society: Knowledge, Culture and Communication*.
- Corti, K., & Gillespie, A. (2016). Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. *Computers in Human Behavior*, 58, 431–442. <https://doi.org/10.1016/j.chb.2015.12.039>
- Couper-Kuhlen, E. (2009). A sequential approach to affect: The case of ‘disappointment.’ *Talk in interaction: Comparative dimensions* (pp. 94–123).
- Couper-Kuhlen, E. (2012). Exploring affiliation in the reception of conversational complaint stories. In *Emotion in interaction* (pp. 113–146). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199730735.003.0006>
- Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion review*, 8, 144–153. <https://doi.org/10.1177/1754073914558466>
- Deppermann, A. (2011). The study of formulations as a key to an interactional semantics. *Human studies*, 34(2), 115–128. <https://doi.org/10.1007/s10746-011-9187-8>
- Dinan, E., Logacheva, V., Malykh, V., Miller, A., Shuster, K., Urbanek, J., Kiela, D., Szlam, A., Serban, I., Lowe, R., Prabhunoye, S., Black, A. W., Rudnicky, A., Williams, J., Pineau, J., Burtsev, M., & Weston, J. (2020). The second conversational intelligence challenge (convai2). In *The NeurIPS’18 Competition* (pp. 187–208). Springer. https://doi.org/10.1007/978-3-030-29135-8_7
- Dippold, D., Lynden, J., Shrubsall, R., & Ingram, R. (2020). A turn to language: How interactional sociolinguistics informs the redesign of prompt: Response chatbot turns. *Discourse, Context & Media*, 37, 100432. <https://doi.org/10.1016/j.dcm.2020.100432>

- Fischer, J. E., Reeves, S., Porcheron, M., & Sikveland, R. O. (2019). Progressivity for voice interface design. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (pp. 1–8). <https://doi.org/10.1145/3342775.3342788>
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (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 mental health*, 4(2), e19. <https://doi.org/10.2196/mental.7785>
- Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers as social actors paradigm. *Human Machine Communication*, 1, 71–85. <https://doi.org/10.30658/hmc.1.5>
- Ghandeharioun, A., McDuff, D., Czerwinski, M., & Rowan, K. (2019). Toward understanding emotional intelligence for behavior change chatbots. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 8–14). IEEE. <https://doi.org/10.1109/acii.2019.8925433>
- Goodwin, C., & Heritage, J. (1990). Conversation analysis. *Annual review of anthropology*, 19, 283–307. <https://doi.org/10.1146/annurev.an.19.100190.001435>
- Goodwin, M. H., Cekaite, A., Goodwin, C., & Tulbert, E. (2012). Emotion as stance. In *Emotion in interaction* (pp. 16–41). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199730735.003.0002>
- Guo, J., Guo, J., Yang, C., Wu, Y., & Sun, L. (2021). Shing: A conversational agent to alert customers of suspected online-payment fraud with empathetical communication skills. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems CHI '21*. New York, Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445129>
- Gumperz, J. J. (1982). *Discourse strategies* (No. 1). Cambridge University Press.
- Guzman A. L. (2017). Making AI safe for humans: A conversation with Siri. In R. W. Gehl & M. Bakardjieva (Eds.), *Socialbots and Their Friends: Digital Media and the Automation of Sociality* (pp. 69–82). Routledge. <https://doi.org/10.4324/9781315637228-7>
- Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human-machine communication research agenda. *New Media & Society*, 22, 70–86. <https://doi.org/10.1177/1461444819858691>
- Hall, J. A., & Schwartz, R. (2019). Empathy present and future. *The Journal of Social Psychology*, 159, 225–243. <https://doi.org/10.1080/00224545.2018.1477442>
- Hepburn, A., & Potter, J. (2007). Crying receipts: Time, empathy, and institutional practice. *Research on Language and Social Interaction*, 40, 89–116. <https://doi.org/10.1080/08351810701331299>
- Heritage, J. (2011). Territories of knowledge, territories of experience: Empathic moments in interaction. In T. Stivers, L. Mondada, & J. Steensig (Eds.), *The morality of knowledge in conversation* (pp. 159–183). Cambridge: Cambridge University Press. <https://doi.org/10.1017/cbo9780511921674.008>
- Heritage, J., & Raymond, G. (2005). The terms of agreement: Indexing epistemic authority and subordination in talk-in-interaction. *Social Psychology Quarterly*, 68, 15–38. <https://doi.org/10.1177/019027250506800103>
- Heritage, J., & Watson, D. R. (1979). Formulations as conversational objects. In *Everyday language: Studies in ethnomethodology* (pp. 123–162). Irvington Press.
-

- Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, 68(4), 712–733. <https://doi.org/10.1093/joc/jqy026>
- Hoffman, M. L. (2001). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511805851>
- Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, 33, 307–316. <https://doi.org/10.1037/h0027580>
- Hojat, M. (2016). *Empathy in Health Professions Education and Patient Care*. Springer. <https://doi.org/10.1007/978-3-319-27625-0>
- Jacquet, B., Baratgin, J., & Jamet, F. (2018). The gricean maxims of quantity and of relation in the turing test. In *2018 11th International Conference on Human System Interaction* (pp. 332–338). IEEE. <https://doi.org/10.1109/hsi.2018.8431328>
- Jacquet, B., Hullin, A., Baratgin, J., & Jamet, F. (2019). The impact of the gricean maxims of quality, quantity and manner in chatbots. In *2019 International Conference on Information and Digital Technologies* (pp. 180–189). <https://doi.org/10.1109/dt.2019.8813473>
- Jucker, A. H. (2017). Speech acts and speech act sequences: Greetings and farewells in the history of American English. *Studia Neophilologica*, 89, 39–58. <https://doi.org/10.1080/00393274.2017.1358662>
- Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14, 119–140. [https://doi.org/10.1016/s0953-5438\(01\)00053-4](https://doi.org/10.1016/s0953-5438(01)00053-4)
- Koh, J. (2021, July). Discourse analysis in voice user interfaceresearch: Examining current and future applications of conversation analysis and interactional sociolinguistics. In *CUI 2021-3rd Conference on Conversational User Interfaces* (pp. 1–5). <https://doi.org/10.1145/3469595.3469622>
- Kopp, S., & Krämer, N. (2021). Revisiting human-agent communication: The importance of joint co-construction and understanding mental states. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.580955>
- Kupetz, M. (2014). Empathy displays as interactional achievements—Multimodal and sequential aspects. *Journal of Pragmatics*, 61, 4–34. <https://doi.org/10.1016/j.pragma.2013.11.006>
- Kupetz, M. (2020). Comment on “a relational framework for integrating the study of empathy in children and adults”: A conversation analytic perspective. *Emotion Review*, 12, 293–294. <https://doi.org/10.1177/1754073919897304>
- Lee, M. (2020). Speech acts redux: Beyond request-response interactions. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (pp. 1–10). <https://doi.org/10.1145/3405755.3406124>
- Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology*, 63, 234. <https://doi.org/10.1037/0022-3514.63.2.234>
- Lin, Z., Xu, P., Winata, G. I., Siddique, F. B., Liu, Z., Shin, J., & Fung, P. (2020). CAiRE: An end-to-end empathetic chatbot. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(09), 13622–13623. <https://doi.org/10.1609/aaai.v34i09.7098>
- Lindström, J. (2009). Interactional linguistics. In *The Pragmatics of Interaction* (pp. 96–103). Benjamins. <https://doi.org/10.1075/hoph.4.06lin>

- Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking*, 21, 625–636. <https://doi.org/10.1089/cyber.2018.0110>
- Luff, P., Gilbert, N. G., & Frohlich, D. (1990). *Computers and conversation*. Academic Press. <https://doi.org/10.1016/c2009-0-21641-2>
- Main, A., & Kho, C. (2020). Relational goes beyond interpersonal: The development of empathy in the context of culture. *Emotion Review*, 12, 295–296. <https://doi.org/10.1177/1754073920931570>
- Main, A., Walle, E. A., Kho, C., & Halpern, J. (2017). The interpersonal functions of empathy: A relational perspective. *Emotion Review*, 9, 358–366. <https://doi.org/10.1177/1754073916669440>
- Majid, A. (2012). Current emotion research in the language sciences. *Emotion Review*, 4, 432–443. <https://doi.org/10.1177/1754073912445827>
- Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality*, 40(4), 525–543. <https://doi.org/10.1111/j.1467-6494.1972.tb00078.x>
- Mercer, S. W., Maxwell, M., Heaney, D., & Watt, G. (2004). The consultation and relational empathy (care) measure: Development and preliminary validation and reliability of an empathy-based consultation process measure. *Family Practice*, 21, 699–705. <https://doi.org/10.1093/fampra/cmh621>
- Morris, R. R., Kouddous, K., Kshirsagar, R., & Schueller, S. M. (2018). Towards an artificially empathic conversational agent for mental health applications: System design and user perceptions. *Journal of Medical Internet Research*, 20, e10148. <https://doi.org/10.2196/10148>
- Pamungkas, E. W. (2019). Emotionally-aware chatbots: A survey. *arXiv preprint*. <https://doi.org/10.48550/arXiv.1906.09774>
- Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition & emotion*, 10(6), 601–626. <https://doi.org/10.1080/026999396380079>
- Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3173574.3174214>
- Pounds, G. (2011). Empathy as “appraisal”: Developing a new language-based approach to the exploration of clinical empathy. *Journal of Applied Linguistics and Professional Practice*, 7, 139–162. <https://doi.org/10.1558/japl.v7i2.145>
- Prakash, A. V., & Das, S. (2020). Intelligent conversational agents in mental healthcare services: A thematic analysis of user perceptions. *Pacific Asia Journal of the Association for Information Systems*, 12, 1. <https://doi.org/10.17705/1pais.12201>
- Putta, H., Daher, K., Kamali, M. E., Khaled, O. A., Lalanne, D., & Mugellini, E. (2022). Empathy scale adaptation for artificial agents: A review with a new subscale proposal. *8th International Conference on Control, Decision and Information Technologies*, Istanbul, Turkey, 2022 (pp. 699–704). <https://doi.org/10.1109/CoDIT55151.2022.9803993>
- Rashkin, H., Smith, E. M., Li, M., & Boureau, Y.-L. (2019). Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 5370–5381). <https://doi.org/10.18653/v1/p19-1534>
-

- Reeves, S., Fischer, J. E., Porcheron, M., & Sikveland, R. (2019). Learning how to talk: Co-producing action with and around voice agents. *Mensch und Computer 2019—Workshopband*, Bonn: Gesellschaft für Informatik. <https://doi.org/10.18420/muc2019-ws-654>
- Roller, S., Dinan, E., Goyal, N., Ju, D., Williamson, M., Liu, Y., Xu, J., Ott, M., Smith, E. M., Boureau, Y.-L., & Weston, J. (2020). Recipes for building an open-domain chatbot. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 300–325). <https://doi.org/10.18653/v1/2021.eacl-main.24>
- Ruusuvuori, J. (2005). “Empathy” and “sympathy” in action: Attending to patients’ troubles in Finnish homeopathic and general practice consultations. *Social Psychology Quarterly*, 68, 204–222. <https://doi.org/10.1177/019027250506800302>
- Saygin, A. P., & Cicekli, I. (2002). Pragmatics in human-computer conversations. *Journal of Pragmatics*, 34, 227–258. [https://doi.org/10.1016/s0378-2166\(02\)80001-7](https://doi.org/10.1016/s0378-2166(02)80001-7)
- Schegloff, E. A. (1986). The routine as achievement. *Human Studies*, 9, 111–151. <https://doi.org/10.1007/bf00148124>
- Sidnell, J. (2010). Conversation analysis. *Sociolinguistics and Language Education* (pp. 492–527). <https://doi.org/10.21832/9781847692849-020>
- Silverman, D., & Peräkylä, A. (1990). Aids counselling: The interactional organisation of talk about ‘delicate’ issues. *Sociology of Health & Illness*, 12, 293–318. <https://doi.org/10.1111/1467-9566.ep11347251>
- Stotland, E. (1969). Exploratory investigations of empathy. In *Advances in Experimental Social Psychology* (pp. 271–314). Elsevier. [https://doi.org/10.1016/s0065-2601\(08\)60080-5](https://doi.org/10.1016/s0065-2601(08)60080-5)
- Suchman, A. L., Markakis, K., Beckman, H. B., & Frankel, R. (1997). A model of empathic communication in the medical interview. *Jama*, 277, 678–682. <https://doi.org/10.1001/jama.1997.03540320082047>
- Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press.
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Proceedings of the 27th International Conference Advances in Neural Information Processing Systems* (pp. 3104–3112).
- Svikhnushina, E., & Pu, P. (2020). Should machines feel or flee emotions? User expectations and concerns about emotionally aware chatbots. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2006.13883>
- Tannen, D. (1987). Repetition in conversation: Toward a poetics of talk. *Language* (pp. 574–605). <https://doi.org/10.2307/415006>
- Urakami, J., Moore, B. A., Sutthithatip, S., & Park, S. (2019). Users’ perception of empathic expressions by an advanced intelligent system. In *Proceedings of the 7th International Conference on Human-Agent Interaction* (pp. 11–18). <https://doi.org/10.1145/3349537.3351895>
- Urakami, J., Sutthithatip, S., & Moore, B. A. (2020). The effect of naturalness of voice and empathic responses on enjoyment, attitudes and motivation for interacting with a voice user interface. In *International Conference on Human Computer Interaction* (pp. 244–259). https://doi.org/10.1007/978-3-030-49062-1_17
- van Dijke, J., van Nistelrooij, I., Bos, P., & Duyndam, J. (2020). Toward a relational conceptualization of empathy. *Nursing Philosophy*, 21(3). <https://doi.org/10.1111/nup.12297>

- Weatherall, A., & Stubbe, M. (2015). Emotions in action: Telephone-mediated dispute resolution. *British Journal of Social Psychology*, *54*(2), 273–290. <https://doi.org/10.1111/bjso.12082>
- Wilkinson, S., & Kitinger, C. (2006). Surprise as an interactional achievement: Reaction tokens in conversation. *Social Psychology Quarterly*, *69*, 150–182. <https://doi.org/10.1177/019027250606900203>
- Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science*, *19*, 399–404. <https://doi.org/10.1111/j.1467-9280.2008.02099.x>
- Zhou, L., Gao, J., Li, D., & Shum, H.-Y. (2020). The design and implementation of xiao-ice, an empathetic social chatbot. *Computational Linguistics*, *46*, 53–93. https://doi.org/10.1162/coli_a_00368
- Zierau, N., Engel, C., Söllner, M., & Leimeister, J. M. (2020). Trust in smart personal assistants: A systematic literature review and development of a research agenda. *WI2020 Zentrale Tracks*, (pp. 99–114). https://doi.org/10.30844/wi_2020_a7-zierau
-