# Journal Pre-proof


Exploring how politeness impacts the user experience of chatbots for mental health support

Robert Bowman, Orla Cooney, Joseph W. Newbold, Anja Thieme, Leigh Clark, Gavin Doherty, Benjamin Cowan

PII: S1071-5819(23)00190-8
DOI: https://doi.org/10.1016/j.ijhcs.2023.103181
Reference: YIJHC 103181

To appear in: *International Journal of Human - Computer Studies*

Received date : 3 July 2023
Revised date : 29 September 2023
Accepted date : 6 November 2023

Please cite this article as: R. Bowman, O. Cooney, J.W. Newbold et al., Exploring how politeness impacts the user experience of chatbots for mental health support. *International Journal of Human - Computer Studies* (2023), doi: https://doi.org/10.1016/j.ijhcs.2023.103181.

# Exploring how Politeness Impacts the User Experience of Chatbots for Mental Health Support

Robert Bowman[a], Orla Cooney[b], Joseph W. Newbold[c], Anja Thieme[d], Leigh Clark[e], Gavin Doherty[a], Benjamin Cowan[b]

[a]School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland
[b]School of Information and Communication Studies, University College Dublin, Dublin, Ireland
[c]Northumbria University, Newcastle-upon-Tyne, United Kingdom
[d]Microsoft Research, Cambridge, United Kingdom
[e]Bold Insight, London, United Kingdom

**Abstract**

Politeness is important in human-human interaction when asking people to engage in sensitive conversations. If politeness manifests similarly in human-chatbot interaction, it may play an important role in the design of sensitive chatbot interactions such as those for providing mental health support. Our mixed methods study (N=39) contributes findings on how the use of politeness by chatbots, for the mental healthcare activity of mood logging, is perceived by users. Our study combined a within-participants controlled experiment, whereby participants interacted with three prototype chatbots differing in their use of politeness, with semi-structured interviews. Our analysis demonstrates that a chatbot's use of politeness can impact how a participant experiences interacting with it, both positively and negatively. While politeness can be experienced as caring, supportive, and encouraging, it can also be experienced as overly apologetic, condescending, and untrustworthy. We discuss the nuances of using politeness in conversational interaction design, setting out a research agenda for polite conversational interaction.

*Keywords:* Chatbots, Conversational User Interface, Politeness, Mental Health

## 1. Introduction

One in four people worldwide are likely to experience difficulties with their mental health in their lifetime [77]; yet, many do not receive sufficient support due to a combination of attitudinal and structural barriers [1]. Accordingly, there is a pressing need for new methods of treatment that account for these barriers; a key strategy identified to address this pressing need is the increased use of digital technology [41, 58].

Conversational agents, such as text-based chatbots, are a promising technology for supporting people with their mental health. Given these interfaces present as an agent with which the user can interact in a manner resembling human-human conversation, there is potential for them to mimic, and deliver at scale, established therapist-delivered, talk-based support [75, 23, 11].

It is imperative that all forms of mental health support are responsibly designed, due to the highly sensitive nature of mental health. By the nature of experiencing difficulties with their mental health, the users of mental health support can be vulnerable. In this context, insensitive interactions risk not only a negative user experience, but harm to the vulnerable person's well-being [25, 59]. Currently it is unclear how best to design effective chatbot interactions [63], especially for sensitive activities.

Our research investigates how politeness can be used to inform the design of chatbot interactions for sensitive activities, such as those for mental health support. Politeness is generally understood to be about behaving appropriately in social interactions, with appropriateness being socioculturally determined [73]. Polite behaviour is generally described as behaviour with the goal of establishing or maintaining the relationships between individuals within a social interaction [73, 46, 15, 20]. It has been identified that politeness within human-chatbot interaction design is important for the user experience, yet it remains unclear how to effectively leverage politeness within design [38, 18, 79, 62]. Research has recommended that we study how the politeness strategies of human-human interaction can be used within human-chatbot interaction design [18, 62]. Our research takes this approach and investigates how the use of a prominent theory of politeness in human-human interaction [15] can be used to support the design of sensitive human-chatbot interactions. This theory describes politeness as being behaviour used to maintain an individual's own feelings of self-worth and autonomy, as well as the feelings of those that they are interacting with. The challenge of maintaining a person's feelings of self-worth has congruence with the challenge of designing sensitive human-chatbot interactions, as sensitive interactions risk impacting the user's feelings of self-worth. Similarly, the challenge of maintaining a person's feelings of autonomy relates to the design of chatbot interactions, which users may experience as inconvenient and thus autonomy impacting. The theory describes approaches for polite interaction; if they work similarly in human-chatbot interaction to how they are theorised to work in human-human interaction, they could be harnessed within the design of chatbot interactions. Little is currently known, however, about how the use of these politeness approaches could impact the user experience of interacting with a chatbot.

In this work we explore how the use of politeness by chatbots can impact the user experience for the activity of mood logging. Mood logging is a self-report activity that is commonly used to support people with their mental health; it involves the user tracking their mood and associated data in order to develop self-awareness, as well as to inform and monitor treatment progress [51]. We conducted a mixed-methods study that combined a within-participants controlled experiment and semi-structured interviews. 39 participants enacted mood logging scenarios with three prototype chatbots. Informed by a theory of politeness [15], the chatbots were designed to use different politeness approaches. One chatbot was designed to use *direct* language; it was included in the study to serve as a non-polite comparison. The other two chatbots were designed to use two different forms of polite language, with one making considerable use of personal pronouns (*Personal* chatbot) and the other not making use of such pronouns (*Passive* chatbot).

Our analysis demonstrates that the use of politeness had a meaningful impact upon how participants experienced interacting with the chatbots. The use of politeness was a reason to expect positive impacts, however our analysis concluded that the impacts could be both positive or negative. The Personal chatbot was perceived as caring; while many participants found this supportive and encouraging, for some participants it incited distrust. The Passive chatbot was perceived as too apologetic, some experienced this chatbot as being condescending. Our analysis suggests that there is opportunity for Personal politeness to support sensitive chatbot interactions such as those for mood logging. Furthermore, the analysis highlights the risks posed by insensitive interactions, as a user feeling condescended to by their mental health support tool could have a negative impact on their wellbeing.

Our work illustrates that politeness can both positively and negatively impact the user experience of interacting with a chatbot for mood logging. We discuss how these impacts may be explained by existing theory. Specifically, our analysis makes clear that both the type of politeness and interaction context matter. To harness the potential of polite chatbots to support activities requiring sensitive interaction, like those for mental health support, we need to be nuanced in our application of politeness theory by adapting it to the characteristics of human-chatbot interaction. Toward this, we contribute a research agenda for embedding politeness into the design of conversational interfaces.

## 2. Related Work

### *2.1. Chatbots for Mental Health Support*

The idea of using conversational agents to provide mental health support to the masses is long standing and enduring. In the 1960's research on natural language processing produced the ELIZA system which could imitate a Rogerian therapist [75]. It was observed that people who interacted with the program, including those fully aware that it was simply a computer program, were willing to self-disclose intimately to and anthropomorphise it [76]. These observations were an early indicator that conversational agents may be able to mimic therapist-delivered talk-based support, and thus have the potential to provide some form of digital mental health support. In response to ELIZA, speculation sparked about the potential benefit of computer programs capable of providing psychotherapeutic support [23, 76]:

> "If the method proves beneficial, then it would provide a therapeutic tool which can be made widely available to mental hospitals and psychiatric centres suffering a shortage of therapists" [23, p.152]

With there still being a persistent gap in mental health treatment provision, and with recent developments in conversational agent technology, the prospect of using conversational agents for mental health support remains alluring. Chatbots for mental health support, such as WoeBot [30] and Wysa [37], are now available to the public and further research has advanced our understanding of the potential of this technology.

#### *2.1.1. Self-Disclosure to Computers*

Key to mental health support, especially mood logging, is the self-disclosure of moods, thoughts, and experiences. Research has investigated self-disclosure to computers, particularly in comparison to self-disclosure to people. Lucas et al. [48, 32] conducted an experiment to compare people's willingness to disclose between a human-operated, and a fully-automated, virtual human (a speech-based human-looking avatar) in the context of a health-screening interview. During the study, participants were asked sensitive and intimate questions such as:

- "Tell me about an event, or something that you wish you could erase from your memory"
- "Tell me about the last time you felt really happy".

Their analysis shows participants were more willing to disclose to the fully-automated virtual human, that is when they thought they were not being observed by another human, than to the human-operated virtual human. An explanation for the greater willingness to disclose to a computer than to a human, supported by anecdotal remarks from participants, is that the computer is perceived as non-judgemental making it a preferable partner to disclose to.

Ho et al. [35] compared the effects on participants of self-disclosure to people with those of self-disclosure to computers. They conducted a Wizard of Oz experiment using a text-based chat interface whereby participants were told they were either interacting with a computer agent or another person. The analysis found that the effects of disclosure to a computer were equivalent to the effects of disclosure to a person across a range of psychological, relational, and emotional measures. The authors note, however, that participants of this study, unlike in the study by Lucas et al. [48, 32], were not asked to disclose about potentially embarrassing or stigmatized topics. For these more sensitive disclosures, people's fears of being judged might be heightened and consequently the effects of disclosure to people and to computers may not be equivalent.

Kim et al. [39] explored teenagers perceptions on chatbots for emotional support through workshops and interviews. Echoing the idea of computers being non-judgemental, this study found that participants perceived chatbots as good listeners. Unlike people, computers can be relied upon to always be available to listen about anything and respond in a patient and empathetic way. Furthermore, participants felt computers could be trusted to not reveal their secrets, particularly more so than their friends. Another perceived advantage of self-disclosure to computers is that computers can offer data-informed advice to support the user with the challenges they are experiencing.

In summary, research suggests that people are willing to self-disclose to conversational interfaces. More so, the perception of computers as non-judgemental and confidential may in some situations make self-disclosure to computers preferable to self-disclosure to other people.

#### *2.1.2. Anthropomorphism*

While computers not being human is an opportunity for self-disclosure, computers being human-like may also be an opportunity to construct supportive human-computer relationships. Research has accordingly investigated supportive and therapeutic human-computer relationships. Bickmore and Cassell [5] developed a model of social dialogue, implemented using an embodied conversational agent, that was designed to build a trusting relationship with the user. Embodied conversational agents using this model of social dialogue are often referred to as relational agents. As noted by the authors, a motivation for building user-trust is that trust is a prerequisite for risky interactions such as disclosing personal information. In this first study [5], the authors demonstrated that the use of the social dialogue model increased trust within more extroverted users. In further work on relational agents, Bickmore and Picard [9] investigated how this use of relational skills, including empathy and social dialogue, may allow for user-computer relationships to develop that are similar in character to effective client-therapist relationships. This study compared the use of a relational and non-relational embodied conversational agent - Laura - designed for health behaviour change (to promote physical activity). The analysis showed that the use of relational behaviours increased the user's experience of working alliance, which is a construct used in psychotherapy to evaluate client-therapist relationships; the analysis also showed participants had a greater desire to continue working with the relational version. For talk-based therapy the client-therapist relationship is understood to be a key determinant of the therapeutic outcome [24]. The potential to develop human-computer relationships similar in character to effective client-therapist relationships, is therefore promising for the potential of chatbots for mental health support.

The extent to which conversational computer interfaces should be anthropomorphised is, however, a topic of debate. Although humanness is core to the interaction metaphor of conversational computer interfaces [28], human-like design can lead to unrealistic user expectations, creating a mismatch between the perceived and actual capabilities of these interfaces [49]. The management of user expectations is particularly critical for mental health applications where mismatches risk not only a poor user experience, but also negatively impacting the user's well-being [11]. Using a qualitative interview study, Clark et al. [22] investigated what characteristics people see as important in conversation with people, and how they vary in interaction with conversational agents. The analysis showed that participants were sceptical of the need for relational interaction with conversational agents and they instead favoured task-oriented, transactional interaction. Similarly, the use of social forms of dialogue can be perceived as 'fake' by users [28].

Overall the suitability of anthropomorphic design for chatbots for mental health support is currently unclear. It is potentially beneficial in that it could promote human-computer relationships that cultivate therapeutic support. Yet, these potential benefits may be outweighed by user preference for more task-oriented interaction, as well as the potential negative impacts of misaligned user expectations.

#### *2.1.3. Mental Health Support and Mood Logging*

To help them with their mental health, people are often advised, or choose, to engage with mental health tracking activities; these activities typically require the user to regularly record their mood, related factors (e.g., sleep and exercise), and life experiences. These mood logging activities are, for example, a common component of cognitive behavioural therapy (CBT) [51]. The efficacy of these activities is often limited by attrition and inaccurate data collection [51, 26].

While users know that mood logging activities have the potential to support them with mental health in the long term, they may not want to engage with these activities in the moment [12]. First, these activities can be inconvenient for the user as they are potentially disruptive to the user's everyday life. Second, although people may be generally willing to self-disclose to computers, such disclosure can still be challenging for the user due the sensitive nature of one's mental health and the associated stigma. To harness the potential of chatbots for mental health support, there is a need to understand how to design interactions that account for the inconvenience and sensitivity of the activity, in order to more effectively engage users [12, 11].

### *2.2. Politeness Theory*

Our research focuses on an established theory of politeness in human-human interaction by Brown and Levinson [15] that has the potential to support the design of sensitive and inconvenient interactions. This politeness theory views the purpose of politeness as being to manage social interactions that may threaten a participant's self-image or autonomy. This theory has relevance to chatbot interactions for mental health support because the sensitivity of these interactions may threaten the user's self-image, and the inconvenience of the interactions may threaten the user's autonomy. The theory describes linguistic strategies to manage these threats within human-human interaction. If these strategies work similarly in human-chatbot interaction, they could be used to design effective chatbot dialogues for mental health support. We next present a summary of this theory and its applicability to the design of chatbots for mental health support.

Politeness theory [15] builds on Goffman's theory of face [31] whereby in social interaction each participant has an expressed self-image which has a social value termed *face*. During interaction participant's behaviours can affect each other's face. Two principles that structure social interaction are:

**Self-respect** people want their expressed self-image (face) to be maintained,

**Considerateness** people do not want other's expressed self-image (face) to be undermined.

Accordingly, people in social interaction perform behaviours in an attempt to maintain their own, and other participants', expressed face; such behaviours are the enactment of politeness [31]. Politeness theory considers face as composed of two parts:

**Positive Face** a person's self-worth and want to be seen as desirable by others,

**Negative Face** a person's autonomy and want to not be impeded by others.

The theory perceives many social actions as intrinsically threatening to participants' faces (i.e., these actions could have detrimental impacts upon a participant's face); it terms such actions as face-threatening acts (FTAs).

Toward maintaining face, politeness theory describes strategies for attempting to mitigate the face-threatening effects of social actions. The theory presents 5 high-level approaches to mitigating the effects of an FTA which, in order of decreasing strength, are:

1. Not performing the act,
2. Performing the act *off-record*,
3. Performing the act with the redressive action of *negative politeness*,
4. Performing the act with the redressive action of *positive politeness*,
5. Performing the act *bald-on-record*

When an act is performed off-record its meaning is ambiguous such that it is unclear if a threat to face exists. The redressive approaches of negative and positive politeness attempt to counteract face-threatening effects by signaling that a threat is not intended or desired. Positive and negative politeness attempt this by tending, respectively, to the positive and negative face of the persons threatened. Both the positive and negative politeness approaches are composed of a set of linguistic strategies. The approach of performing the act bald-on-record is where the act is performed in a direct, clear, and concise manner without redress.

The challenges of engaging users with chatbot interactions for mental health support can be interpreted using this theory of politeness. The request for interaction with a chatbot for mental health support could be viewed as impeding the user and lowering their feelings of autonomy, and therefore this request could be viewed as a negative FTA. As interactions regard the user's mental health, which is a sensitive and stigmatised topic, they could be viewed as impacting the user's feelings of self-worth, and therefore these interactions can also be viewed as positive FTAs. It follows from this interpretation that chatbots for mental health support could be designed to mitigate these FTAs using the approaches, such as positive and negative politeness, described by politeness theory. This FTA mitigation could improve the user experience of activities like mood logging and, consequently, promote engagement.

Politeness theory, however, is a theory of human-human, not human-computer interaction. FTAs, and FTA mitigation strategies, may not work in human-chatbot interaction the same way as they are theorised for human-human interaction. What's more, the application of mental health support may further complicate the use of politeness by a chatbot. As discussed by Lakoff [43], therapeutic discourse is not an ordinary conversation setting. There are expectations for therapeutic conversations of the clients disclosing to the therapist, the therapists asking questions of the client, and therapist offering interpretation of the clients disclosures. These expectations impact what uses of politeness people view as appropriate within therapeutic discourse [43]. To understand if, and how, politeness can be used to design effective chatbot dialogues for mental health support, there is a need to understand how politeness theory manifests in human-chatbot interaction.

### 2.3. Politeness in Conversational Human-computer Interaction

#### 2.3.1. Politeness

The importance of politeness for human-chatbot interaction has been stressed by recent work. A 2021 review on social characteristics in human-chatbot interaction by Chaves and Gerosa [18] identified the characteristic of *manners*. This work built on that of Morrissey and Kirakowski [53] who identified manners as a characteristic for evaluating the naturalness of chatbot interactions. Chaves and Gerosa draw on the definition of manners provided by Morrissey and Kirakowski of "the ability of chatbots to display *polite* behaviour and conversational habits" [53, p. 93 emphasis added]. Further in line with Morrissey and Kirakowski, Chaves and Gerosa report that the main benefit of providing manners is to increase human-likeness. Similarly the review notes that, from an analysis of first-time user's chatbot interactions, Jain et al. [38] found that users expect human-like conversational etiquette from chatbots. Referring to Brown and Levinson's politeness theory [15], Chaves and Gerosa describe that a challenge of manners, or politeness, in human-chatbot interaction is the complexity of identifying when there is a FTA that the chatbot should attempt to mitigate. They conclude that that:

> "adoption of politeness strategies to deal with face-threatening acts is still under-investigated in the chatbot literature" [18, p.742]

Research on the design of collaborative human-agent interactions has led to recommendations for the design of polite interactions. These works have built on ideas of joint activity, which Bowman et al. [12] have previously argued are important for the design of chatbots for activities like mood logging. Cila [19] reviewed literature to identify the qualities of human-agent collaborations and to elicit design considerations; they identified the use of politeness theory as a promising design opportunity. Rapp et al. [62] conducted a qualitative analysis of over 1000 conversations with a customer service chatbot to understand how to design for more collaborative human-chatbot interaction. Their analysis demonstrates that people behave in polite ways by being patient with, and by not behaving in negative ways (e.g., aggressively) toward, the chatbot. The authors recommend designing chatbots that nurture a friendly and polite conversational environment. Rapp et al. [63] conclude, explicitly echoing the conclusions of Chaves and Gerosa [18], that:

> "there is a lack of studies on how to leverage the politeness and friendliness strategies used in human-human social interactions... such design strategies could contribute to making a more comfortable conversation environment and preventing feelings that may cause users to abandon the interaction" [62].

Of particular relevance to our research on the design of chatbots for mental health support is a study by Yang and Aurisicchio [79] that investigates a self-determination theory approach to the design of conversational agents. Self-determination theory (SDT) is a theory of human motivation that argues that the fulfilment of people's psychological needs - specifically feelings of autonomy, competence, and relatedness - impacts their motivation and wellbeing. SDT's consideration of people's feelings of autonomy resembles politeness theory's consideration of negative face (i.e., feelings of autonomy), while SDT's considerations of competence and relatedness overlap with politeness theory's consideration of positive face (i.e., feelings of self-worth). Congruent with the shared aims of self-determination and politeness theory, this investigation into a SDT approach for the design of conversation agents produced the guideline:

> "Talk Politely: Encourage polite and socially appropriate conversation style." [79]

This design guidance is, however, vague, providing little detail for how to use politeness within dialogue design.

In sum, research on human-chatbot interaction is identifying that polite behaviour by chatbots meaningfully impacts the user experience, and subsequently recommends that we design for this behaviour. Yet, it remains unclear how to effectively design polite chatbot behaviour. Motivated by this, prior research has called for studies of how the politeness approaches of human-human interaction can be used within chatbot design [18, 62].

*2.3.2. Politeness Theory*

Studies have investigated the use of approaches based on politeness theory [15]by computers in human-computer interaction. Here we present an overview of relevant findings.

Research suggests that politeness can improve user perceptions of computers as interaction partners. Wang et al. [72] compared a polite (mixture of off-record, positive and negative politeness approaches) with a direct virtual pedagogical robot designed to support students learning to use an engineering simulation software. The analysis shows that the polite system was preferred by students and had better learning outcomes. Torrey et al. [69] investigated the use of hedges which are a form of negative politeness. Hedges introduce uncertainty into requests so that it is not assumed that the receiver can fulfill them, thus reducing the force of the request and thereby reducing the FTA. An experiment was conducted that involved comparing people's perceptions of a speech-based domestic robot helper using, and not using, hedges for the task of supporting a novice human to make cupcakes. Robots using hedges were perceived as more considerate and were liked more.

In a study with older adults, Hu et al. [36] compared a polite (mixture of positive and negative politeness) with a direct (bald-on-record) voice- and touch-interaction display for health support. Quantitatively they observed negligible difference between the polite and direct systems. However, their qualitative results suggest most participants preferred the polite version, others preferred the direct version, and some participants were indifferent. This analysis suggests that perceptions of polite behaviour by computers can differ between users.

User perceptions also differ between the politeness approaches used by a computer as found by the work of Miyamoto et al. [52]. This work compared user perceptions of driving support robots using positive and negative politeness approaches. It was observed that most (two thirds of) participants rated the robot using positive politeness favourably, others (one sixth of) rated the robot using negative politeness favourably, and the rest (one sixth of) rated the robots equally. The study also found that the robot using positive politeness was perceived as significantly more anthropomorphic than the robot using negative politeness.

As with user perceptions of anthropomorphic behaviour [47], user perceptions of different politeness strategies may be impacted by the user's orientation toward social behavior by computers. This is suggested by the work of Lee et al. [45] that compared perceptions of a malfunctioning robotic vacuum cleaner using either positive, negative, or no politeness. They observed that the perceptions of users with a more social orientation were more sensitive to the use of politeness by the robot than those of users with a more utilitarian orientation.

Lee et al. [44] studied perceptions of a polite (mixture of positive and negative politeness) and plain (stated instructions directly) virtual driving assistant in simulated normal and failure driving situations. Their results indicate that politeness improves the user experience in normal situations, but impedes it during failure situations. Similarly, Salem et al. [65] found interaction context (a social or goal-oriented interaction) had a greater impact on user's perceptions of a robot receptionist than the robot's use of politeness (direct (bald-on-record) or positive politeness).

These findings indicate that perceptions of the appropriateness of politeness can be affected by the interaction context.

Research has also considered the potential effects of politeness such as persuasiveness. Hammer et al. [33] compared ratings of persuasiveness and politeness for different utterances finding that politeness and persuasiveness are perceived as different constructs. Srinivasan and Takayama [67] studied how the use of politeness by help-seeking robots would affect how people assisted the robots when requested. They compared direct (bald-on-record), positive politeness, negative politeness, and in-direct (off-record) help requests. The results indicated positive politeness was the most effective. With statistical significance, participants were more willing to help the robot using positive politeness than that using direct language and perceived positive politeness as more appropriate than both direct and indirect language. However, the results did not show statistically significant differences between positive and negative politeness. In a virtual reality setting, Zojaji et al. [82] studied how effective the use of different verbal and non-verbal politeness strategies by a virtual agent were at persuading participants to join a small group conversation. The verbal strategies included direct (bald-on-record), positive politeness, negative politeness, and in-direct (off-record). They found the direct and redressive strategies were similarly persuasive, however the direct strategies were rated less favourable on measurements relating to face loss. The results suggests positive politeness best balances persuasiveness and considerateness.

Although work has compared the different approaches described by politeness theory, using these approaches in isolation is not straight forward. Terada et al. [68] attempted to validate scripts using positive politeness, negative politeness, and off-record approaches; they found the script using negative politeness was not perceived as using negative politeness to a higher degree than other politeness forms. Similarly, when Miyamoto et al. [52] had experts evaluate utterances designed to use either positive or negative politeness; some utterances were classified by the experts as using both forms of politeness. Given the difficulty of isolating different politeness approaches, research should proceed cautiously

Research has also studied the use of politeness theory less centrally. Bickmore and Cassell [5] used politeness theory to inform the design of social talk strategies which were identified as a way to develop the user's trust in the system. This research was motivated by the understanding that trust is a prerequisite for risky interactions, such as self-disclosure. It found that within extroverted users the social talk strategies did increase user-computer trust and bond.

To summarise, these studies we have described demonstrate that the use of politeness approaches by computers can impact the user experience. The impacts of politeness appear to affected by the type of politeness used, user's interaction preferences, and characteristics of the interaction task. These studies have largely investigated robots [52, 67, 69, 45, 33, 65] and embodied virtual agents [5, 72, 82]. A couple of studies [36, 44] have considered voice assistants. How the use of politeness by text-based chatbots is perceived has not yet been a research focus.

### *2.4. Research Gap*

In human-human conversation, a use of politeness is to support interactions that risk inconveniencing, or being sensitive to, a conversation partner [15]. If politeness manifests similarly in chatbot interaction, it could be used within dialogue design to support engaging users with potentially sensitive and inconvenient interactions. Although there is a nascent body of research investigating polite conversational interaction, it is currently unclear how users will perceive the use of politeness by chatbots for mental health support, such as those for mood logging. Accordingly, this study was designed to address the primary research question of:

**RQ1** In the context of mood logging, how does the use of politeness by chatbots affect how they are perceived by users?

In addition, we used the opportunity of the study to explore the secondary research question of:

**RQ2** What opportunities and challenges do people perceive in using chatbots for mood logging?

### 3. Designing Polite Chatbots for Mood Logging

We created three prototype chatbots for mood logging, informed by politeness theory [15], that varied in their use of politeness. To design the chatbots we first identified a set of mood logging queries (Table 1), based on the mood logging tool of an established online mental health platform[1]. These queries, in conjunction with log initiation and completion, gave the basic structure of the mood logging chatbot interactions. We modified language, using different politeness strategies, to create the prompts for the three chatbots which are presented by Table 2.

| Query | Question |
|---|---|
| Mood rating | Rate your mood from 1-5 |
| Mood Situation | Did a situation make you feel this way? |
| Related Factor | How was your sleep last night? |
| Diary Entry | What have you done today? |

Table 1: Base Mood Logging Queries

The first chatbot was termed **Direct**. Similarly to prior work, this chatbot was designed using politeness theory's bald-on-record strategy; it uses direct language and was included to serve as a non-polite comparison. These prompts are largely the same as the base queries.

The second and third chatbots, termed **Passive** and **Personal**, were designed to be polite. The Passive chatbot was designed using politeness theory's redressive action approach of negative politeness. The Personal chatbot was designed using politeness theory's redressive action approach of positive politeness. We chose to use the terms Passive and Personal, rather than Negative and Positive, for two reasons. First, we wanted to emphasise that the Personal chatbot makes use of personal pronouns (e.g., '**I** would love to', '**We** could', 'Would help **me**'), which are innately anthropomorphic, and the Passive does not. Given the apparent importance of anthropomorphism in the mental health domain (Section 2.1.2), this characteristic could be impactful. Second, as discussed in Section 2.3.2, prior work [68, 52] has found shown it is difficult to use positive and negative politeness in isolation. By not using the terms Negative and Positive we want to acknowledge that people may not perceive the Passive chatbot as exclusively using negative politeness and the Personal as exclusively using positive politeness.

We created interactive prototypes in the style of a Microsoft Teams chatbot[2]. Figure 1 presents an image of one of the prototypes. The design of the chatbots was kept minimal; each chatbot was named 'MoodBot' and the profile image was left as the default blank avatar. The prototypes only varied by the chatbot prompts.

[1] https://www.silvercloudhealth.com

[2] We used the tool BotSociety which has since been discontinued.

Table 2: Prompt sequences created by by applying each politeness approach to a sequence of mood logging queries.

| | Chatbot | | |
|---|---|---|---|
| **Prompt** | **Direct** | **Passive** | **Personal** |
| **Log Initiation and Mood Rating** | Rate your mood on a scale of 1-5. | Sorry to bother you, but would you mind logging your mood on a scale of 1-5 please? | *Message 1*: Great to see you! Let's get logging. *Message 2*: I would love to know your mood on a scale of 1 - 5 please? |
| **Mood Situation** | Did a situation make you feel this way? If so, describe the situation. | *Message 1*: Logging your mood can be hard so well done for doing it! *Message 2*: If a situation made you feel this way, it would be useful to write about it. | *Message 1*: Thank you for sharing that with me. I appreciate that mood logging can be hard. *Message 2*: It would be helpful for me to know about a specific situation, if there is one, that made you feel this way. |
| **Diary Entry** | What have you done today? | Could you write about what you've done today? | You'll tell me about what you've done today, I hope? |
| **Related Factor** | How was your sleep last night? | One last thing, could you write a little about how you slept last night? | I'm curious to know how you slept last night. |
| **Log Completion** | Log entry complete. | Log entry complete. That wasn't too hard was it. You should log a quick entry again soon. | Thank you. We have finished the log, I hope to speak to you again soon. |

## 4. Method

To study how people perceive the use of politeness by chatbots, with focus on the sensitive application of mood logging, we conducted an exploratory scenario-based within-participants study with 39 participants. We used a mixed-method analysis approach that combined a quantitative component which allowed us to test for anticipated effects, with a qualitative component which enabled us to gain an understanding as to why, and how, these effects may occur, as well as to explore unanticipated effects. The study involved, in a (online) laboratory setting, a controlled experiment during which every participant enacted a fictional mood logging scenario with each of the three chatbots. Semi-structured interviews followed the experiment. The controlled nature of this study design enabled us to effectively compare how participants experienced interacting with each chatbot. The study was given ethical approval for low risk projects by the university

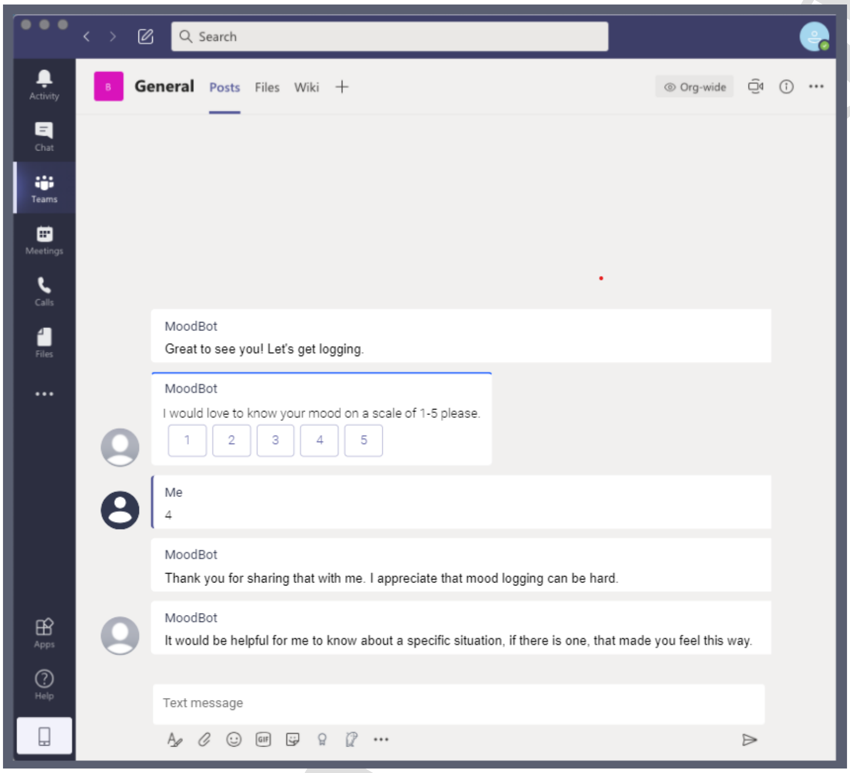


Figure 1: Image of Chatbot Prototype

ethics committee.

### *4.1. Participants*

The study's 39 participants (22 women, 14 men, 2 non-binary, 1 gender fluid[3]) were young adults (age range: 18-35; Mean age= 24.4yrs; SD age=4.2yrs) who were predominantly in university education. Participants came from a diverse set of nationalities, with almost half of the participants reporting their nationality as Irish (41%, N=16). 36 participants self-identified their English fluency as fluent, 2 as high proficiency, and 1 as conversational.

Participants were recruited through university mailing lists and Twitter; they were offered a €15 (or equivalent) Amazon voucher for taking part in the study.

[3]Gender was collected inline with the suggestions of [66]

*4.2. Experiment Design*

The experiment had a within-participants design whereby each participant mood logged for a different fictional scenario with each of the three chatbots. The order of chatbots and scenarios was randomized such that the joint ordering of the chatbots and scenarios (36 possible orders) was counterbalanced. The experiment was designed to analyse how the fixed effect of the chatbot politeness affects users' perceptions of the chatbots.

*4.2.1. Scenarios*

We chose for participants to mood log for pre-defined scenarios of slight positive valence to ensure consistency across participants, minimise repetition, and to minimise risk to participants. The scenarios were intended to be realistic and are presented by Table 3. We consider scenario a random effect.

Table 3: The three scenario descriptions provided to participants to use when mood logging.

| Scenario | Description |
|---|---|
| A | You have just finished work for the week and are about to meet some friends for dinner. |
| B | You received some good feedback on a piece of work that you've been working on for weeks. |
| C | You've had a relaxed day watching television after a busy week. |

*4.2.2. Measures*

Immediately following each chatbot interaction, participants were asked to rate on a 7-point scale (1=Strong Disagree and 7=Strongly Agree) their agreement with the statement that '*the agent was polite*'. This item was to serve as a manipulation test to evaluate how polite each chatbot was perceived, its wording closely resembles that used in previous work [7].

In addition, participants were asked to rate the chatbot across a set of nine 7-point items we term the General Agent Rating (GAR) items (Table 4). We constructed the item set by consolidating items used in previous research to evaluate conversational agents for healthcare applications [4, 9, 7, 10, 64, 80, 81, 55, 57, 56]. Our intent with the GAR items was to broadly explore dimensions of user perceptions of chatbots in order to identify potential effects of the use of politeness. Items were chosen in order to identify effects that could impact the suitability of chatbots for different, particularly sensitive, applications. The items Easy; Express; and Natural concern the experience of communicating with the chatbot while the items Cares; Relationship; and Trust consider how the chatbots are perceived as dialogue partners. The items Like; Satisfied; and Continue concern participants' satisfaction with the interaction; Continue has been used in prior work [7] as a proxy for long-term adherence. We emphasise that, although the GAR items were given together at the end of each chatbot interaction and are displayed together in Table 4, they are not a questionnaire inventory and thus are analysed separately.

*4.2.3. Ethical Considerations*

As the study investigates mood logging strategies, that may not be effective, and concerns the topic of mental health, we must consider the potential for it to be upsetting and distressing to participants. To minimise the risk of participants being negatively impacted by the study, we

Table 4: The General Agent Rating (GAR) items are a set of nine 7-point Likert items for assessing people's perceptions of conversational agents.

| Name | Item | Anchor 1 | Anchor 7 |
|---|---|---|---|
| Easy | How **easy** was talking to the agent? | Very easy | Very difficult |
| Express | How much do you feel you could **express** yourself? | Not at all | Very much |
| Natural | How **natural** was your conversation with the agent? | Not at all | Very natural |
| Cares | How much do you feel the agent **cares** about you? | Not at all | Very much |
| Relationship | How would you characterise your **relationship** with the agent? | Complete stranger | Close friend |
| Trust | How much do you **trust** the agent? | Not at all | Very much |
| Continue | How much would you like to **continue working** with the agent? | Not at all | Very much |
| Like | How much did you **like** the agent? | Not at all | Very much |
| Satisfied | How **satisfied** are you with the agent? | Not at all | Very satisfied |

did not specifically recruit participants experiencing mental health difficulties. We also used predefined scenarios of a positive valence such that participants would not be mood logging about their own experiences or scenarios with a high propensity for negative effect.

### 4.3. Interview Design

After interacting with all three chatbots, participants took part in a semi-structured interview. A semi-structured interview approach was chosen as it allowed interviewers a level of flexibility in the topics discussed. While allowing for this flexibility, the interview covered five key topics: 1) participant's general thoughts on chatbots or conversational interfaces; 2) attitudes around the use of chatbots to track mood; 3) thoughts on the chatbots used specifically in this study; 4) what participants liked or disliked about the language used by the study's chatbots; and 5) potential future use cases for chatbots for mental health support. Discussion of the study's chatbots included screen sharing of the prompts used by the chatbots in order to support participant recall, allowing them to comment on specific prompts and aiding in discussion clarity.

### 4.4. Procedure

Prior to the study session, participants were provided with an information sheet and were asked to complete an informed consent form. Each session was conducted by either the first or second author, using video call, and lasted around 30 to 60 minutes.

When a study session commenced, the participant was first briefed about the study procedure. The participant was then directed to a Qualtrics survey which first collected demographic information before running the controlled experiment. The participant and researcher remained on the video call during this, however the participant completed the survey independently.

Within the Qualtrics survey, participants were presented with each of the three chatbot-scenario pairs. Each pair was presented to the participant independently. The survey would show the participant the scenario in text-form and provide a link alongside it which would open the chatbot in a new tab. The participant was asked to complete a mood log entry with the chatbot as though they were in the given scenario. Immediately following each interaction, before interacting with the next chatbot-scenario pair, they were asked the manipulation and GAR items.

Upon completing the survey a semi-structured interview was conducted. The session concluded with participants being debriefed and thanked for taking part in the study.

### 4.5. Analysis Approach

#### 4.5.1. Quantitative

Participants' ratings of the chatbot on the manipulation check and each GAR item were analysed using linear mixed-effects models (LMEMs). LMEMs are a regression analysis technique that allows for the modelling of both fixed (e.g., politeness) and random effects (e.g., by-participant or by-scenario based effects) when analysing the data. The models were run in `R 4.1` [61] using the `lme4 1.1-27.1` [3] and `lmerTest 3.1-3` [42] packages. Following best practice guidelines [2] and similar analysis in related work [70, 78, 29], LMEMs for each questionnaire item were initially constructed using a maximal model approach (i.e., participant and scenario level random intercepts and slopes included). These models were then simplified as necessary until convergence was achieved. The full model syntax for the converged models along with full reporting of fixed and random effects is reported in the supplementary material.

To identify effects of the polite chatbots compared to the direct chatbot, the direct chatbot was set as the baseline condition within the chatbot fixed effect. Analysis was also conducted whereby the Passive chatbot was set as the baseline condition, allowing us to directly compare the ratings of the Passive and Personal chatbots. Consistent with advice on exploratory statistical analysis in HCI [16] and similar to recent prior work [70] we do not explicitly correct the p value for the number of statistical tests conducted. The use of a correction with the number of tests would lead to Type II errors. However, due to the number of analyses conducted, findings that are statistically significant at a p value of or around .05 should be interpreted with caution.

#### 4.5.2. Qualitative

Our qualitative analysis approach for data collected from the 39 participant interviews draws on reflexive thematic analysis [13, 14] while aligning with the structure of our research study. The first two authors of the paper, who conducted the interviews, conducted the analysis collaboratively. We wanted to conduct the analysis collaboratively to include multiple perspectives in the analysis to promote exploration and reflexivity; our collaboration approach was guided by an approach described by Braun et al. [14]. All interviews were audio recorded and auto-transcribed. Resulting transcripts were then manually reviewed and corrected by the researcher who conducted the respective interview. We familiarised ourselves with the data set through reading the transcripts and note taking. Next, working with separate copies of the entire data set, we coded the data inductively, and mostly semantically, which aligns with our research aims of exploring and describing participant's perceptions. One researcher used spreadsheets and the other NVivo. We initially worked independently when coding the data and then started to increasingly confer to discuss, compare, and reflect on our codes as the analysis progressed. From our codings, we started to each construct candidate themes which we then developed into a shared set of themes through comparison and discussion. Throughout the analysis process we perceived our individual analyses to be very similar, we did not identify any major differences between our interpretations of the data. We attribute this, in part, to the structured nature of the data collection and the descriptive orientation of the analysis. Our semi-structured interview study had two main purposes. Firstly, we wanted to better understand how participants specifically perceived the use of politeness by each of the three chatbots; which determined our first three themes. Secondly, our theme construction involved creating themes more typical of reflexive thematic analysis to describe participant's perceptions of chatbots for mental health support more generally.

## 5. Quantitative Analysis

### *5.1. Manipulation Check - Chatbots Using Politeness are Perceived as More Polite*

Descriptively, participants rated both the Passive (M=5.64; SD=1.44) and Personal (M=6.05; SD=1.07) chatbots are more polite than the Direct (M=5.13; SD=1.58). As shown by Table 5, both the Passive and Personal chatbots were rated at more polite than the direct with statistical significance ($\alpha$=0.05). Although descriptively the Personal chatbot was rated as more polite than the Passive, an analysis did not find this difference to be of statistical significance [Unstandardized $\beta$= 0.41, t=1.65, p=0.1043; CI (95%) 0.02-1.00].

Table 5: Politeness LMEM tests compared against Direct. $\beta$ values are unstandardised. ($*$, <0.05; $**$, <0.01; $***$, <0.001)

| Chatbot | $\beta$ | CI (95%) | t | p |
|---|---|---|---|---|
| Passive | 0.51 | 0.02 - 1.00 | 2.06 | 0.0430* |
| Personal | 0.92 | 0.44 - 1.41 | 3.71 | 0.0004*** |

### *5.2. General Agent Rating Items - Personal Politeness Is Favourably Perceived*

Table 6 presents descriptive statistics for the General Agent Rating items. Considering the mean ratings, the Direct chatbot was rated lowest and the Personal chatbot highest for every item except Easy. For the Easy item, the Direct chatbot was rated highest and the polite chatbots equally.

To compare the ratings of each chatbot we conducted analysis using LMEMs (Section 4.5.1) whereby the rating of the Passive and Personal chatbots were compared to the ratings of the direct chatbot. The reported $\beta$ values are unstandardised.

Table 6: Descriptive Statistics for Politeness rating and GAR items (Mean (SD)). Item names are shorthand, based on the bold text in Table 2

| GAR Item | Direct | Passive | Personal |
|---|---|---|---|
| Easy | 3.15 (2.08) | 2.92 (1.68) | 2.92 (1.87) |
| Express | 3.90 (1.74) | 4.36 (1.68) | 5.03 (1.35) |
| Natural | 3.26 (1.67) | 3.74 (1.68) | 4.36 (1.83) |
| Care | 3.15 (1.69) | 3.62 (1.87) | 4.33 (1.77) |
| Relationship | 2.41 (1.46) | 2.97 (1.53) | 3.38 (1.82) |
| Trust | 3.33 (1.54) | 3.64 (1.72) | 4.21 (1.38) |
| Continue | 3.79 (1.84) | 3.97 (1.78) | 4.56 (1.67) |
| Like | 3.46 (1.76) | 3.97 (1.53) | 4.77 (1.61) |
| Satisfied | 3.92 (1.92) | 4.26 (1.76) | 5.08 (1.46) |

As shown by Table 7, for all GAR items — except *Easy* — the analysis shows the Personal chatbot was rated statistically significantly higher on each item than the Direct chatbot. There were however no statistically significant differences between the Passive and Direct chatbot ratings (all p >.05). These statistics suggest that for the GAR items, the chatbot using Personal politeness was perceived differently to the Direct chatbot, while the chatbot using Passive politeness was not.

Table 7: LMEM tests Personal compared to Direct (∗, <0.05; ∗∗, <0.01; ∗ ∗ ∗, <0.001)

| GAR Item | $\beta$ | CI (95%) | t | p |
|---|---|---|---|---|
| Easy | -0.23 | -0.94 - 0.48 | -0.63 | 0.529 |
| Express | 1.13 | 0.57 - 1.68 | 3.97 | 0.0002*** |
| Natural | 1.11 | 0.54 - 1.68 | 3.80 | 0.0003*** |
| Care | 1.18 | 0.52 - 1.84 | 3.49 | 0.0008*** |
| Relationship | 0.97 | 0.39 - 1.56 | 3.27 | 0.0016** |
| Trust | 0.88 | 0.33 - 1.43 | 3.17 | 0.0023** |
| Continue | 0.77 | 0.17 - 1.37 | 2.51 | 0.014* |
| Like | 1.31 | 0.73 - 1.89 | 4.40 | 0.00003*** |
| Satisfied | 1.15 | 0.56 - 1.74 | 3.83 | 0.0003*** |

In view of these findings we chose to conduct further LMEM analysis (Section 4.5.1) whereby the Passive chatbot was set as the baseline condition to enable direct comparison of the Passive and Personal chatbot ratings. As shown by Table 8 the differences were significant for 6 of the 9 GAR items. This indicates that, for the GAR items, the difference in perception between the Personal and Passive chatbots is greater than that between the Passive and Direct.

Table 8: LMEM tests Personal compared to Passive (∗, <0.05; ∗∗, <0.01; ∗ ∗ ∗, <0.001)

| GAR Item | $\beta$ | CI (95%) | t | p |
|---|---|---|---|---|
| Easy | -0.00 | -0.77 - 0.71 | -0.01 | 0.994 |
| Express | 0.67 | 0.11 - 1.22 | 2.35 | 0.0215* |
| Natural | 0.61 | 0.04 - 1.18 | 2.08 | 0.0406* |
| Care | 0.72 | 0.05 - 1.38 | 2.12 | 0.0372* |
| Relationship | 0.41 | -0.17 - 0.99 | 1.38 | 0.1722 |
| Trust | 0.56 | 0.01 - 1.10 | 1.99 | 0.0498* |
| Continue | 0.59 | -0.01 - 1.19 | 1.93 | 0.0576 |
| Like | 0.79 | 0.21 - 1.38 | 2.68 | 0.0091** |
| Satisfied | 0.82 | 0.23 - 1.41 | 2.73 | 0.0080** |

The results for the Easy item show that participants did not perceive any chatbot as easier to talk to. However, the interactions with the Personal chatbot were rated with statistical significance as more natural than with both the Direct [$\beta$=1.11, p=0.0003***] and Passive [$\beta$=0.61, p=0.0406*] chatbots. Similarly, participants ratings demonstrate that participants felt more able to express themselves to the Personal chatbot than to both the Direct [$\beta$=1.13, p=0.0002*** and Passive [$\beta$=0.67, p=0.0215*] chatbots. These findings suggest that Personal politeness can support communication with chatbots relative to Passive politeness and, more so, directness.

Participants characterised, with statistical significance, their relationship with Personal chatbot as more similar to a close friend than their relationship with the Direct chatbot [$\beta$=0.97, p=0.0016**]. While descriptively participants characterised their relationship the Passive chatbot as between that of Direct and Personal, neither of the differences in ratings between the Passive and other two chatbots were of statistical significance. In contrast, participants rated with statistical significance the Personal chatbot as not only more caring than the Direct chatbot [$\beta$=1.18, p=0.0008***], but also as more caring than the Passive chatbot [$\beta$=0.72, p=0.0372*]. The Personal chatbot was trusted more by participants than the Direct [$\beta$=0.88, p=0.0023**]; the Personal chatbot was also trusted more than the Passive chatbot [$\beta$=0.56, p=0.0498*], however we caution that this p-value is close to the significance level. These findings indicate that the use of Personal politeness can support a more intimate relationship between the user and the chatbot.

The results for the Like and Satisfied items are of the largest effect sizes and most significant p-values. Participants liked [$\beta$=1.31, p=0.00003***] and were satisfied [$\beta$=1.15, p=0.0003***] considerably more with the Personal chatbot than the Direct. Although with a smaller effect size, the ratings show participants also liked [$\beta$=0.79, p=0.0091**] and were satisfied [$\beta$=0.82, p=0.0080**] more with the Personal than the Passive chatbot. These findings strongly indicate that Personal politeness can improve the user experience of chatbot interaction. What's more, participants rated that they would like to continue working with the Personal chatbot to an extent significantly greater than they would like to continue working with the Direct chatbot [$\beta$=0.77, p=0.014*]. Participants did not rate that they would like to continue working with the Personal chatbot to an extent significantly greater than the Passive chatbot [$\beta$=0.59, p=0.0576].

## 6. Qualitative Analysis

In this section, we describe the two main thematic findings that we constructed through our analysis pertaining to: (1) participants perceptions of politeness in each of the three chatbot variants; and (2) participants more general though on prospects of using a chatbot for the mental health task of mood logging.

### *6.1. User Perceptions of Chatbots using Politeness*

Our analysis is organised according to participant's perceptions of each chatbot's politeness approach: (1) Direct - the expected robot; (2) Passive - condescending and too apologetic; and (3) Personal - the conflict of being caring.

#### *6.1.1. Direct: The Expected Robot*

A few participants expressed a preference for the Direct chatbot, perceiving it as affording quick and efficient logging.

> *"I think the first one [Direct] was just very straightforward maybe, and it felt more like what you would expect from a chatbot."* [P17]

Similarly, some participants, while generally preferring one of the polite chatbots, expressed that they would prefer the Direct chatbot in situations where they want to complete a mood log quickly. For example, when they are short on time or when they are in a positive mood on which they do not feel the need to reflect.

The Direct chatbot was often described as robotic and contrasted with the two polite chatbots which were assigned more 'human' attributes by participants; this suggests that the use of politeness by chatbots has an anthropomorphising effect.

*"[Direct] is very much just robot text, text, text and I think it'll put people off."* [P09]

This robotic nature was deemed as potentially damaging to the success of a chatbot for mood logging. Participants described experiencing it as disconnected and emotionless, which generally dissuaded them from wanting to use it.

*6.1.2. Passive Politeness: Condescending and Too Apologetic*

A common opinion among participants was that the Passive chatbot was overly polite and more formal than its Personal counterpart. Although a few participants appreciated the chatbot acknowledging that it may be inconveniencing the user, many participants critiqued this apologetic stance. It was particularly felt that the "sorry to bother you" initiation was inappropriate. It was described that this initiation starts the interaction in a defensive way, and on a negative note, which implies that mood logging is something the user does not want to do. A number of participants pointed out that users of the chatbot would have likely already downloaded the application of their own free will, so the chatbot apologising to them for carrying out its purpose is unnecessary. In some cases this resulted in feelings of irritation.

*"I have the choice to ignore it, if I wanted to if it - if it bothers me. So to me when a chatbot comes to me being like 'sorry to bother you', it just doesn't feel natural"* [P39]

The perceived formality of the chatbot led participants to liken the interaction experience to that of being in a doctor's office or the workplace. These exemplify situations or places whereby politeness is often used as a means of reducing imposition on the hearer, as opposed to being indicative of a close relationship between the speaker and hearer.

*"Almost like being in a doctor's office you know, like 'Sorry to bother you', you know, 'one last thing' kind of thing, like it was a bit more formal"* [P38]

As well as being perceived as more formal, many participants found the approach of the Passive chatbot to be condescending, passive aggressive, or sarcastic in tone, the same way a child may feel if scolded by their parents. It was suggested this could exacerbate any existing low moods of users engaging with the chatbot. Indeed, concern was even expressed that this tone could be potentially harmful.

*"I think the dangerous thing about that, in a mental health setting, is that you don't know what this person has actually written. So if they've pulled up something that's very vulnerable or that was really difficult for them to write, like 'Oh, I had a terrible day and I had a fight with my mom' or something really that was quite intense and then you get a response like that ['that wasn't too hard'], it might come off the wrong way, of being condescending even though that wasn't the intention at all."* [P03]

*6.1.3. Personal Politeness: The Conflict of Being 'Caring'*

Participants commonly described the Personal chatbot as warm, comforting, and more emotional, in some cases likening it to a person or friend that wished to hear about the participant's day. This perception led to an impression that the chatbot cared about the participants and was itself genuinely eager to hear their responses to each prompt.

> *"Like 'you'll tell me about what you've done today, I hope', I like that 'I hope' part at the end, it's like, kind of like 'Oh, if you wish to' like you feel like this person like cares about how you want to release your information or your emotions it's not just 'you'll tell me'."* [P08]

Although some participants referred to the chatbot as if it were a person, others expressed this sentiment while concurrently acknowledging that they knew the chatbot was a machine.

> *"[Personal] just like sort of more like 'great to see you' so it was, it kind of felt like, I know it's just a bot, but it kind of felt like it was happy to see me"* [P24]

This 'caring' tone, in turn, led many participants to feel more open to disclosing their feelings to the chatbot.

> *"When I interacted with [Personal] it felt really sweet, and I was like aw okay I'll tell you about my day."* [P23]

It was commented that the caring and compassionate nature of this chatbot could make it well suited for when the user is in an low mood and may be more open to spending more time reflecting on this mood and why they are experiencing it. For some, this 'caring' tone was recognised, but had an adverse effect. These participants felt the chatbot cared *too* much about their answers and were subsequently reluctant to share information.

> *"Even the 'it would be helpful for me to know', like why? What are you doing with my data? Why would it be helpful for you? I don't want it to be helpful for you, I want it to be helpful for me"* [P29]

Furthermore, a few participants described this behaviour as 'pushy'. There appears to a be a juxtaposition in that, although the Personal chatbot led people to feel that the chatbot was eager for them to engage, some of the phrases used left people wondering as to the intentions of the chatbot and its role during interactions with the user.

### *6.2. Using Chatbots for Mood Logging*

This analysis explores participant's perceptions of chatbot use for mood logging more generally. We consider: (1) a chatbot's capability to understand the user; and (2) how users conceptualize who (or what?) they are disclosing to.

#### *6.2.1. Understanding the User*

Participants perceived potential for chatbots to support users with self-understanding, yet this potential was considered dependent upon the capabilities of chatbots to understand the user. Participants felt that chatbots could encourage and support user reflection. It was perceived that chatbots could prompt users to reflect on topics that are relevant to the user's experience, but that the user would not independently think to reflect upon.

> *"They guide you kind of down the line of thought [...] it's kind of like prompting you to reflect on things that might be affecting your mood. Whereas when you're by yourself, I guess you can just not really pinpoint why you're feeling that way."* [P01]

Yet, for this to be effective, and for the user experience to be reasonable, participants felt that chatbots would need to engage users in way that makes them feel sufficiently heard and understood. It was suggested that chatbots should respond in supportive and emotionally appropriate ways.

> *"You know, if I had a crappy day or something I would kind of want a little bit of, kind of like sympathetic feedback in a way. Like even just a comment from the Bot like 'Oh', you know 'sorry that your mood is only...' you know, 'two out of five' or something like that. Because, you know, I can kind of just write it down myself if I didn't get those kind of aspects. I want to kind of have this you know, a little bit of feedback even though it's, you know, a mood bot.* [P36]

However there was also scepticism about the appropriateness of, and capabilities for, chatbots behaving in an understanding way. This included critiques about it being inappropriate for chatbots to comment of the difficulty of the activity for the user (i.e., how hard or not hard the mood logging was) that extended to concerns about chatbot interactions causing the user distress. Judgements made by a chatbot about the difficulty of a disclosure, or similarly assumptions about how the user feels, may be misaligned with the user's actual experience. The misalignment risks an invalidating experience, or one of feeling misunderstood, for the user whereby the user may feel the chatbot is implying they should feel differently. Accordingly, there are perhaps limits to the extent to which a chatbot should act as though it understands the user's experience.

Participants perceived opportunity for chatbots to demonstrate understanding of users, and to support users with self-understanding, through use of the user's data. Chatbots could, for instance, incorporate references to the user's data into conversation to promote a user experience of feeling heard and understood.

> *"what would encourage me to keep using it would be if it would like continue on learning from what I say to it, like, like it will remember things about me. Like let's say I feel really happy a week ago and the chatbot remember and is like 'Oh, remember last week, you were feeling like that'."* [P05]

Many participants spoke of chatbots integrating with other data sources about the user, such as their calendar or trackers relating to health and fitness, and using this information to support users in more relevant ways. For example, a participant described how a chatbot could make use of the user's calendar to support them with anxiety:

> *"It would ask you, you know, 'you've got a meeting on at two - how are you feeling about that?' ... then follow up later on, like 'you said you were nervous about the meeting... how did you end up feeling?' and then you could end up getting this feedback at the end of the week, saying like 'you end up getting really anxious about meetings but they go better than you think"'* [P18]

Participants felt that chatbots could help people interpret and gain insights from their data, but were often vague as to how. Some participants described functionality where the chatbot would analyse the user's natural language disclosures in relation to their mental health to identify influential patterns of behaviour or factors. Functionality was also described where the chatbot would regularly (e.g., weekly) create textual summaries of the user's logging for the user to reflect upon. The use of visualisations was also often suggested, possibly because this is what people have come to expect of data analysis. One participant suggested the use of word clouds to illustrate words associated with different moods.

*6.2.2. Disclosure to Who?*

Participants' thoughts about the disclosure opportunity presented by a chatbot for mood logging varied around who they conceptualised the disclosure as being to. A small number of participants described the disclosure as being to, or with, themselves. In the case of these conceptualisations the chatbot is acting as medium for expression, similar to a journal.

> *"It's almost like yeah like, like my own, like me talking to myself, you know? The chatbot becomes a reflection of myself, then I can just log my mood like that. I guess because I'm aware of the fact that it's an inanimate object, or it's a program, I'm more comfortable to kind of talk and, again it's like self-reflection, you know. Rather than confiding within somebody it's almost confiding within myself."* [P01]

Most participants, however, described the disclosure as being to, or with, a chatbot whereby the chatbot is conceptualised more as a conversation partner. Participants perceived benefits of disclosure to chatbots relative to disclosure to another person. With a chatbot, people do not need to worry about their conversation partner judging them, being disinterested, or reacting in a obviously biased or emotional way. For some, this can make the act of disclosure easier and subsequently allow for more honest disclosure. Consequently, for many, self-disclosure of this kind of information to a chatbot can be preferable over disclosure to a person.

> *"It's not going to judge you [...] it doesn't care about like how - like if you're sad or you're happy, it'll just be there."* [P05]

> *"Personally I feel it'd be a bit easier disclosing it to the chatbot, because sometimes when you're talking to a person, you know the body language, the eye contact, you can kind of see if they're not into it, or if they're not, if they don't really care."* [P38]

Furthermore, it was expressed that the conceptualisation of disclosure to a chatbot, rather than to one's self, may even promote more honest disclosure. A perceived reason for this is that when people describe their moods to themselves they may describe more preferable moods, potentially due to self-stigmatisation or toward promoting the experience of more preferable moods.

Some participants felt a chatbot for self-disclosure would be particularly beneficial for people who do not have another person to disclose to. Similarly, it was felt chatbots could benefit people who struggle to disclose to other people, potentially helping them to become more comfortable with self-disclosure and thereby supporting them to disclose to other people in the future. Other participants expressed contrary views that enabling people to disclose to chatbots, rather than people, may not be for the best. There was concern that disclosing to a chatbot could itself be a lonely or sad experience. There was also concern that people may come to prefer interacting with a non-judgemental chatbot over a potentially-judgemental person, and that this may lead them to reduce their interactions with other people.

Often in conjunction with describing the benefits of disclosure to a chatbot, relative to another person, participants spoke of how other people may view or use their disclosure data. In the knowledge of this, some participants commented that they would moderate their disclosures to a chatbot. A few participants described ways other people could make use of the user's mood logging data to support the user. Participants envisioned their data being shared with their therapist to help understand the onset of a mental health difficulty or to monitor treatment progress. One participant also perceived an opportunity for summaries of their data to be shared with their partner. Additionally, there were expectations described by a few participants of monitoring to identify, and then support, people at higher levels of risk (e.g., of suicide). Many participants were however concerned about who was collecting their data and what their intentions were. This concern caused many participants to question if chatbots should be trusted with their disclosures. Some participants were concerned about how people with access to their data may judge them.

> *"A negative of it would be, if I confide in it personally, say like something extremely personal, and then I feel weird because maybe somebody looking at it, on the other end of the screen, is laughing with their friends."* [P01]

Most concerns were about how the participant's data may be used to manipulate them. A frequent concern was the use of user's data to inform advertising.

> *"I wouldn't want, you know, a conversational agent or a chatbot to know if I'm not feeling, you know, well, or if I'm not in a good mood and then just, it just starts, you know, throwing lots of advertisements towards me."* [P20]

Participants expressed that without a reputable backing - such as that of scientists, a government, or health service - they would doubt the effectiveness of the application. Distrust was particularly expressed about the prospect of a chatbot provided by a big tech or unknown organisation.

## 7. Discussion

In this section we will make clear that although politeness can positively impact the user experience, both the type of politeness and interaction context matter. We argue that to harness the potential of polite chatbots, we need to be careful in our application of politeness theory, adapting it to the characteristics of chatbot interaction. We set out a research agenda for embedding politeness into the design of conversational interfaces.

### *7.1. Politeness Can Positively and Negatively Impact the User Experience*

Our combined quantitative and qualitative analysis demonstrates clear differences between how the chatbots using different politeness approaches were experienced by our participants. In contrast with the direct chatbot which was described as robotic, the polite chatbots were assigned more human-like attributes which indicates that the use of politeness has an anthropomorphic effect. This result aligns with the conclusions of Morrissey and Kirakowski [53] and Chaves and Gerosa [18] that politeness (or manners) increases the human-likeness of chatbots. Our study demonstrates that *how* politeness is used (i.e., the type of politeness) can substantially impact how a user experiences interacting with the chatbot. For the most part, interactions with the Personal chatbot were experienced as, and praised for, being friendly and caring. In contrast, for the most part interactions with the Passive chatbot were experienced as formal and criticised for being too apologetic and condescending. This is an important result as while both chatbots were designed to be polite - as prior work has recommended - very different user experiences were produced. With the Personal chatbot being designed to use positive politeness and the Passive chatbot to use negative politeness, the generally favourable experience of the Personal chatbot echos the work of Miyamoto et al. [52] that showed positive politeness to be more favourably perceived in human-robot interaction. Consequently our analysis supports the reoccurring conclusion that politeness in human-chatbot interaction is impactful for the user experience [38, 18, 79, 62]. Our work emphasises that to harness politeness effectively within chatbot interaction design, we need to understand the impacts of different politeness approaches on the user experience.

The positive experiences of the Personal chatbot suggest that Personal politeness could be well suited for sensitive activities like those for mental health support. Personal politeness will likely suit interactions that, more generally, we would expect to benefit from more friendly and caring design. It could, for instance, be useful to develop sensitive and caring chatbot dialogues to support the self-reporting of sexual abuse experiences [50] or the reporting of online threats [60]. Furthermore, many participant's characterisation of their experience interacting with the Personal chatbot resemble the "polite and friendly" conversational environment advocated for by Rapp et al. [62] for promoting collaborative human-chatbot interactions.

Passive politeness, however, does not seem well suited for mood logging. Participants disliked the apologetic behaviour of the Passive chatbot. The apologetic behaviour seems to run contrary to participant expectations of a mood logging chatbot, in that users will likely have expressly opted into the activity of mood logging before interacting with the chatbot, making the apology unnecessary. This misalignment of expectations suggests that Passive politeness may be poorly suited to situations where the user is already aware that the system will request them to interact. The Passive chatbot was also experienced as being condescending which, as commented by study participants, could cause the user distress, particularly if experienced during a sensitive interaction. This highlights how chatbot interactions for mental health support risk having a detrimental impact on the user's wellbeing if they are not carefully designed, including in how they use politeness. It is possible that for different applications, Passive politeness may better support the user experience than Personal Politeness. The apologetic behaviour of Passive politeness may be considered more appropriate for interactions that people consider as more inconvenient; for example, it may suit a customer service interaction to assist the user with a faulty product or missing delivery. Interaction with Passive chatbot was also perceived as formal, and likened to being in a doctors office, which may make it more appropriate than the friendly behaviour of Personal politeness for more formal interactions. For instance, in the context of mental health support, Passive politeness may suit more administrative interactions such as those for the organisation of sessions with a therapist.

Although many participants disliked the robotic nature of the Direct chatbot, some participants favoured it as they perceived it to afford a fast and efficient interaction. This preference for the Direct chatbots is similar to that expressed by the participants of Clark et al.'s study for interaction with voice user interfaces to be task-oriented [22]. Our analysis also showed that the caring nature of the Personal chatbot, despite being generally well perceived, incited distrust in some participants. Similarly when studying perceptions of caring behaviours by conversational interfaces, Bickmore and Picard [8] discovered a mixture of perceptions within their user groups, with many either liking or disliking the caring behaviour. These between-participant differences in how the chatbots were perceived suggests individual differences in how the use of politeness by chatbots is experienced by users. These differences may relate to the participant's orientation to social behaviour by chatbots. Liao et al. [47] showed that people have different orientations to social behaviour in human-chatbot interaction and Lee et al. [45] found participant's orientation to social behaviour human-robot interaction affected how sensitive they were to robot's use of politeness. Our analysis, in conjunction with this prior work, suggests that for politeness to be used effectively within human-chatbot interaction it will need to be adaptable to user's orientation to social behaviour by chatbots.

Our analysis also suggests that, rather than using one strategy consistently, politeness prompts should be adaptable to contextual factors. For instance, our participants suggested that the Direct chatbot could be preferable when they are in a good mood, so as to facilitate a brief interaction. Yet if they are in a low mood, Personal politeness may be preferable as it affords a more compassionate interaction. This ties in with findings of Lee et al. [44] (studying a virtual driving assistant) and Salem et al. [65] (studying a robot receptionist) which relate the appropriateness of politeness to characteristics of the interaction task. In particular, the preference we identified of the Direct chatbot for brief interactions resembles the preference identified by Lee et al. of direct instructions during failure situations. Our findings broaden this relationship to consider, in addition to task characteristics, the momentary mental state and situation of the user. An implication of this is that, to make effective use of politeness, chatbots may need to consider aspects of both the interaction task and the user's state, so as to dynamically adapt their use of politeness to these. These challenges are also faced by proactive conversational interfaces [17, 12, 11], where modelling of the user's cognitive state, context, and task is seen as critical to ensuring that conversational interfaces engage appropriately. Effective adaption may require carefully incorporating inferences about the user's momentary state into dialogue designs that support the quick identification of the what interaction characteristics suit the user in that moment [12].

In sum, the use of politeness can both positively and negatively impact the chatbot user experience. Our findings highlight the need to carefully design and evaluate the use of politeness by chatbots, particularly for sensitive applications. For politeness to be used effectively, it may need to be adaptive to the user and the interaction context. Although future work needs to apply these strategies in wider contexts, our findings show promise in the potential for politeness to support human-chatbot interactions.

### 7.2. Face Threatening Acts in Chatbot Interaction

As discussed above, the use of politeness by chatbots impacted the user experience in a range of ways. Here we consider how our analysis of these impacts relates to politeness theory [15].

As mentioned above, across both our quantitative and qualitative data analysis, Personal politeness was clearly seen as caring. This perception of the chatbot is consistent with how politeness theory informed its design. Our Personal politeness chatbot makes use of positive politeness, the redressive actions of which attempt to satisfy people's want to be desirable, typically by showing an interest in them (i.e., expressing care) [15]. Indeed, many participants felt this caring behaviour was supportive and felt encouraged to interact with the chatbot. Yet this was not seen positively by some participants, leading to a feeling that the chatbot was untrustworthy and subsequently making them reluctant to interact with it. As previously noted, this analysis resembles the mixed perception of caring behaviour by conversational agents previously identified by Bickmore and Picard [8]. A possible explanation for this is that users may not see the chatbots as having the ability to be face-threatening in the first place. Our analysis (Section 6.2.2), and prior work [75, 48, 35, 39], have demonstrated that many people perceive self-disclosure to computers as preferable to self-disclosure to other people due in part to computers being perceived as non-judgemental. For some people, as they perceive computers to be non-judgemental, they may not perceive them as able to threaten their feelings of self-worth (i.e., positive face). This may complicate the use of politeness within human-chatbot interaction. For instance, if users do not perceive the interaction to be face-threatening, then a chatbot's redressive action when using politeness could suggest a potential face threat that the user needs to be aware of. This could in turn signal a reason for the user to distrust the chatbot, as some of our participants did.

Our qualitative analysis shows that participants experienced the Passive politeness chatbot as overly apologetic and condescending. This chatbot being perceived as apologetic is consistent with it's use of negative politeness which includes the technique of apologising. As with the negative perceptions of the Personal chatbot, a face threat mismatch may also explain the negative perceptions of the Passive chatbot. Our Passive politeness chatbot made large use of negative politeness to mitigate threats to the user's autonomy. However, as the user has agency over the chatbot, they may not perceive it as capable of threatening their autonomy. Accordingly, the chatbot's negative politeness redressive actions (e.g., "Sorry to bother you") would be seen as unnecessary and thus overly apologetic.

Our explanation of the negative impacts of politeness on the user experience being the result of face threat mismatches corresponds with the discussion of Chaves and Gerosa [18]. Chaves and Gerosa describe dealing with face-threatening acts as being a challenge of manners in human-chatbot interaction. They argue that while it is natural for people to identify and mitigate FTAs, this is challenging for chatbots due the complexity of identifying such acts. Our explanation suggests that, in part, this is complex because chatbots are not perceived as having the same face-threatening capabilities as people.

The concept of face threat and how this influences perceptions of politeness must also consider the modality and fidelity of the conversational interface. Although our qualitative analysis (Section 6.1.1) found the polite chatbots to be perceived as more human-like, other conversational user interface modalities and platforms (e.g., voice assistants and robots) may be perceived as more human-like than text-based chatbots, resulting in different face threatening characteristics. For instance, work by Kocielnik et al. [40] compared text to voice interaction for reflection, finding that some participants perceived a pressure to respond to the voice interface's questions immediately, allowing them less time to think of their answers. A complexity to this is that users may vary by the human- or machine-like characteristics they attribute to an interface, potentially impacting the ability of the interface to be face threatening [36]. Further research needs to not only consider the differing characteristics of face in conversational human-computer interaction compared to human conversation, but also how modality and platform fidelity may impact the perception of face threat and the impact this may have on the user experience.

### *7.3. Tensions in Using Chatbots for Mood Logging*

Our analysis demonstrates that participants positively perceived the prospect of using chatbots for activities like mood logging. Participants described how chatbots could integrate into people's daily and digital lives, providing the user with emotional support and data-driven insight, helping the user develop self-understanding and supporting the user with challenges they may be experiencing. However, our analysis also surfaced concerns about this prospect which exist in tension with the perceived opportunities.

There is tension around the ways in which it may be appropriate, or desirable, for a chatbot to support a person with their mental health. For the most part, participants were appreciative of the caring and friendly behaviour of the Personal chatbot. Similar to prior work [48, 39], many participants also perceived that chatbots, being machines, could be relied upon to respond to users in a non-judgemental, engaged, and understanding manner. Due to these machine characteristics, concern was expressed about people finding it preferable to self-disclose to, and seek support from, chatbots rather than people. This preference for chatbots was one that some of our participants expressed and that has been identified in prior work [48, 39]. Furthermore, people increasingly receiving support from chatbots may produce unrealistic expectations of how people can and should support each other. It is therefore a concern that the provision of support-giving chatbots may negatively impact how people give and receive support from others. Although many participants preferred the Personal chatbot to the Direct chatbot, the robotic behaviour of the Direct chatbot may promote a more appropriate human-chatbot relationship for mental health support where the chatbot presents less as a friend and more as a tool. Similarly, as discussed in our Understanding the User theme, it may be more appropriate for a chatbot to demonstrate understanding and provide support by harnessing the user's data more and expressing care less. Considering how chatbots can support people with their mental health in conjunction with how people support people with their mental health may help us to navigate this tension. For example, one of our participants' perceived opportunities is being able to share some of their data with their partner. The work of Murnane et al. [54], investigating how people's social relationships and their use of technology for mental health management relate, offers a starting point for such considerations. Given this tension, future work should consider both the appropriateness and potential impacts of people receiving different types of support from chatbots.

The participants of our study were cognisant of the fact that chatbots are provided by people. There were concerns about the motives of the chatbot providers and about how user data may be used; some concerns were exacerbated by the Personal chatbot expressing a want to know about the user. These concerns undermine the perception of chatbots being trustworthy and reliable machines and thus their potential for providing mental health support. To mitigate these concerns it will be important for systems to be transparent about their use of user data, as advocated for by guidelines on the design and evaluation of mental health technologies [25].

### *7.4. Future Directions: Towards Polite Conversational Interaction*

The use of politeness clearly impacts people's experience of interacting with chatbots, as well as other conversational interaction modalities. However, as we discuss in Section 7.2, it appears that politeness does not manifest in human-chatbot interaction the same way it is theorised for human conversation. Toward making effective use of politeness when designing conversational interfaces, we need to develop new theory, or modify existing theory from human-human interaction, to reflect the nature of politeness in conversational human-computer interaction. Below we outline three key focus areas for future developments in this area. We note that these are by no means exhaustive, but are generated based on our analysis and experiences of studying polite conversational design within the context of mood logging:

#### *7.4.1. Distinguishing Politeness Approaches*

The results of this study and others [52, 45] demonstrate that different politeness approaches can have different effects. Although a politeness approach combining a diversity of politeness strategies may ultimately be most effective, this will require an understanding of the contexts in which different strategies are most appropriate. When developing such an understanding we need to be careful if using multiple strategies in conjunction with each other as this risks competing effects that could conceal each other. Research should therefore seek to distinguish different politeness strategies and identify their individual effects. A sensible starting point would be to apply the politeness theory of Brown and Levinson [15] at the more granular level of positive and negative politeness sub-strategies. For instance, Torrey et al. [69] focused on the negative politeness sub-strategy of hedging. Given the findings of this study, we may likely find some of these strategies are not appropriate for use by conversational interfaces. Critically, we will need to take care when describing how we apply politeness strategies in order for researchers and designers to be able to harness them.

#### *7.4.2. Measuring Perceptions of Politeness*

Due in part to the concurrent emergence of much research in this area, there is inconsistency in the measurement of the perception of politeness and concepts relevant to its effects. To support the comparison of findings, and to develop a unified understanding, we should identify or create appropriate measures and use them consistently. Often work has not attempted to measure politeness. Some work [36, 7] has used items resembling our '*the agent was polite*' manipulation check item. Other work has not detailed the measure used [45], used a multi-item politeness scale [44, 65], and also sought to measure the multiple constructs such as positive and negative politeness [36, 82]. To effectively study politeness we may need to identify the dimensionality of how this is perceived when interacting with conversational interfaces. This work could take the form of recent research that has looked to identify the dimensionality of agent personality [71], partner models [27] and human-likeness [28]. Such research could then form a basis for developing a validated scale to measure the dimensions relevant to perceptions of politeness in human-computer interaction for use in future work.

*7.4.3. Exploring Adaptation and Context Awareness*

Our analysis, as discussed in Section 7.1, and prior work suggest that a number of factors may affect how people perceive the use of politeness by conversational interfaces which include user characteristics (e.g., social-agent orientation [47, 45]), the user's momentary state (e.g., mood and activity [17]), and task factors (e.g., urgency, task-type [44, 65]). As discussed in Section 7.2, perceptions may also be impacted by factors including the interaction modality (i.e., chatbot, voice assistant, or robot) and be mediated by anthropomorphism. Research should seek to experiment with these variables while accounting for, if not controlling, the others. As with the measurement of politeness, we should attempt to be consistent in our measurement of these constructs.

*7.5. Ethical Implications*

An unanticipated outcome was that some participants negatively experienced interacting with the Passive chatbot. If the study been less controlled this could have caused distress to participants (e.g., as a result of a participant feeling the chatbot was being condescending during a sensitive interaction) which highlights the importance of staged evaluation [25, 59], whereby systems are first carefully evaluated in low risk settings.

As Passive politeness was negatively experienced it has the potential to be used maliciously (e.g., chatbots designed to annoy or distress users). To protect users from intentional, or unintentional, harmful design we may want to develop techniques, or tools, for the detection of potentially harmful design characteristics. Tools could identify, or help designers to identify, opportunities to apply or the presence of different linguistic (e.g., politeness) techniques and then support designers to reason about the likely impacts on the user experience for that interaction context. Such tools may serve to help designers to identify parts of dialogue that warrant careful user testing and to demonstrate that a system is suitable for more real world use.

The Personal chatbot was perceived as more human-like and caring; some participants found this encouraged them to interact. While this outcome resembles an aim of this research, it also raises ethical concerns. Personal politeness could contribute to users over-trusting, or becoming over-reliant on, conversational agents. Echoing broader concerns about conversational agents [74], this could become avenue for exploiting users. In the case of mental health support, it could contribute to undesirable usage, such as users becoming dependent upon an agent.

*7.6. Limitations*

Our study aimed to explore how the use of politeness impacts the chatbot user experience for mood logging. Similar to other work on politeness use in interface design (e.g., [52, 69, 82]) we used a within-participant controlled experiment whereby participants enacted pre-defined scenarios with the chatbots. This allowed us to gather data on participant perceptions of multiple politeness approaches, whilst also allowing participants to compare these under similar conditions and constraints. The use of pre-defined scenarios also enabled us to minimise the risk of participants being negatively impacted while engaging with participants on the topic on mental health [34]. This approach is consistent with staged evaluation approaches for the development of mental healthcare technology [25], as well as evaluation approaches for healthcare dialog systems [6]. While mood logging systems should be evaluated longitudinally and clinically before real world deployment, this evaluation serves to first explore design possibilities and address the user experience. Informed by this, future work can look to explore ways of understanding how politeness impacts engagement with mood logging in more natural settings.

The quantitative component of our study used a set of 10 one item (1 Politeness and 9 General Agent Rating) ratings. This choice to use a set of single item measures, which as discussed in Section 4.2.2 have been used in relevant prior work, suited the exploratory nature of this study. The item set allowed us to explore a range of relevant dimensions without overwhelming participants with rating items. As research in this area starts to consider more natural settings, it should also seek to adopt validated measures. However, as highlighted in previous work [21], the use of validated self-report scales within research on conversational interfaces is challenging due to there currently being a dearth of validated self-report scales for conversational interaction. We note that further work must look to design and utilise validated scales within this domain (e.g., for measuring politeness).

Our study's specific focus on text-based chatbots for mood logging may affect the generalisability of our analysis. For instance the sensitivity of mood logging may have a notable effect on how users experience the use of politeness. Much of the previous literature, which we have used to situate this work, has studied politeness use by other conversational interface modalities (e.g., robots and voice user interfaces). As discussed in Section 7.2, interface modality and fidelity may impact the effects of politeness, thus it is unclear how our analysis will generalise across conversational interfaces. This is made more challenging due to the fact that politeness research currently uses a variety of measures, making it difficult to compare findings. It is imperative that future work looks to develop and validate measures that can be used to measure concepts relevant to the user experience of conversational interaction that may be influenced by politeness. This would then make it easier to compare and contrast findings across domains and devices.

## 8. Conclusion

Politeness is used in human conversation to support interactions that risk inconveniencing, or being sensitive for, a conversation partner. Using a mixed methods study involving 39 participants, this research studied how politeness impacts the chatbot user experience to see if it could be used to support engagement with sensitive activities like mood logging. While our quantitative findings suggest Personal politeness to be more suitable than Passive politeness or directness, our qualitative findings contribute a more nuanced perspective. Personal politeness was perceived as caring; although this was often desirable, it incited distrust in some participants. Passive politeness, meanwhile, was considered too apologetic. Our analysis suggests the use of politeness needs to be adaptive to the interaction context. In addition, we find that while participants perceived benefits to using a chatbot for mood logging, they were concerned as to whether it should be trusted. We emphasise with this work that to harness the potential of polite chatbots to support sensitive activities like mood logging, we need to be careful in our application of politeness through adapting it to the characteristics of conversational human-computer interaction.

### Acknowledgements

This research was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real, grant 18/CRT/6224) and Microsoft Research through its PhD Scholarship Programme. This research was additionally conducted in part with the financial support of the Science Foundation Ireland Adapt Research Centre (grant 13/RC/2106_P2).

### References

[1] Andrade, L.H., Alonso, J., Mneimneh, Z., Wells, J., Al-Hamzawi, A., Borges, G., Bromet, E., Bruffaerts, R., De Girolamo, G., De Graaf, R., et al., 2014. Barriers to mental health treatment: Results from the WHO World Mental Health surveys. Psychological Medicine 44, 1303–1317. doi:10.1017/S0033291713001943.

[2] Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language 68, 255–278.

[3] Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67, 1–48. doi:10.18637/jss.v067.i01.

[4] Bickmore, T.W., Caruso, L., Clough-Gorr, K., Heeren, T., 2005. 'it's just like you talk to a friend'relational agents for older adults. Interacting with Computers 17, 711–735.

[5] Bickmore, T.W., Cassell, J., 2001. Relational agents: A model and implementation of building user trust, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 396–403. URL: https://doi.org/10.1145/365024.365304, doi:10.1145/365024.365304.

[6] Bickmore, T.W., Giorgino, T., 2006. Health dialog systems for patients and consumers. Journal of Biomedical Informatics 39, 556–571. doi:https://doi.org/10.1016/j.jbi.2005.12.004. dialog Systems for Health Communications.

[7] Bickmore, T.W., Mauer, D., Crespo, F., Brown, T., 2007. Persuasion, task interruption and health regimen adherence, in: Persuasive Technology, Springer, Berlin, Heidelberg. pp. 1–11.

[8] Bickmore, T.W., Picard, R.W., 2004. Towards caring machines, in: CHI '04 Extended Abstracts on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1489–1492. URL: https://doi.org/10.1145/985921.986097, doi:10.1145/985921.986097.

[9] Bickmore, T.W., Picard, R.W., 2005. Establishing and maintaining long-term human-computer relationships. ACM Trans. Comput.-Hum. Interact. 12, 293–327. URL: https://doi.org/10.1145/1067860.1067867, doi:10.1145/1067860.1067867.

[10] Bickmore, T.W., Ring, L., 2010. Making it personal: end-user authoring of health narratives delivered by virtual agents, in: International Conference on Intelligent Virtual Agents, Springer. pp. 399–405.

[11] Bowman, R., Cooney, O., Thieme, A., Cowan, B.R., Doherty, G., 2022a. Pervasive therapy: Designing conversation-based interfaces for ecological momentary intervention. IEEE Pervasive Computing .

[12] Bowman, R., Cowan, B.R., Thieme, A., Doherty, G., 2022b. Beyond subservience: Using joint commitment to enable proactive cuis for mood logging, in: 4th Conference on Conversational User Interfaces, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3543829.3544512, doi:10.1145/3543829.3544512.

[13] Braun, V., Clarke, V., 2013. Successful Qualitative Research: A Practical Guide for Beginners. Sage.

[14] Braun, V., Clarke, V., Hayfield, N., Terry, G., 2019. Thematic analysis, in: Liamputtong, P. (Ed.), Handbook of Research Methods in Health Social Sciences. Springer.

[15] Brown, P., Levinson, S.C., 1987. Politeness: Some universals in language usage. volume 4. Cambridge university press.

[16] Cairns, P., 2019. Doing Better Statistics in Human-Computer Interaction. Cambridge University Press. chapter 13. pp. 155–166.

[17] Cha, N., Kim, A., Park, C.Y., Kang, S., Park, M., Lee, J.G., Lee, S., Lee, U., 2020. Hello there! is now a good time to talk? opportune moments for proactive interactions with smart speakers. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4. URL: https://doi.org/10.1145/3411810, doi:10.1145/3411810.

[18] 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.

[19] Cila, N., 2022. Designing human-agent collaborations: Commitment, responsiveness, and support, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1–18.

[20] Clark, H.H., 1996. Using language. Cambridge university press.

[21] Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., Aylett, M., Cabral, J., Munteanu, C., Edwards, J., et al., 2019a. The state of speech in hci: Trends, themes and challenges. Interacting with computers 31, 349–371.

[22] Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., Munteanu, C., Wade, V., Cowan, B.R., 2019b. What makes a good conversation? challenges in designing truly conversational agents, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–12. URL: https://doi.org/10.1145/3290605.3300705.

[23] Colby, K.M., Watt, J.B., Gilbert, J.P., 1966. A computer method of psychotherapy: Preliminary communication. The Journal of Nervous and Mental Disease 142, 148–152.

[24] Coyle, D., Doherty, G., Matthews, M., Sharry, J., 2007. Computers in talk-based mental health interventions. Interacting with computers 19, 545–562.

[25] Doherty, G., Coyle, D., Matthews, M., 2010. Design and evaluation guidelines for mental health technologies. Interacting with computers 22, 243–252.

[26] Doherty, K., Balaskas, A., Doherty, G., 2020. The design of ecological momentary assessment technologies. Interacting with Computers 32, 257–278.

[27] Doyle, P.R., Clark, L., Cowan, B.R., 2021. What do we see in them? identifying dimensions of partner models for speech interfaces using a psycholexical approach, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: https://doi-org.ucd.idm.oclc.org/10.1145/3411764.3445206, doi:10.1145/3411764.3445206.

[28] Doyle, P.R., Edwards, J., Dumbleton, O., Clark, L., Cowan, B.R., 2019. Mapping perceptions of humanness in intelligent personal assistant interaction, in: Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3338286.3340116, doi:10.1145/3338286.3340116.

[29] Edwards, J., Liu, H., Zhou, T., Gould, S.J., Clark, L., Doyle, P., Cowan, B.R., 2019. Multitasking with alexa: how using intelligent personal assistants impacts language-based primary task performance, in: Proceedings of the 1st International Conference on Conversational User Interfaces, pp. 1–7.

[30] 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, e7785.

[31] Goffman, E., 1967. Interaction ritual: Essays on face-to-face interaction. .

[32] Gratch, J., Lucas, G.M., King, A.A., Morency, L.P., 2014. It's only a computer: the impact of human-agent interaction in clinical interviews, in: Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems, pp. 85–92.

[33] Hammer, S., Lugrin, B., Bogomolov, S., Janowski, K., André, E., 2016. Investigating politeness strategies and their persuasiveness for a robotic elderly assistant, in: International Conference on Persuasive Technology, Springer. pp. 315–326.

[34] Hirsch, T., 2020. Practicing without a license: Design research as psychotherapy, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–11. URL: https://doi.org/10.1145/3313831.3376750, doi:10.1145/3313831.3376750.

[35] 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, 712–733.

[36] Hu, Y., Qu, Y., Maus, A., Mutlu, B., 2022. Polite or direct? conversation design of a smart display for older adults based on politeness theory, in: CHI Conference on Human Factors in Computing Systems, pp. 1–15.

[37] Inkster, B., Sarda, S., Subramanian, V., 2018. An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR mHealth and uHealth 6, e12106.

[38] Jain, M., Kumar, P., Kota, R., Patel, S.N., 2018. Evaluating and informing the design of chatbots, in: Proceedings of the 2018 designing interactive systems conference, pp. 895–906.

[39] Kim, J., Kim, Y., Kim, B., Yun, S., Kim, M., Lee, J., 2018. Can a machine tend to teenagers' emotional needs? a study with conversational agents, in: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–6. URL: https://doi.org/10.1145/3170427.3188548, doi:10.1145/3170427.3188548.

[40] Kocielnik, R., Avrahami, D., Marlow, J., Lu, D., Hsieh, G., 2018. Designing for workplace reflection: a chat

and voice-based conversational agent, in: Proceedings of the 2018 designing interactive systems conference, pp. 881–894.

[41] Kola, L., 2020. Global mental health and COVID-19. The Lancet Psychiatry 7, 655–657. doi:10.1016/S2215-0366(20)30235-2.

[42] Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software 82, 1–26. doi:10.18637/jss.v082.i13.

[43] Lakoff, R.T., 1989. The limits of politeness: Therapeutic and courtroom discourse .

[44] Lee, J.G., Lee, K.M., Ryu, S.H., 2019. Vehicle politeness in driving situations. Future Internet 11, 48.

[45] Lee, Y., Bae, J.e., Kwak, S.S., Kim, M.S., 2011. The effect of politeness strategy on human-robot collaborative interaction on malfunction of robot vacuum cleaner, in: RSS workshop on HRI.

[46] Leech, G.N., 2014. The pragmatics of politeness. Oxford Studies in Sociolinguistics.

[47] Liao, Q.V., Davis, M., Geyer, W., Muller, M., Shami, N.S., 2016. What can you do? studying social-agent orientation and agent proactive interactions with an agent for employees, in: Proceedings of the 2016 ACM Conference on Designing Interactive Systems, Association for Computing Machinery, New York, NY, USA. p. 264–275. URL: https://doi.org/10.1145/2901790.2901842, doi:10.1145/2901790.2901842.

[48] Lucas, G.M., Gratch, J., King, A., Morency, L.P., 2014. It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior 37, 94–100.

[49] Luger, E., Sellen, A., 2016. "like having a really bad pa": The gulf between user expectation and experience of conversational agents, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 5286–5297. URL: https://doi.org/10.1145/2858036.2858288, doi:10.1145/2858036.2858288.

[50] Maeng, W., Lee, J., 2022. Designing and evaluating a chatbot for survivors of image-based sexual abuse, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: https://doi-org.ucd.idm.oclc.org/10.1145/3491102.3517629, doi:10.1145/3491102.3517629.

[51] Matthews, M., Doherty, G., 2011. In the mood: Engaging teenagers in psychotherapy using mobile phones, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 2947–2956. URL: https://doi.org/10.1145/1978942.1979379, doi:10.1145/1978942.1979379.

[52] Miyamoto, T., Katagami, D., Shigemitsu, Y., Usami, M., Tanaka, T., Kanamori, H., Yoshihara, Y., Fujikake, K., 2021. Influence of social distance expressed by driving support agent's utterance on psychological acceptability. Frontiers in Psychology 12, 254.

[53] Morrissey, K., Kirakowski, J., 2013. 'realness' in chatbots: establishing quantifiable criteria, in: Human-Computer Interaction. Interaction Modalities and Techniques: 15th International Conference, HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part IV 15, Springer. pp. 87–96.

[54] Murnane, E.L., Walker, T.G., Tench, B., Voida, S., Snyder, J., 2018. Personal informatics in interpersonal contexts: towards the design of technology that supports the social ecologies of long-term mental health management. Proceedings of the ACM on Human-Computer Interaction 2, 1–27.

[55] Olafsson, S., Kimani, E., Asadi, R., Bickmore, T.W., 2017. That's a rap, in: International Conference on Intelligent Virtual Agents, Springer. pp. 325–334.

[56] Olafsson, S., O'Leary, T.K., Bickmore, T.W., 2020. Motivating health behavior change with humorous virtual agents, in: Proceedings of the 20th ACM international conference on intelligent virtual agents, pp. 1–8.

[57] Parmar, D., Olafsson, S., Utami, D., Bickmore, T.W., 2018. Looking the part: The effect of attire and setting on perceptions of a virtual health counselor, in: Proceedings of the 18th International Conference on Intelligent Virtual Agents, Association for Computing Machinery, New York, NY, USA. p. 301–306. URL: https://doi.org/10.1145/3267851.3267915, doi:10.1145/3267851.3267915.

[58] Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P.Y., Cooper, J.L., Eaton, J., Herrman, H., Herzallah, M.M., Huang, Y., Jordans, M.J., Kleinman, A., Medina-Mora, M.E., Morgan, E., Niaz, U., Omigbodun, O., Prince, M., Rahman, A., Saraceno, B., Sarkar, B.K., De Silva, M., Singh, I., Stein, D.J., Sunkel, C., UnÜtzer, J., 2018. The Lancet Commission on global mental health and sustainable development. The Lancet 392, 1553–1598. doi:10.1016/S0140-6736(18)31612-X.

[59] Peters, D., Vold, K., Robinson, D., Calvo, R.A., 2020. Responsible ai—two frameworks for ethical design practice. IEEE Transactions on Technology and Society 1, 34–47.

[60] Piccolo, L.S.G., Troullinou, P., Alani, H., 2021. Chatbots to support children in coping with online threats: Socio-technical requirements, in: Designing Interactive Systems Conference 2021, Association for Computing Machinery, New York, NY, USA. p. 1504–1517. URL: https://doi-org.ucd.idm.oclc.org/10.1145/3461778.3462114, doi:10.1145/3461778.3462114.

[61] R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

[62] Rapp, A., Boldi, A., Curti, L., Perrucci, A., Simeoni, R., 2023. Collaborating with a text-based chatbot: An exploration of real-world collaboration strategies enacted during human-chatbot interactions, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–17.

[63] Rapp, A., Curti, L., Boldi, A., 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. International Journal of Human-Computer Studies 151, 102630.

[64] Ring, L., Utami, D., Bickmore, T.W., 2014. The right agent for the job?, in: International Conference on Intelligent Virtual Agents, Springer. pp. 374–384.

[65] Salem, M., Ziadee, M., Sakr, M., 2013. Effects of politeness and interaction context on perception and experience of hri, in: International Conference on Social Robotics, Springer. pp. 531–541.

[66] Spiel, K., Haimson, O.L., Lottridge, D., 2019. How to do better with gender on surveys: a guide for hci researchers. Interactions 26, 62–65.

[67] Srinivasan, V., Takayama, L., 2016. Help me please: Robot politeness strategies for soliciting help from humans, in: Proceedings of the 2016 CHI conference on human factors in computing systems, pp. 4945–4955.

[68] Terada, K., Okazoe, M., Gratch, J., 2021. Effect of politeness strategies in dialogue on negotiation outcomes , 195–202URL: https://doi.org/10.1145/3472306.3478336.

[69] Torrey, C., Fussell, S.R., Kiesler, S., 2013. How a robot should give advice, in: 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE. pp. 275–282.

[70] Völkel, S.T., Buschek, D., Eiband, M., Cowan, B.R., Hussmann, H., 2021. Eliciting and analysing users' envisioned dialogues with perfect voice assistants, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3411764.3445536, doi:10.1145/3411764.3445536.

[71] Völkel, S.T., Schödel, R., Buschek, D., Stachl, C., Winterhalter, V., Bühner, M., Hussmann, H., 2020. Developing a personality model for speech-based conversational agents using the psycholexical approach, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–14. URL: https://doi-org.ucd.idm.oclc.org/10.1145/3313831.3376210, doi:10.1145/3313831.3376210.

[72] Wang, N., Johnson, W.L., Mayer, R.E., Rizzo, P., Shaw, E., Collins, H., 2008. The politeness effect : Pedagogical agents and learning outcomes 66, 98–112. doi:10.1016/j.ijhcs.2007.09.003.

[73] Watts, R.J., 2003. Politeness. Cambridge University Press.

[74] Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Hendricks, L.A., Rimell, L., Isaac, W., Haas, J., Legassick, S., Irving, G., Gabriel, I., 2022. Taxonomy of risks posed by language models, in: 2022 ACM Conference on Fairness, Accountability, and Transparency, Association for Computing Machinery, New York, NY, USA. p. 214–229. URL: https://doi.org/10.1145/3531146.3533088, doi:10.1145/3531146.3533088.

[75] Weizenbaum, J., 1966. Eliza—a computer program for the study of natural language communication between man and machine. Communications of the ACM 9, 36–45.

[76] Weizenbaum, J., 1976. Computer power and human reason: From judgment to calculation. .

[77] World Health Organisation, 2001. The world health report 2001: Mental disorders affect one in four people. URL: https://www.who.int/news/item/28-09-2001-the-world-health-report-2001-mental-disorders-affect-one-in-four-peopl

[78] Wu, Y., Edwards, J., Cooney, O., Bleakley, A., Doyle, P.R., Clark, L., Rough, D., Cowan, B.R., 2020. Mental workload and language production in non-native speaker ipa interaction, in: Proceedings of the 2nd Conference on Conversational User Interfaces, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3405755.3406118, doi:10.1145/3405755.3406118.

[79] Yang, X., Aurisicchio, M., 2021. Designing conversational agents: A self-determination theory approach, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3411764.3445445, doi:10.1145/3411764.3445445.

[80] Zhou, S., Bickmore, T., Paasche-Orlow, M., Jack, B., 2014. Agent-user concordance and satisfaction with a virtual hospital discharge nurse, in: International conference on intelligent virtual agents, Springer. pp. 528–541.

[81] Zhou, S., Zhang, Z., Bickmore, T.W., 2017. Adapting a persuasive conversational agent for the chinese culture, in: 2017 International Conference on Culture and Computing (Culture and Computing), IEEE. pp. 89–96.

[82] Zojaji, S., Peters, C., Pelachaud, C., 2020. Influence of virtual agent politeness behaviors on how users join small conversational groups, in: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, pp. 1–8.

# Highlights

**Exploring how Politeness Impacts the User Experience of Chatbots for Mental Health Support**

Robert Bowman, Orla Cooney, Joseph W. Newbold, Anja Thieme, Leigh Clark, Gavin Doherty, Benjamin Cowan

- For the effective use of politeness in human-chatbot interaction, a nuanced application of politeness theory is needed.
- Politeness can positively impact the chatbot user experience; the Personal politeness chatbot was experienced as caring and encouraging.
- Politeness can negatively impact the chatbot user experience; the Passive politeness chatbot was experienced as too apologetic and condescending.

CREDIT:

**Robert Bowman:** Methodology, Investigation, Formal Analysis, Writing Original Draft, Review & Editing, Project Administration
**Orla Cooney:** Methodology, Investigation, Formal Analysis, Writing Original Draft, Review & Editing, Project Administration
**Joseph W. Newbold:**Conceptualization, Methodology, Writing Original Draft, Review & Editing
**Anja Thieme:** Conceptualization, Methodology, Writing Original Draft, Review & Editing, Funding acquisition, Supervision
**Leigh Clark**: Conceptualization, Methodology, Writing Original Draft, Review & Editing
**Gavin Doherty:** Conceptualization, Methodology, Writing Original Draft, Review & Editing, Funding acquisition, Supervision
**Benjamin Cowan:** Conceptualization, Methodology, Formal Analysis, Writing Original Draft, Review & Editing, Funding acquisition, Supervision

**Declaration of interests**

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Robert Bowman reports financial support was provided by Science Foundation Ireland. Robert Bowman reports financial support was provided by Microsoft Research. Orla Cooney reports financial support was provided by Science Foundation Ireland. Gavin Doherty reports financial support was provided by Science Foundation Ireland. Benjamin Cowan reports financial support was provided by Science Foundation Ireland. Robert Bowman reports a relationship with Amwell that includes: employment. Gavin Doherty reports a relationship with Amwell that includes: consulting or advisory and equity or stocks. Associate Editor IJHCS: Gavin Doherty. Associate Editor IJHCS: Benjamin Cowan