

Affective Conversational Agents: Understanding Expectations and Personal Influences

Javier Hernandez⁺, Jina Suh, Judith Amores,
Kael Rowan, Gonzalo Ramos, and Mary Czerwinski

Microsoft Research, Microsoft, Redmond, USA

Abstract—The rise of AI conversational agents has broadened opportunities to enhance human capabilities across various domains. As these agents become more prevalent, it is crucial to investigate the impact of different affective abilities on their performance and user experience. In this study, we surveyed 745 respondents to understand the expectations and preferences regarding affective skills in various applications. Specifically, we assessed preferences concerning AI agents that can perceive, respond to, and simulate emotions across 32 distinct scenarios. Our results indicate a preference for scenarios that involve human interaction, emotional support, and creative tasks, with influences from factors such as emotional reappraisal and personality traits. Overall, the desired affective skills in AI agents depend largely on the application’s context and nature, emphasizing the need for adaptability and context-awareness in the design of affective AI conversational agents.

Index Terms—AI Conversational Agents, Affective Computing, User Preferences

I. INTRODUCTION

The development of conversational AI agents has been marked by significant milestones that have transformed the landscape of human-computer interaction. From early pioneering systems like Eliza in the 1960s [1], which simulated conversation by using pattern matching and substitution techniques, to the complex, context-aware agents that dominate today’s technology ecosystem [2]. This progression has been shaped by a rich history of research [3], [4], as well as popular culture, where science fiction movies and literature have envisioned a future where human-like machine interactions are seamlessly integrated into our lives [5]. As AI conversational agents continue to advance and permeate various domains, understanding their implications and potential becomes increasingly important for researchers, developers, and users alike.

Text-based large language models have recently emerged as a major driving force in the field of AI conversational agents, demonstrating unique capabilities that set them apart from their predecessors. These models excel at encoding semantics, allowing them to communicate naturally with people and generate contextually relevant responses [6]. Furthermore, they exhibit emergent behaviors [7], which stem from their ability to learn from massive amounts of textual data [8]. Examples of such models include BERT [9], which has been employed for

various natural language processing applications such as sentiment analysis and question-answering systems [10], and more recently, OpenAI’s GPT-4, which has shown proficiency in tasks ranging from text summarization to code generation [11]. These models can operate as chatbots that can be customized to take different personas and skills, such as having the ability to recognize emotions associated with different perspectives, adjust their responses accordingly, and generate their own simulated states, enabling the development of more human-like and emotionally intelligent AI systems when appropriate.

Existing literature highlights simulated affective empathy or emotional empathy as one important factor in these human-agent interactions [12]–[14]. Affective empathy, as one of the core areas of the field of Affective Computing [15], is one component of general empathy, which involves emotions as well as the capacity to share and understand another’s state of mind, with continuous exchanges between emotion and intention in dialogue [16], [17]. Incorporating capabilities that embody affective empathy into systems has been shown to impact user satisfaction, trust, and acceptance of AI systems, as well as facilitate more effective communication and collaboration between humans and AI agents [18], [19]. As conversational AI agents are further adopted across various application domains, it becomes crucial to understand the potential role and expectations of affective empathy in designing and building AI conversational agents with simulated affective empathy.

Motivated by these considerations, this work seeks to answer the following research questions:

- RQ1** What are the specific preferences of individuals regarding the manifestation of affective AI conversational agents across a diverse range of applications, and how do these preferences differ based on the context and nature of the application?
- RQ2** What human factors such as personality traits and emotional regulation skills contribute to the variation in personal preferences for affective AI agents, and how do these factors influence the acceptance of affective qualities?

To address these questions, we conducted a survey with 745 respondents, exploring their preferences for AI agents that can perceive, respond to, and simulate emotions [15]. Our findings indicate that most of the participants would like

⁺Corresponding author: javierh@microsoft.com.

agents with affective skills but the desired level of affective intelligence, as well as the particular abilities (perceived, respond to, and simulate) depends largely on the nature of the application and personal human factors.

The paper is organized as follows. First, we review prior work on AI conversational agents, examining their applications, characteristics studied, and potential ethical concerns. Second, we delve into the methodology, detail the survey design and provide descriptive statistics about the data. Third, we present our analysis of the results, including preferences and contributing factors. Fourth, we discuss the limitations and implications of our work, along with potential avenues for future research. Finally, we provide some concluding remarks and relevant ethical considerations.

II. RELATED WORK

AI conversational agents have been extensively explored in various application areas, including search, education, health care, entertainment, customer service, and social support, to provide information, assistance, feedback, or companionship [15], [20]–[26]. A key challenge in developing these agents is whether or not to incorporate aspects of Affective Computing which enables agents to recognize, interpret, and respond to human emotions. This capability has been shown to enhance user experience, engagement, and satisfaction by allowing AI agents to establish and maintain social and emotional bonds with users [27]–[31]. Designing empathetic agents, however, requires not only selecting and integrating appropriate modalities to convey empathy but also considering the conversational style that best aligns with user preferences [32]–[35].

AI agents often employ natural language and other modalities, such as gestures and facial expressions, to interact with humans. These agents can take various forms, ranging from humanoid to non-humanoid and even non-physical entities [20]. Numerous comprehensive studies have investigated different aspects of agent voice, indicating that humans apply similar expectations to AI agents as they do to other humans, but these expectations may depend on factors such as the agent’s quality, age, and embodiment [36], [37]. Research on empathic AI agents, including those modeling empathy and prosocial behavior, has led to the development of systems that detect and regulate learners’ affective states, adapt to personality traits, elicit and express emotions, and facilitate social and emotional learning in various contexts [14], [14], [38]–[41]. However, it is crucial to note that people with different conversational styles and personality traits may evaluate expressive agents differently, emphasizing the need for further research on the role of conversational style and user preferences in empathetic agents’ effectiveness [42]–[45].

Despite the progress made in Affective Computing and AI conversational agents, several challenges and ethical implications still need to be addressed. These include accurately detecting and expressing emotions by the agent, the risk of over- or under-empathizing emotions, and potential emotional manipulation, which can have consequences for privacy,

accountability, and human dignity [32], [46]. Furthermore, most research to date has focused on specific scenarios or domains, without systematically exploring user preferences and expectations regarding affective empathy across a diverse range of applications, user backgrounds, and contexts.

III. METHODOLOGY

This section outlines our approach to investigating affective empathy in AI conversational agents which includes the details surrounding the survey design and data collection.

A. Survey Design

To explore people’s perspectives on affective empathy in AI conversational agents, we designed a survey with the following sections:

Demographics and background information. We collected basic demographic information about the participants, including their gender, age, job role, education level, and whether English was their first language.

Personality. To examine the role of personality on personal preferences, we included the 10-item Big Five Personality Inventory [47] which decomposes personality into five main components: openness, conscientiousness, extroversion, agreeableness, and neuroticism.

Emotional Experience. To assess various aspects of participants’ emotional experience, we included the 20-item Toronto Alexithymia Scale [48] which provides an overall score for the inability to recognize or describe one’s own emotions (a.k.a., alexithymia) as well as separate scores for the difficulty to describe and identify feelings, as well as the ability for externally oriented thinking. We also included the 10-item Emotion Regulation Questionnaire (ERQ) [49] which provides scores for cognitive reappraisal and expressive suppression. These measures were selected based on their relevance to emotional intelligence, communication, and empathy, which are key components in the context of AI conversational agents.

Prior experience. To help contextualize the findings of this work, we evaluated participants’ familiarity with the technology by asking them to rate the duration and frequency of their interactions with conversational AI agents, as well as describe some of their primary use cases.

Preferences. To capture preferences on affective empathy for AI conversational agents, participants were asked to indicate whether they would like to interact with an AI agent with simulated affective empathy to achieve different goals. Based on the seminal book of Affective Computing [15], we divided digital affective skills into three main components: the ability to perceive emotions from the user, the ability to respond to users’ emotions, and the ability to simulate its own emotions. For each of these abilities, we provided a definition and some examples (see Table I). These dimensions were chosen to cover some of the core aspects of affective empathy [15], ensuring that our measures provided a nuanced understanding of participants’ preferences in AI conversational agents. For each of these abilities, participants were then asked to rate

TABLE I
AFFECT IN CONVERSATIONAL AGENTS

<i>Ability</i>	<i>Definition</i>	<i>Positive Example</i>	<i>Negative Example</i>
Perceive	AI conversational agents could recognize the user’s emotions via input cues such as tone, word choice, and context.	The agent detects excitement from the user’s message, “I’m so thrilled with the results of this project!”	The agent detects disappointment from the user’s message, “I’m really bummed that things didn’t work out as planned.”
Respond	In the future, agents could adjust their behavior and responses based on the user’s emotions, expressing empathy or offering support.	The AI enthusiastically replies, “Congratulations on the amazing results! Your hard work has paid off. What’s the next step for this project?”	The AI compassionately replies, “It is too bad that things didn’t go as planned. It’s understandable to feel disappointed. How can I help you move forward?”
Simulate	Future agents could generate and express their own simulated states to enhance interactions with users.	The AI simulates experiencing joy, saying: “Your success brings me joy too! I’m thrilled to be a part of your journey and support your continued growth.”	The AI simulates experiencing sadness, saying, “I feel saddened to learn about the challenges you’re facing. Please know that I’m here to help you navigate through these tough times and find a way to overcome them.”

whether they would like to see them when interacting with an AI agent across 32 application areas.

We selected the list of application areas based on popular applications of text-based chatbots (e.g., browsing and retrieving content, generating personalized stories and narratives), as well as prevalent application areas of emotion recognition [50] (e.g., providing empathetic emotional support and assistance, facilitating customer service interactions). These applications can be grouped into 10 main thematic categories (e.g., customer service and sales, creativity, and content generation) as shown on Figure 1 (top).

Finally, to better understand potential individual variations in terms of affective preferences across all applications, we defined a Preference for Affective Conversations score (PAC) for each participant. We computed this score as follows:

$$PAC_p = \frac{\sum_{i=1}^n I_{ip}}{n} \times 100\% \quad (1)$$

where n is the total number of applications, and I_{ip} is an indicator variable that takes the value 1 if participant p indicated at least one empathy preference for the i^{th} application, and 0 otherwise.

Open-ended scenarios. Since the curated list of application areas might not have captured all possibilities, we asked participants to identify the ideal and worst application scenarios for an AI agent with affective empathy. For the ideal scenario, participants also provided information about the signals they would be willing to share with the AI conversational agent, as well as their preferences regarding the platform and potential channel of communication. To analyze open-ended questions, we conducted a thematic analysis regarding potential areas of application and concerns about affective AI agents. Each survey response was manually coded according to the themes we discovered to facilitate quantitative analysis.

B. Data Collection and Overview

A link to the survey was sent via email to a random set of information workers of a large technology company in the USA to request their preferences in terms of affective AI agents. Participation was voluntary without any form of

compensation, and completing the survey took approximately 15 minutes. The study was pre-approved by the Institutional Review Board.

A total of 745 information workers took part in the user study, representing diverse job roles, education levels, age groups, and gender identities. The majority of participants were in the software development/engineering field (55.84%), followed by product management (15.57%), and data science/analytics (3.62%). Other job roles accounted for 21.36% of the participants. The participants’ education levels were predominantly bachelor’s degrees (46.44%) and master’s degrees (35.30%). A smaller percentage held other education levels, with a few preferring not to disclose their education (1.34%). In terms of age distribution, the largest group of participants fell within the 26-35 age range (32.89%), followed by those aged 36-45 (28.72%) and 46-55 (20.81%). A smaller percentage belonged to other age groups (1.74%). Gender-wise, the majority of participants identified as men (70.20%), while 25.64% identified as women, and the remaining 1.47% identified as non-binary/gender diverse, self-described, or preferred not to disclose their gender.

IV. RESULTS

This section presents our findings on various aspects of affective empathy, exploring preferred application areas, differences across preferred affective skills, and factors that influence potential acceptance.

A. What prior experience does our sample population have with AI conversational agents?

Prior experience with agents can play a role in how people perceive and understand interactions with agents. In our survey, 56.24% of the participants reported having less than 6 months of experience with AI conversational agents, followed by 13.56% with 6 months to 1 year, 8.86% with 1-2 years, and 9.13% had more than 5 years of experience. Regarding the frequency of interaction, 6.58% of participants had never interacted with AI conversational agents, 15.57% had interacted rarely (once or twice in their life), 37.45% interacted occasionally (a few times a year), 22.68% interacted

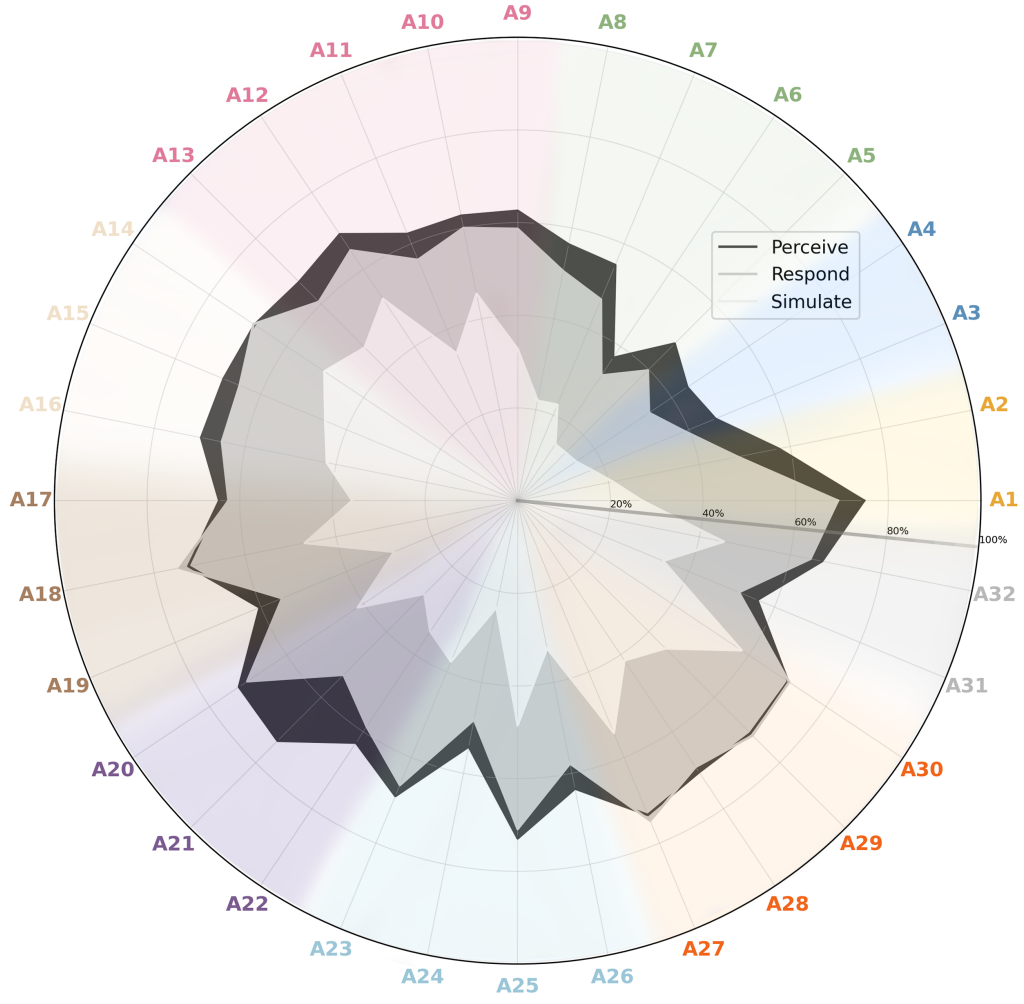
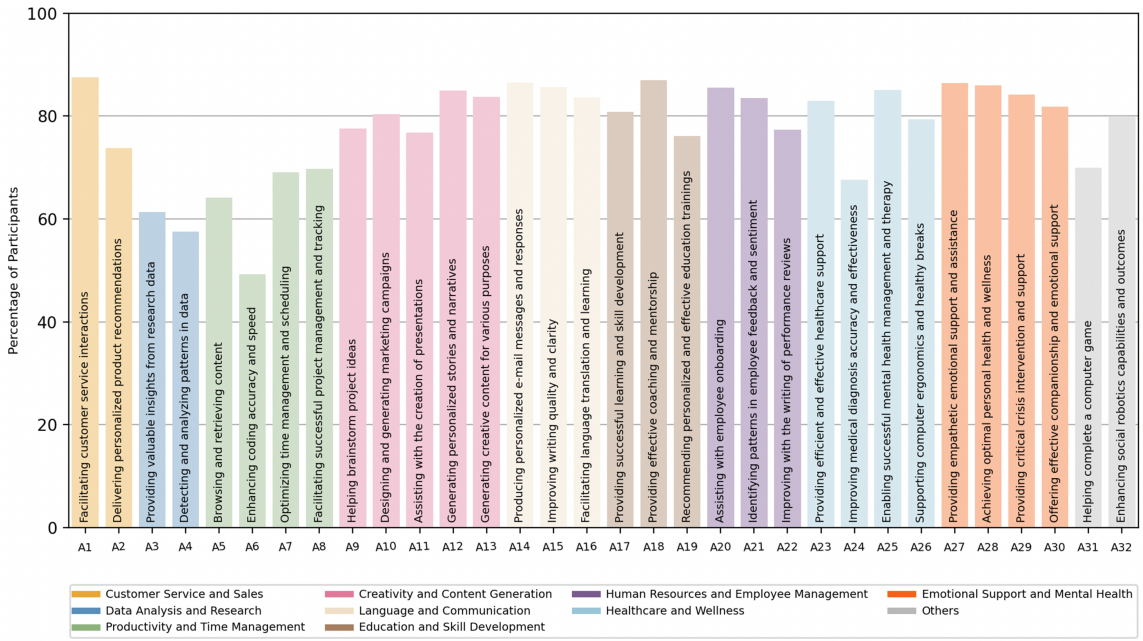


Fig. 1. Overall preference for Affective AI conversational agents across different application areas (top) and affective skills (bottom).

frequently (several times per month), and 17.72% interacted very frequently (multiple times per week or more). This highlights that most of the participants had at least occasional interactions with AI conversational agents.

Analysis of the free-form question regarding frequently explored applications revealed a variety of settings where respondents have used conversational AI agents, including customer service, personal assistance, healthcare, education, finance, e-commerce, entertainment, and writing, among others. Overall, these responses indicate that conversational AI agents are already widely used in numerous domains, suggesting a growing reliance on AI technology for various tasks.

B. What are the most preferred application areas for incorporating affective empathy?

Figure 1 (top) shows the percentage of participants who indicated that they would like to see some kind of affective empathy, which includes any of the three abilities (i.e., perceiving, responding to, or simulating emotions). As can be seen, there is large variability across applications with some consistency within categories. For instance, only around 50% of the participants expressed a preference for *affective empathy when enhancing coding accuracy and speed* (A6), but 84% favored its incorporation when *producing personalized email messages and responses* (A14). Interestingly, around 80% of the respondents expected to see some form of emotional skills in AI agents for a large number of applications (50%),

The most preferred categories for affective empathy were found to be *language and communication, emotional support and mental health* ($84.5\% \pm 1.5\%$), and *creativity and content generation* ($80.8\% \pm 3.2\%$). These categories align with the inherently human aspects of interaction, empathy, and creativity, where the presence of affect in AI agents could be perceived as more beneficial. Conversely, the least preferred categories for affective empathy were *data analysis and research* ($59.5\% \pm 1.5\%$), as well as *productivity and time management* ($63\% \pm 8.6\%$). In these categories, participants may prioritize the efficiency and accuracy of AI agents over agents' ability to demonstrate affective skills, viewing affective empathy as less essential to the tasks at hand.

As part of the survey, we incorporated open-ended questions to explore potential areas where affective AI agents might be advantageous or disadvantageous. The thematic analysis of beneficial scenarios revealed preferences for mental health and emotional support (27.7%), healthcare and wellness (21.9%), customer service and experience (20.7%), coaching and parenting (20.1%), content creation and language education/translation (10.7%), personal assistants and productivity (9.3%), entertainment and gaming (3.9%), and personalized recommendations (3.8%). For the most prominent theme on mental well-being and support for individuals, respondents emphasized applications that assist individuals in various social contexts, such as therapy and counseling, as well as applications that offer companionship for those experiencing loneliness, isolation, or challenging life situations.

The thematic analysis of non-beneficial cases uncovered a mix of less-than-ideal scenarios and general concerns related to the concept of affective empathy. The distribution of themes included scenarios that require a high level of objectivity (32.5%), those necessitating emotional support and healthcare in general (18.6%), situations demanding emotional authenticity and human interaction (17.2%), general privacy and security concerns associated with emotional data (7.5%), marketing and customer service scenarios due to potential manipulation (5.8%), and general opposition against emotional capabilities (5.5%). For the most prominent theme of objectivity, respondents provided examples that were primarily technical, data-driven, or objective, such as coding, data analysis, and research.

C. How do preferences vary across affective skills and application areas?

Figure 1 (bottom) shows the percentage of users who preferred to interact with AI agents having the ability to perceive (black), respond to (grey), and simulate (white) emotions across all the applications. The graph reveals that people generally expect AI agents to possess the ability to perceive emotions ($66.33\% \pm 12.07\%$ on average), followed closely by the ability to respond to emotions ($61.87\% \pm 12.86\%$). Approximately half of the respondents expect AI agents to be capable of simulating emotions ($35.47\% \pm 12.45\%$), although there is noticeable variance across settings. For certain applications, such as *producing personalized email messages and responses* (A14), *providing effective coaching and mentorship* (A18), *providing critical crisis intervention and support* (A29), and *offering effective companionship and emotional support* (A30), the scores for perceiving and responding to emotions were nearly identical.

Participants expressed a preference for AI agents that perceive and respond to emotions in applications involving human interaction or emotional support. Conversely, they expressed a preference for agents that can simulate emotions in creative or content generation tasks and emotional support. A lower preference for affective empathy, particularly for simulating emotions, was observed in data-focused or analytical tasks. Overall, the desired level of affect in AI agents appears to be highly dependent on the context and nature of the application. This underscores the importance of tailoring AI agent capabilities to meet the specific requirements and expectations of users across various domains.

When examining the open-ended questions, respondents shared diverse perspectives on the expected capabilities of affective AI agents. They acknowledged the potential benefits of AI agents that can accurately perceive and interpret users' emotional states, but also expressed concerns about privacy, data security, and the possibility of misinterpretation or emotional biases. Respondents also highlighted the advantages of AI agents capable of simulating emotional expressions to create more engaging, human-like, and relatable interactions. However, they raised concerns about authenticity and sincerity, as well as the potential for simulated empathy to be perceived

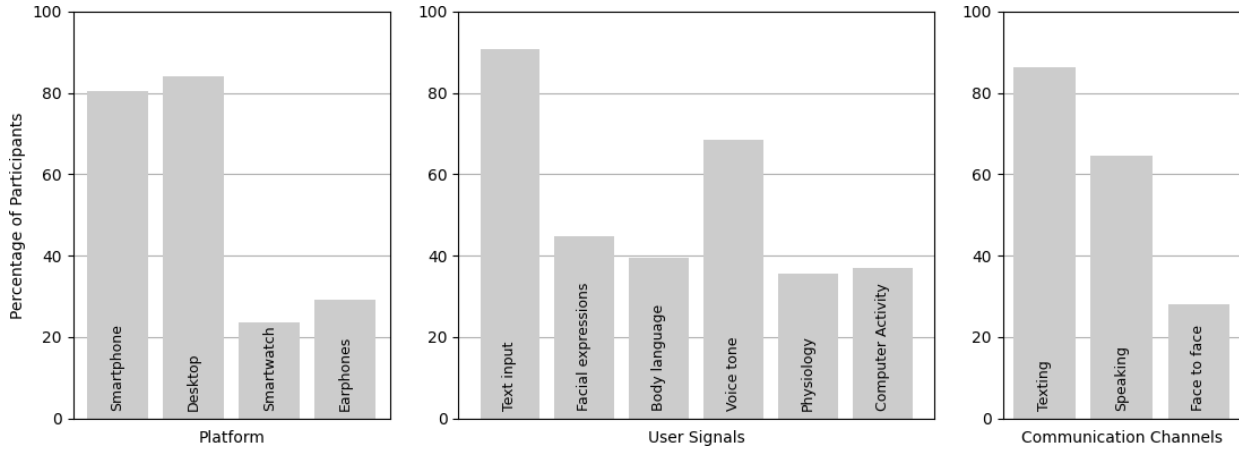


Fig. 2. Preferences in terms of platforms (left), shared user signals (center), and communication channels (right) for affective AI conversational agents.

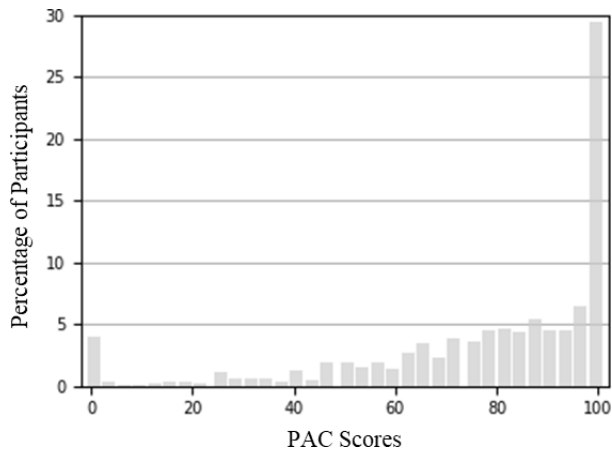


Fig. 3. Distribution of Preference for Affective Conversations (PAC) scores across participants. Higher PAC values indicate stronger preference towards seeing more affective empathy on AI conversational agents.

as manipulative, disingenuous, or even harmful. They worried that AI agents could exploit users’ emotional vulnerabilities. Interestingly, the contrasting views of benefits and concerns in areas such as mental health, emotional support, and customer service demonstrate the intricate and nuanced nature of public sentiment toward affective AI agents.

D. What are the preferred methods of communication with affective AI agents?

When evaluating preferences, respondents were asked to consider the application in which affective empathy made the most sense to them. Considering this application, Figure 2 (left) shows the preferred platform for interacting with AI agents. In this case, participants showed a significantly higher preference for desktop platforms and smartphones than other devices such as smartwatches and earphones ($\chi^2(1) = 938.57, p < 0.001$). Similarly, Figure 2 (center) shows the signals that participants would be happy to share to enable affective empathy as well as the percentage of participants who selected them. As can

be seen, the results indicate a significant preference for sharing text and voice data over other modalities such as facial expressions, body language, and physiological signals ($\chi^2(1) = 655.14, p < 0.001$). Finally, Figure 2 (right) shows the preferred communication channels. As can be seen, people preferred significantly more text-based interactions (messages) and audio (speaking) interactions than face-to-face interactions ($\chi^2(1) = 465.36, p < 0.001$).

These findings show some important preferences in terms of platforms, user signals, and communication channels when designing affective AI agents. Our survey did not allow us to delve much deeper into why these preferences emerged, as we were unaware of them before this survey. We intend to follow up with future research on device/signal preferences. Still, by being aware of and potentially accommodating these preferences, developers can create more preferred, engaging, and affective agents in the future.

E. How do emotional experience and personality traits impact acceptance of affective empathy?

Figure 3 shows the distribution of PAC scores across all participants, revealing large variance among individuals. For instance, around 30 participants did not select any affective skill across all applications, whereas around 220 people selected at least one empathy preference for all the applications.

To understand whether personal factors could explain PAC scores, we performed a linear mixed effects models in which we used the PAC scores as the dependent variable, and both personality and emotional experience information as the independent variables. All dependent variables were z-scored to ensure that a consistent scale was met across them. Table II shows the coefficients and statistical values, showing that multiple variables significantly contributed to the preference score. In particular, individuals with higher cognitive reappraisal skills contributed more significantly to a higher PAC score, suggesting that people who are able to reappraise their feelings effectively are more likely to prefer affective empathy. Similarly, higher ratings in difficulty identifying feelings decreased PAC scores, although less significantly. In

TABLE II
FACTORS CONTRIBUTING TO THE PREFERENCE FOR AFFECTIVE
CONVERSATIONS
(\cdot : $p < 0.1$, \ast : $p < 0.05$, $\ast\ast$: $p < 0.01$, $\ast\ast\ast$: $p < 0.001$).

	<i>Coefficient</i>	<i>t-value</i>	<i>p-value</i>
<i>Intercept</i>	0.000	0.000	1.000
<i>Cognitive Reappraisal</i>	0.125	3.273	0.001 $\ast\ast\ast$
<i>Expressive Suppression</i>	-0.064	-1.491	0.136
<i>Difficulty Describing Feelings</i>	-0.020	-0.260	0.795
<i>Difficulty Identifying Feelings</i>	-0.109	-1.850	0.065 \cdot
<i>Toronto Alexithymia Scale</i>	0.132	1.385	0.167
<i>Extroversion</i>	0.109	2.919	0.004 $\ast\ast\ast$
<i>Agreeableness</i>	0.067	1.799	0.072 \cdot
<i>Conscientiousness</i>	0.095	2.542	0.011 \ast
<i>Neuroticism</i>	-0.022	-0.579	0.563
<i>Openness</i>	0.054	1.486	0.138

terms of personality, participants ranking high in extroversion and conscientiousness and, in a lesser amount, agreeableness, contributed positively to the score. This finding indicates that people who score high on these factors tend to prefer affective empathy on AI agents.

Finally, we examined the correlation between PAC scores and the likelihood of expressing concerns on the open-ended questions. As expected, we found that the opposition to the agent having affective capabilities was significantly and negatively correlated with users’ preferences for affective capabilities overall ($r=-0.36$), similarly with perceiving ($r=-0.19$), responding to ($r=-0.28$), and simulating ($r=-0.24$) capabilities. Similarly, concerns about emotional authenticity and human interaction were significantly and negatively correlated with the preferences for simulating capabilities ($r=-0.19$).

V. DISCUSSION

This study provides valuable insights into people’s preferences regarding affective empathy in conversational AI agents across a variety of applications. We found that there is a wide range of empathy preferences depending on the application, with higher affective expectations for tasks that involve human interaction, emotional support, and creative or content generation tasks. In particular, *assisting in the writing of personalized e-mails and messages* (A14), *facilitating customer service interactions* (A1), and *providing effective coaching and mentorship* (A18) received the highest ratings. Moreover, we identified preferences in terms of platforms, sharing of sensing modalities, and potential channels of communication in scenarios where affective empathy may be helpful. Understanding these preferences can help researchers and practitioners identify areas of opportunity and prioritize potential research and development efforts, ultimately leading to more effective and engaging AI conversational agents.

Our work demonstrates that certain personal characteristics can influence users’ preferences in terms of affective empathy. We found that factors such as cognitive reappraisal skills and personality traits, such as extroversion and conscientiousness, can significantly contribute to users’ expectations of affective empathy in AI agents. Our research method preventing us from

knowing why these particular personality traits and characteristics drive certain preferences, but much further work needs to be done to explore the relationship. Also, these findings open an opportunity to explore the automated recognition of user preferences based on their personal information, which could be relevant in adjusting agents’ abilities and better matching user expectations. Ultimately, this personalized approach to AI agent design could help enhance user experiences and create more meaningful interactions.

As part of the study, we recruited participants from a large technology company. While this allowed us to recruit people with prior experience with AI conversational agents, it will be important to validate our findings against other demographics and cultures. Future studies should consider more diverse populations to capture the complexities of emotion, such as cultural differences and the wider range of user preferences. Additionally, further exploration of other factors that might influence affective preferences, such as culture or environment, could provide a more comprehensive understanding of users’ expectations.

The main focus of this work was to increase our understanding of people’s preferences on expected affect in AI agents. As such, we did not go into the actual implementation and experience of AI agents with affective skills. Further research is needed to assess the potential impact of AI agents with affective empathy on achieving the different goals, as well as to evaluate user satisfaction and agent effectiveness in real-world scenarios. Moreover, there is room for investigation methods to implement and assess different affective skills, such as using fine-tuning, acting-related prompts, or other novel approaches that balance affective skills with other agent functionalities.

VI. CONCLUSIONS

This work increases our understanding about people’s preferences regarding affective empathy in AI conversational agents across a variety of applications. By surveying 745 respondents, we characterized their expectations for affect, preferred channels of communication and user signals, as well as potential influences of personal information on these preferences. Our findings revealed a wide range of preferences and highlighted the importance of context and application type in shaping users’ expectations. Our research sheds light on the evolving field of conversational AI agents and the role of Affective Computing to help foster more meaningful and empathetic interactions.

VII. ETHICAL STATEMENT

Integrating affective empathy into AI agents is expected to present complex challenges that could potentially lead to user harms, including forced conformity, emotional harm, manipulation, and invasion of privacy. These risks may vary depending on the degree of empathy and the application domain, making it essential to carefully consider them when developing mitigation strategies. Some potential mechanisms

to address these risks involve promoting responsible communication and informed consent, as recommended by prior research [50]–[53]. When interpreting and applying the findings of this study, it is crucial to be aware of its limitations, such as the distribution of the studied population and differences when compared to other populations. By acknowledging these limitations and incorporating ethical considerations into the research process, we can work towards developing AI agents that not only purposefully exhibit affective empathy but also respect users’ autonomy and well-being, fostering a more ethically robust and user-centric perspective in future research endeavors.

REFERENCES

- [1] J. Weizenbaum, “Eliza—a computer program for the study of natural language communication between man and machine,” *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.
- [2] M. Hutson, “Artificial intelligence faces reproducibility crisis,” *Science (New York, NY)*, vol. 359, no. 6377, p. 725, 2018.
- [3] A. M. Turing, “Computing machinery and intelligence,” *Mind*, vol. 59, no. 236, pp. 433–460, 1950.
- [4] J. R. Searle, “Minds, brains, and programs,” *Behavioral and Brain Sciences*, vol. 3, no. 3, pp. 417–424, 1980.
- [5] I. Asimov, *I, Robot*. Gnome Press, 1950.
- [6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez *et al.*, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.
- [7] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, and Y. Zhang, “Sparks of artificial general intelligence: Early experiments with gpt-4,” 2023.
- [8] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, “Language models are unsupervised multitask learners,” *OpenAI Blog*, vol. 1, no. 8, p. 9, 2019.
- [9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.
- [10] C. Sun, X. Qiu, and X. Huang, “Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence,” *arXiv preprint arXiv:1903.09588*, 2019.
- [11] OpenAI, “Gpt-4 technical report,” 2023.
- [12] S. Brave, C. Nass, and K. Hutchinson, “Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent,” *International journal of human-computer studies*, vol. 62, no. 2, pp. 161–178, 2005.
- [13] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, “Empathic robots for long-term interaction,” in *International Conference on Social Robotics*. Springer, 2014, pp. 285–294.
- [14] A. Paiva, I. Leite, H. Boukricha, and I. Wachsmuth, “Empathy in virtual agents and robots: A survey,” *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 7, no. 3, pp. 1–40, 2017.
- [15] R. W. Picard, *Affective Computing*. The MIT Press, 1997.
- [16] K. Harrelson, “Intention and empathy,” *Philosophical Psychology*, vol. 33, no. 8, pp. 1162–1184, 2020.
- [17] A. Welivita, Y. Xie, and P. Pu, “A large-scale dataset for empathetic response generation,” in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 1251–1264.
- [18] S. Marsella, J. Gratch, P. Petta *et al.*, “Computational models of emotion,” *A Blueprint for Affective Computing-A sourcebook and manual*, vol. 11, no. 1, pp. 21–46, 2010.
- [19] L.-F. Rodríguez and F. Ramos, “Development of computational models of emotions for autonomous agents: a review,” *Cognitive Computation*, vol. 6, pp. 351–375, 2014.
- [20] T. Bickmore and R. W. Picard, “Establishing and maintaining long-term human-computer relationships,” in *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 12, no. 2. ACM, 2005, pp. 293–327.
- [21] L. Powell, M. Z. Nizam, R. Nour, Y. Zidoun, R. Sleibi, S. K. Warrior, H. A. Suwaidi, and N. Zary, “Conversational agents in health education: Protocol for a scoping review,” *JMIR Research Protocols*, vol. 11, 2022.
- [22] L. Martinengo, A. I. Jabir, W. I. W. T. Goh, N. Y. W. Lo, M. ho Ringo Ho, T. Kowatsch, R. A. Atun, S. Michie, and L. T. Car, “Conversational agents in health care: Scoping review of their behavior change techniques and underpinning theory,” *Journal of Medical Internet Research*, vol. 24, 2022.
- [23] S. Kusal, S. A. Patil, J. Choudrie, K. Kotecha, S. Mishra, and A. Abraham, “Ai-based conversational agents: A scoping review from technologies to future directions,” *IEEE Access*, vol. 10, pp. 92 337–92 356, 2022.
- [24] A. Osta, A. I. Kokkinaki, and C. Chedrawi, “Online health communities: The impact of ai conversational agents on users,” in *European, Mediterranean and Middle Eastern Conference on Information Systems*, 2021.
- [25] J. Simpson, F. Gaiser, M. Macik, and T. Breßgott, “Daisy: A friendly conversational agent for older adults,” *Proceedings of the 2nd Conference on Conversational User Interfaces*, 2020.
- [26] D. Ghosh and I. Faik, “Practical empathy: The duality of social and transactional roles of conversational agents in giving health advice,” in *International Conference on Interaction Sciences*, 2020.
- [27] A. Sharma, I. W. Lin, A. S. Miner, D. C. Atkins, and T. Althoff, “Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach,” in *Proceedings of the Web Conference 2021*, 2021, pp. 194–205.
- [28] I. Leite, C. Martinho, and A. Paiva, “Social robots for long-term interaction: a survey,” in *International conference on social robotics*. Springer, 2013, pp. 291–301.
- [29] I. Leite, A. Pereira, G. Castellano, S. Mascarenhas, C. Martinho, and A. Paiva, “Modelling empathy in social robotic companions,” in *Proceedings of the 19th International Conference on Advances in User Modeling*, ser. UMAP’11. Berlin, Heidelberg: Springer-Verlag, 2011, p. 135–147.
- [30] K. Chawla, R. Clever, J. Ramirez, G. M. Lucas, and J. Gratch, “Towards emotion-aware agents for negotiation dialogues,” *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1–8, 2021.
- [31] H. C. van Cuylenburg and T. N. D. S. Ginige, “Emotion guru: A smart emotion tracking application with ai conversational agent for exploring and preventing depression,” *2021 International Conference on UK-China Emerging Technologies (UCET)*, pp. 1–6, 2021.
- [32] L. Stark and J. Hoey, “The ethics of emotion in artificial intelligence systems,” in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, ser. FAccT ’21. New York, NY, USA: Association for Computing Machinery, 2021, p. 782–793.
- [33] D. Tannen, “Conversational style,” *Psycholinguistic models of production*, pp. 251–267, 1987.
- [34] P. Thomas, M. Czerwinski, D. McDuff, N. Craswell, and G. Mark, “Style and alignment in information-seeking conversation,” in *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, 2018, pp. 42–51.
- [35] H. Chin and M. Y. Yi, “Should an agent be ignoring it?: A study of verbal abuse types and conversational agents’ response styles,” *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019.
- [36] K. Seaborn, N. P. Miyake, P. Pennefather, and M. Otake-Matsuura, “Voice in human-agent interaction: a survey,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–43, 2021.
- [37] J. James, C. I. Watson, and B. MacDonald, “Artificial empathy in social robots: An analysis of emotions in speech,” in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2018, pp. 632–637.
- [38] O. C. Santos, “Emotions and personality in adaptive e-learning systems: an affective computing perspective,” *Emotions and personality in personalized services: Models, evaluation and applications*, pp. 263–285, 2016.
- [39] Ö. N. Yalçın and S. DiPaola, “Modeling empathy: building a link between affective and cognitive processes,” *Artificial Intelligence Review*, vol. 53, no. 4, pp. 2983–3006, 2020.
- [40] H. Rashkin, E. M. Smith, M. Li, and Y.-L. Boureau, “Towards empathetic open-domain conversation models: A new benchmark and dataset,” *arXiv preprint arXiv:1811.00207*, 2018.
- [41] L. Zhou, J. Gao, D. Li, and H.-Y. Shum, “The design and implementation of xiaoice, an empathetic social chatbot,” *Computational Linguistics*, vol. 46, no. 1, pp. 53–93, 2020.
- [42] P. Thomas, D. McDuff, M. Czerwinski, and N. Craswell, “Expressions of style in information seeking conversation with an agent,” in *Proceedings*

of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 1171–1180.

- [43] M. R. Leary and A. B. Allen, "Personality and persona: Personality processes in self-presentation," *Journal of personality*, vol. 79, no. 6, pp. 1191–1218, 2011.
- [44] N. R. Richendoller and J. B. Weaver III, "Exploring the links between personality and empathic response style," *Personality and Individual Differences*, vol. 17, no. 3, pp. 303–311, 1994.
- [45] P. Costa, R. Alves, I. Neto, P. Marvao, M. Portela, and M. J. Costa, "Associations between medical student empathy and personality: a multi-institutional study," *PloS one*, vol. 9, no. 3, p. e89254, 2014.
- [46] N. Bostrom, *Superintelligence: Paths, dangers, strategies*. Oxford University Press, 2014.
- [47] B. Rammstedt and O. P. John, "Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german," *Journal of Research in Personality*, vol. 41, no. 1, pp. 203–212, 2007.
- [48] R. M. Bagby, J. D. Parker, and G. J. Taylor, "The twenty-item toronto alexithymia scale—i. item selection and cross-validation of the factor structure," *Journal of Psychosomatic Research*, vol. 38, no. 1, pp. 23–32, 1994.
- [49] J. J. Gross and O. P. John, "Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being," *Journal of Personality and Social Psychology*, vol. 85, no. 2, pp. 348–362, 2003.
- [50] J. Hernandez, J. Lovejoy, D. McDuff, J. Suh, T. O'Brien, A. Sethumadhavan, G. Greene, R. Picard, and M. Czerwinski, "Guidelines for assessing and minimizing risks of emotion recognition applications," in *2021 IEEE 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2021.
- [51] L. Stark and J. Hoey, "The ethics of emotion in artificial intelligence systems," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, pp. 782–793.
- [52] D. Ong, "An ethical framework for guiding the development of affectively-aware artificial intelligence," in *2021 IEEE 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2021.
- [53] R. Cowie, *Ethical issues in affective computing*. Oxford University Press, 2015, pp. 334–348.