Computers in Human Behavior: Artificial Humans 6 (2025) 100223

Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-artificial-humans

First interactions with generative chatbots shape local but not global sentiments about AI

Eva-Madeleine Schmidt [a,b,*], Clara Bersch [a,b], Nils Köbis [c], Jean-François Bonnefon [d,e], Iyad Rahwan [a], Mengchen Dong [a]

[a] Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany
[b] Max Planck School of Cognition, Leipzig, Germany
[c] Research Center for Trustworthy Data Science and Security, University of Duisburg-Essen, Duisburg, Germany
[d] Toulouse School of Economics, CNRS, University of Toulouse Capitole, Toulouse, France
[e] Institute for Advanced Study in Toulouse, University of Toulouse Capitole, Toulouse, France

A B S T R A C T

As artificial intelligence (AI) chatbots become increasingly integrated into everyday life, it is important to understand how direct interaction with such systems shapes public sentiment toward AI more broadly. Leveraging a unique window in April 2023—when many individuals still had little or no experience with such systems—we combined experimental manipulation (chatbot exposure vs. no exposure) with natural variation in real-world AI usage. In a preregistered proof-of-concept experiment ($N = 220$), we investigated whether a short conversation with a GPT-3.5-based chatbot influenced participants' sentiments across multiple dimensions of AI perception. We assessed system-specific fear, user engagement, anthropomorphization, and potential spillover effects to other domains, including AI in medicine, recruitment and governance. Results show that direct interaction reduced fear and increased enjoyment of the chatbot itself, while fostering a more critical, realistic understanding of its abilities. However, spillover effects were limited: exposure led to reduced fear of AI in familiar, concrete domains (e.g., medical applications), but not in more abstract or speculative areas. Hope about AI's societal potential remained unaffected. Our findings highlight that sentiments toward AI are multidimensional and context dependent. Exposure to AI chatbots can shift immediate attitudes but does not necessarily generalize to broader AI perceptions, underscoring the need for more targeted engagement strategies in shaping public understanding and trust.

1. Introduction

The use of artificial intelligence (AI) chatbots has rapidly expanded in the last few years, permeating both professional and personal areas of life (Rahwan et al., 2019). Among these, ChatGPT stands out as a widely adopted tool, with millions of users leveraging it daily for diverse applications (Roth, 2024), including education, customer service, and healthcare. By transforming productivity (Fauzi et al., 2023), writing practices (Mahapatra, 2024; Song & Song, 2023), and even verbal communication (Yakura et al., 2024), ChatGPT exemplifies the profound impact that chatbots and AI in general can have.

As the scope, depth, and duration of interactions with AI chatbots grows (Starke et al., 2024) it becomes increasingly relevant to understand how *these interactions shape public opinions toward AI:* The development and adoption of beneficial AI technologies, as well as a critical assessment of their potential risks, depend on public acceptance, trust, and active integration into everyday life (De Freitas et al., 2023; Glikson & Woolley, 2020; Liao, 2020). The widespread use of AI chatbots may further influence how people perceive and interact with other AI technologies, by creating a feedback loop: If people form positive (or negative) sentiments about AI in general, based on their current interaction with AI chatbots, they can be more likely to seek (or avoid) future interaction with other AI systems and support (or oppose) future AI development.

One theoretical framework that offers insight into how exposure might shape attitudes toward AI systems is *Intergroup Contact Theory* (ICT; Pettigrew & Tropp, 2006). ICT proposes that positive interactions between members of different social groups can reduce prejudice and improve intergroup relations. And the reduction of negative sentiments toward outgroup members can happen via two main mechanisms. The first is mere exposure, which increases familiarity and liking, thereby fostering more positive sentiments. The second is uncertainty reduction, which alleviates anxiety about the outgroup by providing concrete information about its behaviors and abilities, thereby reducing negative sentiments (Pettigrew, 2008).

Although originally developed for human–human intergroup contexts, ICT has been extended to human–robot interaction (Haggadone et al., 2021). Haggadone and colleagues demonstrated that brief,

* Corresponding author. Max Planck Institute for Human Development Lentzeallee 94, 14195, Berlin, Germany.
E-mail address: eschmidt@mpib-berlin.mpg.de (E.-M. Schmidt).

https://doi.org/10.1016/j.chbah.2025.100223
Received 28 August 2025; Received in revised form 16 October 2025; Accepted 25 October 2025
Available online 27 October 2025

structured exposure to a humanoid robot could meaningfully reduce social distance toward robots as a social group. In other words, brief contact with one robot was sufficient to override preexisting negative attitudes toward robots as a broader outgroup category.

In the present study, we examine whether similar principles might apply to contemporary AI chatbots. Without being given specific instructions, do people naturally adopt a positive or negative interactive tone when conversing with an AI chatbot? And what are the causal, downstream effects of even a brief conversation? Applying ICT to this context suggests that even brief, text-based interaction with an AI chatbot could reduce system-specific fears and foster more realistic attitudes through increased familiarity or reduced uncertainty. However, unlike the physical robots examined in earlier ICT work, contemporary AI chatbots, such as ChatGPT, differ in important respects: they are non-embodied and they rely on advanced generative language capabilities that enable open-domain, adaptive conversation. These features may fundamentally alter the nature and impact of contact, potentially leading to different outcomes from those reported in earlier robot-focused studies.

Although ICT provides a valuable framework for predicting when and how contact with AI chatbots might shift attitudes, it primarily addresses the direction of change (i.e., reducing negative sentiment or increasing positive sentiment). Importantly, compared to robots studied in prior work, AI chatbots such as ChatGPT can be more cognitively sophisticated and socially versatile, and people often generalize their beliefs about an AI system's capabilities or limitations from one domain to another (Longoni et al., 2023). However, we currently lack robust causal evidence on how contact with one AI system can have spillover effects and shape attitudes toward other AI systems or applications.

This gap can be attributed to two key factors. The first is *limited first-hand experience to advanced AI tools by the general public in the past:* Before tools like ChatGPT became widely available, public perceptions of AI were largely shaped by cultural narratives, media portrayals, and speculative fiction (Ge et al., 2024; Yam et al., 2023). These narratives can include science fiction stories of "robot rebels" and fears of AI achieving consciousness and becoming a threat, as epitomized by the "Terminator" archetype (Cave & Dihal, 2019). Media exposure in different countries can also shape public sentiments surrounding advanced technologies, cultivating utopian visions from the East and dystopian visions from the West (Dong et al., 2023; Yam et al., 2023).

The second key factor is *methodological limitations in existing research.* Previous studies on algorithmic aversion or algorithmic appreciation often rely on hypothetical scenarios, presenting simplistic descriptions of AI systems, without enabling participants to develop comprehensive impressions of AI through real-world interactions (Dong et al., 2024; Mahmud et al., 2022). Importantly, existing correlational studies cannot provide direct evidence on how AI exposure shapes general sentiments toward AI (Molina & Sundar, 2022; Shinners et al., 2023). For example, people with higher educational background and socioeconomic status tend to report less fear of AI (Liang & Lee, 2017). This could be explained by exposure to AI in two opposite directions: On the one hand, these people may be more likely to access and adopt advanced technologies; on the other hand, they may encounter less direct exposure to AI due to a lower likelihood of facing work automation and algorithmic management.

To fully capture the complexity of public attitudes toward AI, it is also important to consider the structure of these sentiments across different dimensions and contexts. Here, we argue that understanding public sentiment toward AI requires a multidimensional perspective. People's attitudes can span a wide range of concerns—from privacy risks to existential threats—as well as hopes, such as increased efficiency and fairer systems free from human bias. Rather than approaching public sentiments about AI as a uniform and predominantly negative construct (e.g., algorithm aversion[14]), we present a more balanced framework as presented in Table 1. This framework highlights two key points: (1) sentiments about AI are often more nuanced than simple reluctance, encompassing both positive and negative components, and (2) these valenced sentiments can be structured at multiple levels and shaped to varying degrees by the use and diffusion of AI chatbots.

Taken together, to illustrate the multidimensional nature of public sentiments about AI and the influence of AI chatbot exposure, we present a unique dataset collected during the early diffusion of ChatGPT. Data collection began in April 2023, when many individuals had little or no direct experience with such systems. By combining experimental manipulation (chatbot exposure vs. no exposure) with natural variation in real-world usage, we examined both direct responses to the chatbot and potential spillover effects on broader AI-related attitudes. We find that direct interaction with an AI chatbot reduces negative sentiments and increases positive sentiments about the chatbot itself. However, these effects do not extend systematically to other domains. Spillover effects are nuanced, reducing negative sentiments in some contexts (e.g., medical applications) while leaving other sentiments, such as hope about AI's societal impact, largely unchanged.

2. Methods

2.1. Participants

A total of 220 participants (47.7 % identifying as women, 48.6 % identifying as men, and 3.6 % others) from the United States were recruited through Prolific. Participants were required to be fluent in English and received monetary compensation at rates consistent with Prolific's fair-pay guidelines (9 GBP per hour). Prolific has been shown to yield high-quality and reliable data, comparable to or exceeding that of other online participant pools (Peer et al., 2021). The dataset and the code for the analysis are available on the Open Science Framework repository [https://osf.io/5szfk/?view_only=217dfc43e7ca4f9baa7f3362e18809d2]. Anonymized conversational data and code for setting up the GPT-3.5 chatbot will be made available to readers upon request. The study was approved by the Ethics Board at the Max Planck Institute for

Table 1
Dimensions of sentiments about Artificial Intelligence (AI).

Dimension	Target	Manifestations	Examples
Product-specific sentiments	Characteristics of specific AI products	Fear privacy and autonomy loss, errors, or bias. Hope for personalization and efficiency.	Concerns about chatbots using personal data or producing biased recommendations.
Ethical adoption	Use and misuse of AI tools in specific domains	Concern about prejudice and lack of accountability. Hope for consistency and fairness.	Misuse for recognition, prediction, decision-making, in medicine, HR, and content recommendation.
Human identity sentiments	AI systems imitating humans	Fear of loss of empathy or human connection, discomfort with human-like AI. Hope for companionship and mental health support.	Fear of impersonal interactions, or discomfort with human-like AI exhibiting "creepy" traits.
Human replacement	Autonomous roles typically held by humans	Fear of displacement in critical social roles. Hope for AI enhancing productivity.	Concerns about automation in medicine, education, or creative professions.
Existential impact	Broader societal and global impacts	Threats to safety, stability, or human oversight. Hope for societal improvement and solving global challenges.	Scenarios involving autonomous weapons or superintelligent AI.

Human Development and was preregistered at AsPredicted.org [https://aspredicted.org/KYV_VB4]. Data collection took place in April 2023.

2.2. Procedure

After providing informed consent, participants were randomly assigned to one of two conditions. In the *Exposure condition* ($n = 110$), participants engaged in a 7-min chat with a GPT-3.5-powered chatbot before completing the survey. In the *Control condition* ($n = 110$), participants proceeded directly to the survey after receiving a brief explanation of what a chatbot is (see Supplementary Material, Material 1A).

The chatbot interface was developed using React, with MongoDB for data storage and Node.js for managing chat sessions and interactions with the OpenAI API. To maintain conversation context, each API request contained five user-chatbot exchanges. The interface contained a simple text field and message window without any visual avatar. Upon entering the chat, a pop-up window instructed participants to click Start to initiate the conversation, after which the chatbot greeted them and invited open conversation (e.g., about hobbies or personal interests). Participants were asked to chat for at least 7 min, after which they proceeded to the questionnaire; full instructions and additional interface are provided in the Supplementary Material (Material 1B and C).

After talking to the chatbot or imagining the conversation, participants completed a survey reflecting the multi-dimensional construct of sentiments about AI, including, for example, fear of the current AI system, threat of AI applied in high-stakes domains (e.g., medicine and recruitment), and sentiments regarding AI's potential in governance and social contexts.

2.3. Measures

The study employed a multilevel measurement approach to capture both direct and indirect effects of chatbot exposure. Measures ranged from immediate, system-specific evaluations (e.g., fear of the chatbot, anthropomorphization, user experience) to broader assessments of attitudes toward AI applications in various domains, human-like occupational roles, and societal contexts. This structure allows us to examine whether short-term exposure effects extend beyond the local interaction.

2.3.1. System-specific fear, anthropomorphization, and user experience

To assess participants' immediate responses to the chatbot, we measured fear, anthropomorphization, and user experience. To measure fear, participants rated their agreement with the statement, "The chatbot's capabilities are intimidating" in the Exposure condition and "Such a chatbot's capabilities are intimidating" in the Control condition (from "1 = not at all" to "100 = extremely"). Anthropomorphization was measured as the extent to which users perceive the system as having human-like qualities (e.g., "The chatbot has learned to think like a human"). User experience and attitudes toward the AI chatbot were evaluated using a self-developed 7-item scale, where participants in the Exposure condition rated statements like, "I enjoyed talking to the chatbot," and those in the Control condition rated hypothetical versions of the same items (e.g., "I would enjoy talking to such a chatbot"). Together, these measures provide an assessment of participants' immediate emotional and experiential responses to interacting with the chatbot.

2.3.2. Fear of AI tools in specific domains and human-like roles, and hope about AI's societal applications

To explore spillover effects, we examined whether interaction with the AI chatbot influenced sentiments toward AI tools in broader contexts. Specifically, we assessed three key aspects of AI-related sentiments: fear of AI tools used in specific domains, fear of AI in human-like roles, and hope about AI's societal applications.

First, we assessed fear of AI tools used in specific domains. This dimension reflects concerns about the practical risks posed by AI tools being (mis) used in real-world decision-making. These sentiments may stem from product-specific concerns and ethical adoption (see Table 1) and may arise from fears of errors, bias, or privacy risks in AI applications. To measure this, we assessed participants' fear of AI systems in domains such as medicine, recruitment, and product recommendation using the Threats of Artificial Intelligence (TAI; Kieslich et al., 2021) scale. Participants rated items like, "When you think about the use of AI in medicine, how threatening do you consider computer systems with artificial intelligence that diagnose diseases?" (from "1 = not at all" to "100 = extremely").

Second, we assessed fear of autonomous AI acting in human-like, occupational roles (e.g., doctor, judge, life coach, financial advisor). This dimension captures speculative concerns about AI occupying roles traditionally held by humans, reflecting human replacement and human identity concerns (see Table 1). To address this, participants rated their level of discomfort with AI occupying various occupational roles (from "1 = not at all" to "100 = extremely").

Third, we measured positive sentiments toward AI in societal applications. This series of statements reflects hopes about AI's potential to improve governance, enhance social systems, and create beneficial societal outcomes. This measure allowed us to evaluate whether interaction with the AI chatbot influenced participants' optimism about AI's societal applications. Participants rated their agreement with statements about AI's potential benefits in personal and governance contexts, such as "AI technologies can facilitate human communication" or "AI can reduce corruption in governance" (from "1 = not at all" to "100 = extremely"). Together, these measures capture broader, domain-specific and societal-level sentiments, allowing examination of whether local exposure effects extend beyond the immediate chatbot interaction. Overall, this multilevel structure enables us to distinguish local exposure effects on chatbot-specific perceptions from potential spillover to broader attitudes about AI.

In addition, we assessed general fear of AI by asking participants to reflect on how the conversation with the chatbot affected their overall fear of AI ("How do you think the conversation with the chatbot affected your general fear of AI?" in the Exposure condition, versus "How do you think a conversation with such a chatbot would affect your general fear of AI?" in the Control condition).

3. Analysis

We conducted both parametric and nonparametric tests to compare the mean values between the Exposure and the Control conditions. Since the parametric tests were preregistered, we report them in the main text. The nonparametric tests, which are based on less stringent assumptions and are robust to deviations from the normal distribution, are reported in the Supplementary Materials, which yielded similar results. For all tests, we ran linear regression with condition as the predictor unless otherwise stated. All analyses were conducted in R Studio (version 4.0.4).

4. Results

4.1. Demographic information and experience with the AI chatbot

Randomization checks confirmed that treatment and control groups did not differ significantly on demographics (age, political stance, gender, education), nor did the distribution of prior experience differ systematically between the groups. Independent-samples *t*-tests showed no significant differences between the treatment and control groups in age, $t\,(215.05) = -0.90$, $p = .370$, 95 % CI [–4.76, 1.78], or political stance, $t\,(217.87) = 0.06$, $p = .955$, 95 % CI [–7.01, 7.42]. Fisher's exact tests indicated no significant group differences in prior ChatGPT experience, Replika experience, gender, or education.

To explore how prior familiarity with ChatGPT might influence participants' responses, participants were categorized into two groups based on their self-reported ChatGPT experience: No Experience (32.3 %, $n = 71$) vs. Some Experience (67.7 %, $n = 149$).

4.2. Sentiments toward the AI chatbot

Overall, participants who interacted with the AI chatbot reported reduced fear of the system, held a more critical view of its human-like qualities (anthropomorphization), and reported positive engagement with the chatbot (user experience). Additional statistical analyses for each of these dimensions are provided in the following sections.

4.2.1. Fear of the AI chatbot

As shown in Fig. 1, participants in the Exposure condition exhibited significantly less fear than those in the Control condition ($\beta = -15.182$, $SE = 3.7$, $t = -4.103$, $p < .001$) with a medium effect size ($d = 0.55$, 95 % CI [0.28, 0.82]). This reduced fear was observed both for participants who had no experience at all (β = - 13.154, $SE = 6.462$, t = - 2.035, p = .0456, Cohen's d = 0. 49 [95 % CI [0.01, 0.96]]) and those who had some prior experience (β = - 15.846, $SE = 4.547$, t = - 3.485, $p < .001$, Cohen's $d = 0.57$ [95 % CI [0.24, 0.90]]).

4.2.2. User experience of the AI chatbot

We conducted an analysis of variance (ANOVA) using aligned rank-transformed data to account for the non-normality of our data and to allow for robust testing of main and interaction effects in factorial designs. The analysis revealed significant effects of condition (F (1,1526) = 7.2701, $p = .0071$), question item (F (6,1526) = 77.9315, $p < .0001$), and their interaction (F (6,1526) = 8.3502, $p < .0001$). These results indicate that attitudes toward the AI chatbot differ between the Exposure and Control groups, but the significant interaction suggests that these differences are not consistent across all question items.

To explore these item-specific differences further, we conducted Wilcoxon rank-sum tests for individual items, as this non-parametric method is well-suited for comparing distributions in data that deviate from normality. Significant differences between the Exposure and Control groups were observed for several engagement-related measures. Participants in the Exposure group reported greater enjoyment interacting with the chatbot ($W = 4699.5$, $p = .0042$), were more likely to express a willingness to talk to the chatbot again ($W = 4137.5$, $p < .0001$), and were more willing to pay for the service ($W = 7410.5$, $p = .0017$) compared to those in the Control group.

However, no significant differences were found between the groups for perceptions of the chatbot's ability to provide social advice ($W = 6351.5$, $p = .5233$), teaching capabilities (W = 6816, $p = .1043$), trustworthiness ($W = 5487$, $p = .2325$), or factual accuracy ($W = 5711.5$, $p = .4734$). These findings suggest that direct interaction improves certain aspects of user engagement, such as enjoyment, willingness to re-engage, and perceived value, but does not significantly alter perceptions of the AI system's competence in providing advice, teaching, or factual reliability.

4.2.3. Anthropomorphization

Participants in the Exposure condition did not perceive the AI system as thinking like a human significantly more or less than those in the Control condition ($\beta = 1.927$, $SE = 4.125$, $t = -0.467$, $p = .641$). Similarly, no significant differences were found regarding perceptions of the system's consciousness (β = - 3.055, $SE = 3.841$, $t = -0.795$, $p = .427$). However, participants in the Exposure condition were less likely to believe the system was forming an opinion about them (β = - 10.982, $SE = 3.974$, $t = -2.763$, $p < .01$). Additionally, they found the system significantly less impressive compared to how participants in the Control condition imagined a chatbot would perform (β = - 12.109, $SE = 3.789$, $t = -3.196$, $p < .01$). These findings suggest that direct interaction with the AI system reduces idealized perceptions of its abilities, potentially fostering a more realistic understanding of its capabilities. While there was no credible evidence that participants altered their views on the system's fundamental human-like traits, they appeared less

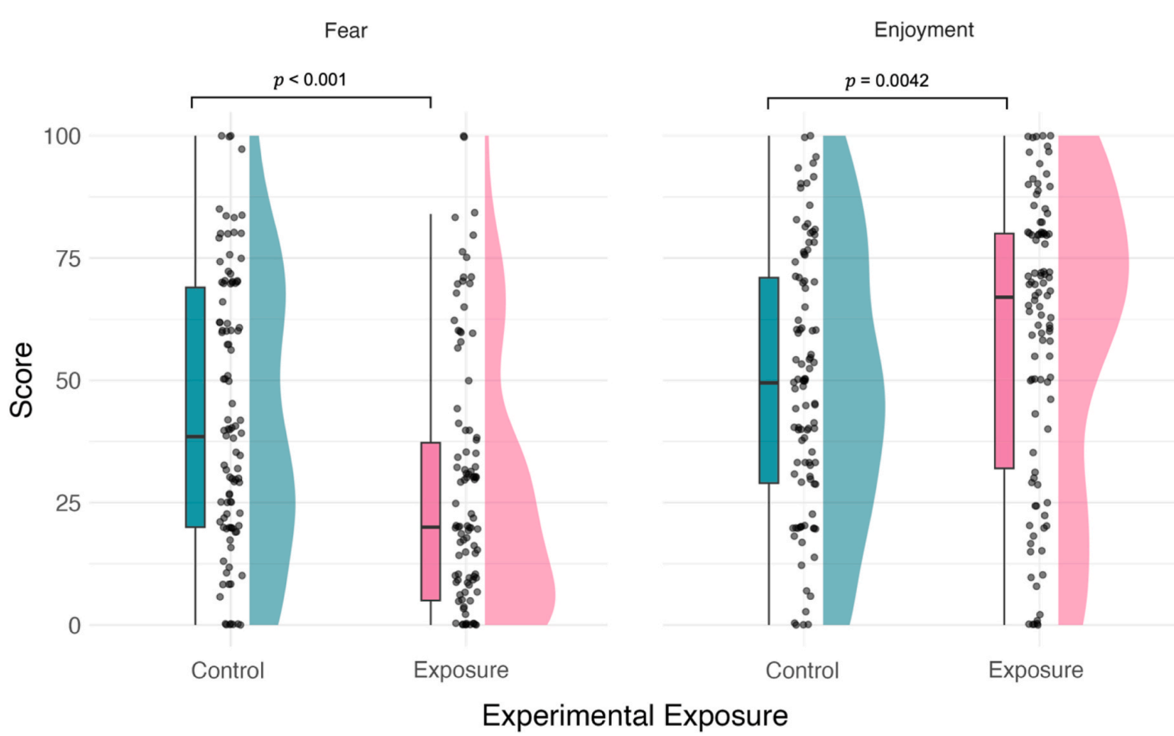

Fig. 1. Comparison of Fear and Enjoyment Scores by Experimental Exposure

Note. Fear and enjoyment scores are displayed for participants in the Control: No Interaction (Survey Only) condition and the Exposure: Interaction with Chatbot condition. Error bars represent the interquartile range, and data points are jittered for visibility. Participants who interacted with the chatbot reported significantly lower fear (negative sentiment) and significantly higher enjoyment (positive sentiment) compared to those in the control condition.

likely to attribute advanced social or evaluative qualities to the AI system after interacting with it.

4.3. Sentiments toward AI tools in specific contexts

Overall, participants who interacted with the AI chatbot reported reduced fear of AI tools, particularly in concrete domains such as medical applications, and showed a trend toward reduced fear in human-like occupational roles. However, there was no credible evidence for differences in general fear of AI or to hope about AI in broader personal and governance contexts, as outlined below.

4.4. Fear of AI tools in specific domains

Overall, as shown in Fig. 2, participants in the Exposure condition reported significantly less fear of AI tools compared to the Control group ($\beta = -6.402$, $SE = 2.492$, $t = -2.569$, $p = .0109$), with a small to medium effect size (Cohen's $d = 0.35$, 95 % CI [0.08, 0.61]).

Further analysis for each subdomain revealed that in the medical domain, participants in the Exposure condition reported significantly lower fear levels compared to those in the Control condition ($\beta = -8.682$, $SE = 2.944$, $t = -2.949$, $p = .00353$), with a small to medium effect size (Cohen's $d = 0.40$, 95 % CI [0.13, 0.66]). However, there was no credible evidence for such a difference in the recruitment domain ($M_{Exposure} = 47.680$, $M_{\text{Control}} = 53.080$, $\beta = -5.400$, $p = .127$) nor in the product recommendation domain ($M_{\text{Exposure}} = 32.87$, $M_{\text{Control}} = 37.995$, $\beta = -5.125$, $p = .0777$).

We further explored whether these effects differed between participants with and without prior experience with AI tools. The results showed similar patterns across both groups: while participants in the Exposure condition had descriptively lower fear of AI tools compared to those in the Control condition, these differences were not statistically significant (β = - 6.978, $p = .0631$). Detailed subgroup analyses are provided in the Supplementary Material (Material 2).

4.4.1. Fear of autonomous AI in human-like roles

We examined whether interacting with the chatbot influenced participants' fear of AI in various human occupational roles (e.g., doctor, therapist; see Fig. 3) using a linear regression analysis. There was no credible evidence that participants in the Exposure condition reported less overall fear compared to the Control group ($\beta = -6.018$, $SE = 3.235$, $t = -1.861$, $p = .0641$). Furthermore, while they reported significantly less fear for specific roles, including doctors ($p = .008$), legal advisors ($p = .037$), life coaches ($p = .013$), and financial advisors ($p = .008$, these differences did not remain significant after applying a Bonferroni correction to control for multiple comparisons.

4.4.2. Hope in governance and social contexts

We found positive exposure effects on enjoyment for the chatbot interacted, but no evidence that these positive effects generalized to broader sentiments in personal contexts (e.g., communication, companionship) or governance contexts (e.g., efficiency, anti-corruption). Detailed results are provided in the Supplementary Material (Material 3).

4.4.3. General fear of AI

Across the full sample, we observe no significant effect of exposure on general fear of AI ($\beta = -0.100$, $p = .97$). There was no credible evidence for interactions with prior experience ($ps >. 505$) These results suggest that direct interaction with the chatbot did not significantly influence participants' general fear of AI, regardless of their prior experience with such systems.

4.5. Exploratory qualitative word analysis

Given that AI chatbots can be applied in different domains and the conversational topics may also influence sentiments about AI, we analyzed the chat content in the Exposure condition to examine word frequencies and identify key conversational themes and sentiment. Detailed methodological descriptions can be found in the Supplementary Material (Material 4). This analysis is exploratory in nature and not intended as confirmatory evidence; rather, it provides contextual insight into how participants spontaneously engaged with the chatbot in a brief, unstructured setting.

Participants primarily engaged in casual discussions, reflecting a conversational tone focused on personal interests, such as movies, food, and other hobbies. The most frequently mentioned words were "like" (266 mentions), "movies" (95), "favorite" (89), "love" (60), "movie" (58), and "food" (46). Latent Dirichlet Allocation (LDA) was applied to uncover hidden topics within participant messages. The analysis

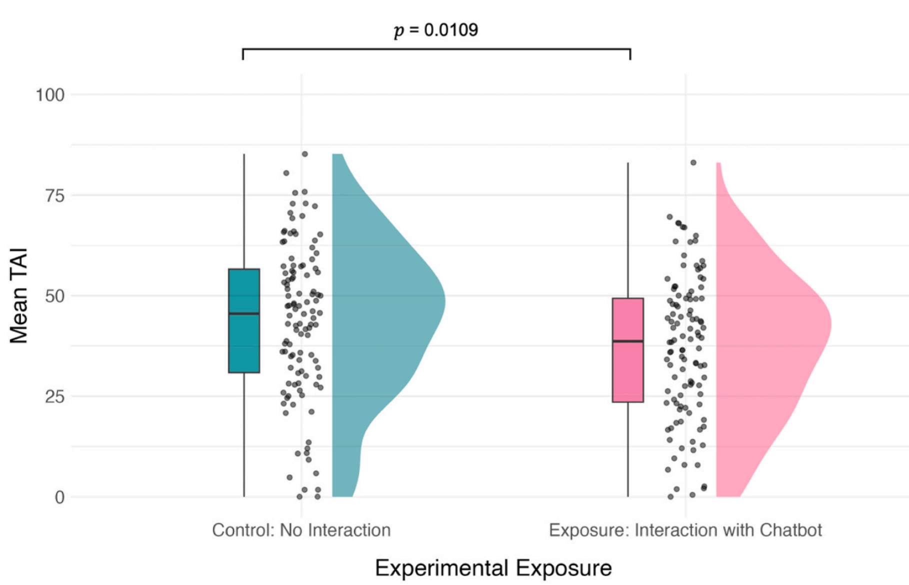

Fig. 2. TAI Scores by Experimental Exposure
Note. Mean Threats of Artificial Intelligence (TAI) scores are displayed for participants in the Control: No Interaction (Survey Only) condition and the Exposure: Interaction with Chatbot condition. Error bars represent the interquartile range, and data points are jittered for visibility. Participants who interacted with the chatbot reported significantly lower perceived threats of AI compared to those in the control condition.

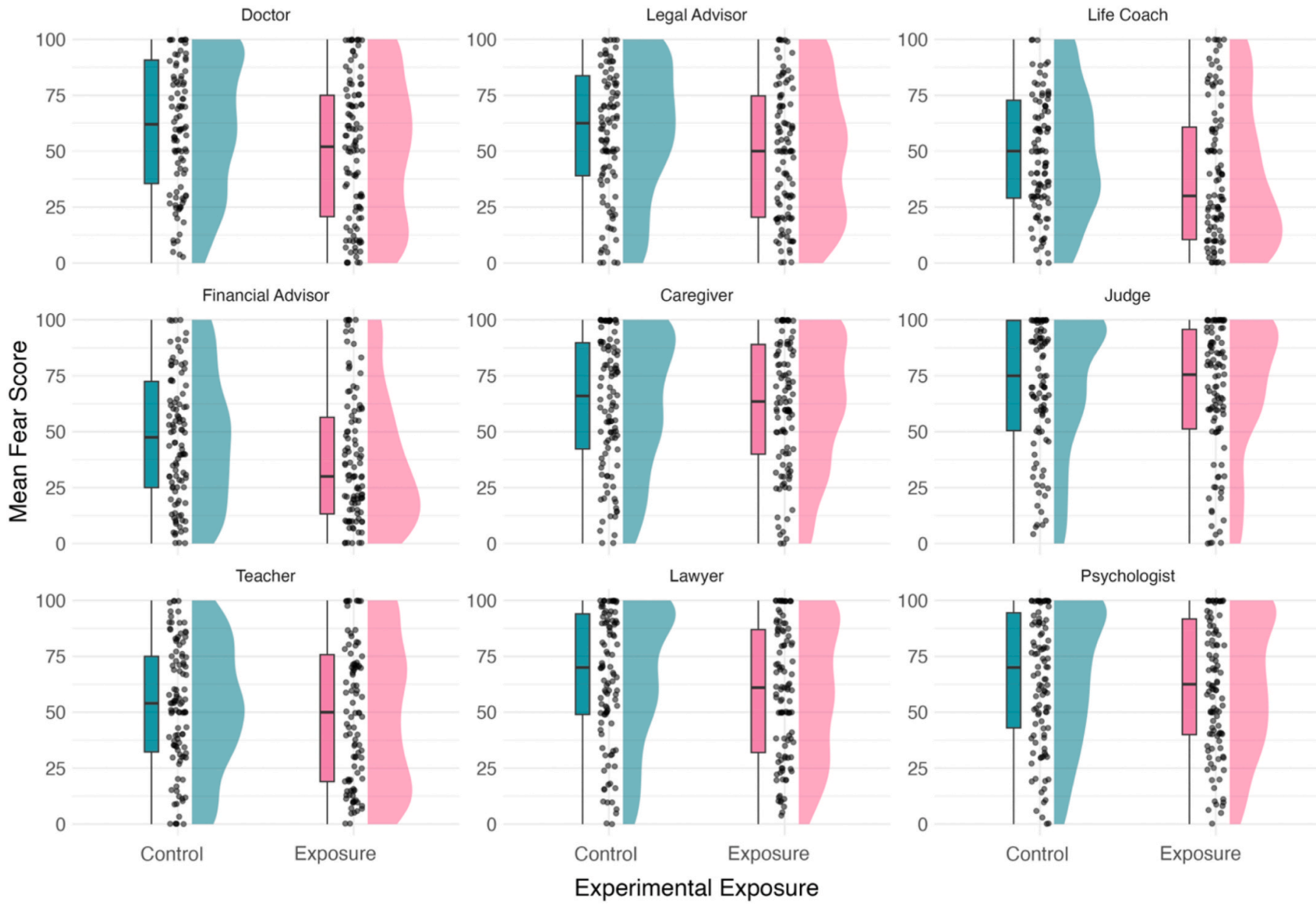

Fig. 3. Mean Fear Scores for Different Occupational Roles by Experimental Exposure
Note. This figure displays participants' reported fear scores for various AI-driven occupational roles across two conditions: Control: No Interaction (Survey Only) and Exposure: Interaction with Chatbot. Mean fear scores are shown for nine roles: Doctor, Legal Advisor, Life Coach, Financial Advisor, Caregiver, Judge, Teacher, Lawyer, and Psychologist. Error bars represent the interquartile range, and data points are jittered for visibility. On a descriptive level, participants in the Exposure condition reported lower fear scores across most roles, but there was no credible evidence for statistical significance.

revealed themes centered on interests, preferences, and daily activities, particularly around favorite pastimes and entertainment. The frequent occurrence of terms like "favorite" and "love" underscores the emphasis on sharing likes, opinions, and experiences. Overall, the tone of the interactions suggests light, informal engagement typical of small talk.

The dataset of the conversations leans slightly toward positive sentiment—with an average sentiment score of 0.179, measured with the VADER Sentiment Analyzer from the NLTK library. To gain a more detailed understanding of participant emotions, a multi-class sentiment analysis was conducted using the Hugging Face model j-hartmann/emotion-english-distilroberta-base. This analysis revealed that neutral responses were the most frequent (51.6 %). Joy was the second most frequent emotion (26.5 %), indicating that a subset of messages expressed positive sentiment. These findings provide a nuanced view of the conversational dynamics, with participants primarily engaging in neutral or mildly positive exchanges. However, there are no significant correlations between the sentiments from the conversations with the AI chatbot and our key outcome variables. Detailed analyses and additional statistical outputs are provided in the Supplementary Material (Material 5).

4.6. Robustness check: Fatigue effects

As expected, participants in the exposure group (which included the 7-min chatbot interaction) spent more time in the study ($M = 872.59$ s) than those in the control group ($M = 548.70$ s), Welch's t (202.69) $= -9.15$, $p < .001$, 95 % CI [–393.66, −254.13]. The effect was large, Cliff's $\delta = -0.67$, 95 % CI [–0.77, −0.54].

Next, as a robustness check, we tested whether task duration was correlated with the dependent variables. Most associations were small and non-significant. Notably, longer durations were positively correlated with greater enjoyment of the agent, r (218) $= 0.26$, $p < .001$, 95 % CI [0.13, 0.38], willingness to interact again, r (218) $= 0.30$, $p < .001$, 95 % CI [0.17, 0.41], perceptions that the agent thinks like a human, r (218) $= 0.19$, $p = .005$, 95 % CI [0.06, 0.31], and trust in the agent, r (218) $= 0.14$, $p = .037$, 95 % CI [0.01, 0.27]. Further main analyses controlling for time are provided in the Supplementary Material (Material 6).

5. Discussion

Here we adopt a multidimensional perspective on public sentiments toward AI at a time of increasing integration of AI chatbots, such as ChatGPT, into everyday life. Taking advantage of a time window when people were experiencing first contact with generative AI chatbots, we show the nuanced causal influence of such exposure on different dimensions of sentiment.

Specifically, direct interaction with the AI chatbot reduced negative sentiments and increased positive sentiments toward the system itself, regardless of participants' prior experience. These effects, however, did not consistently generalize to other AI-related contexts. Integrating insights from *Intergroup Contact Theory* (ICT; Pettigrew & Tropp, 2006), the reduction in negative responses is consistent with prior literature, suggesting that even brief contact with an outgroup member — here, an AI chatbot — can lower negative sentiments towards this group by reducing uncertainty and increasing familiarity (Pettigrew & Tropp, 2006).

In our case, the observed reduction in negative responses may reflect a shift from speculative concerns to more grounded impressions based on firsthand experience with the chatbot's actual capabilities and limitations. In this sense, interaction appeared to foster a more realistic view of the chatbot, without substantially altering participants' underlying beliefs about its human-likeness.

The largely casual, small-talk nature of our interactions may have been sufficient to influence system-specific fears but insufficient to trigger broader spillover to more abstract or less directly related AI contexts. Notably, exposure led to reduced fear of AI in medical applications, but similar effects were not observed in other domains such as recruitment or product recommendation. Likewise, interaction had minimal influence on people's hope regarding AI's broader societal benefits, including those in personal or governance contexts (e.g., companionship, anti-corruption).

Our findings complement emerging large-scale analyses of public responses to ChatGPT. Early social media sentiment analyses revealed a predominantly positive, yet ambivalent sentiment profile, with users expressing curiosity and appreciation alongside concern and fear (Korkmaz et al., 2023; Zhou et al., 2024). Similarly, Alam and Escobari (2024) observed a subsequent rise in negative sentiment in public discourse following ChatGPT's release. Adding on these general patterns, Lu and colleagues (2025) show domain-specific patterns, with optimism about efficiency coexisting with persistent concerns about accuracy and professional displacement in medical contexts. In contrast, our experimental evidence shows that direct, personal interaction tends to reduce fear and foster engagement. Together, these results suggest that while public narratives about AI may amplify mixed sentiments, individual experience can temper such concerns through direct familiarity and reduced uncertainty.

Taken together, these findings provide preliminary evidence that exposure to AI chatbots may influence sentiment in a domain-specific manner, with stronger effects on engagement-related perceptions (e.g., enjoyment, human-likeness) than on competence-related judgments (e.g., trust, factual accuracy). This aligns with prior work which suggests that interactive experiences can increase social attributions to technology (Epley et al., 2007) and with recent reviews suggesting that human trust in AI is shaped not only by general attitudes but also by contextual features such as the task domain and the perceived representation of the AI (Glikson & Woolley, 2020). From a design perspective, our results suggest that brief, unstructured interactions (like the short exposure implemented here) can effectively enhance engagement and perceptions of social presence but are less likely to shift deeper competence-related attitudes such as trust or perceived accuracy. This implies that chatbot applications aimed at promoting user connection or comfort (e.g., educational or wellbeing contexts) may benefit from even minimal exposure, whereas applications requiring confidence in the system's expertise may need more sustained interaction, transparency, or task-specific scaffolding to build trust.

This pattern aligns with prior literature showing that sentiments like trust in automation is highly context-dependent, varying according to situational cues, task characteristics, and prior experience (Hoff & Bashir, 2015). While some evidence suggests that experiences with one AI can influence perceptions of others, such effects are not uniform. For example, Manoli and colleagues (2025) found that immoral actions by an AI agent could spill over to attitudes toward other AIs—reducing perceived moral agency and trustworthiness—but primarily when AIs were presented as part of a homogeneous category, and not when they were individuated. These findings reinforce that broader sentiment change may depend on how the AI is framed, perceived group membership, and the interaction context—helping explain why our short AI chatbot exposure produced domain-specific rather than widespread effects.

It is also worth noting that participants approached the chatbot with a naturally positive tone, and the conversations largely revolved around everyday topics such as personal interests (e.g., movies, food, hobbies), as revealed by our qualitative analysis. Future research could examine whether directing participants toward more cognitively or ethically charged discussions (such as AI's societal roles or decision-making capabilities) might produce stronger or different shifts in sentiment. Nevertheless, even without such targeted instructions, we found that a brief, unstructured interaction was sufficient to change certain sentiment dimensions. While this approach allowed us to capture spontaneous, low-stakes engagement, it does not necessarily represent the diversity of real-world chatbot use cases, which can vary greatly in purpose, duration, and emotional intensity.

Participants were recruited via Prolific, a platform that yields higher-quality data than most crowdsourcing sites but nonetheless over-represents younger, educated, and digitally literate populations, limiting generalizability beyond Western online samples. Although the present findings are based on GPT-3.5, the underlying mechanisms—uncertainty reduction through concrete interaction and familiarity effects through repeated exposure—reflect processes established in intergroup contact research (Pettigrew, 2008; Pettigrew & Tropp, 2006) and human–robot interaction (Haggadone et al., 2021). These dynamics are therefore likely to apply across other chatbots and future model iterations, though the magnitude and direction of effects may vary with system design and capability. Importantly, effects should not be assumed to unfold monotonically: greater exposure may foster normalization, but it may also provoke backlash if systems are perceived as overly capable or intrusive.

It should be noted that the contrast with the exposure condition centers on a low-fidelity scenario involving imagined interaction, which requires less time and elicits lower levels of engagement. This design isolates the causal effect of abstract expectation (versus direct interaction), reflecting the reality around the time of study that many participants had limited actual experience with AI chatbots and relied instead on imagined encounters. Depending on the research questions of interest, further studies may compare actual interaction with AI versus rule-based chatbots, or personal versus observed interaction with AI chatbots, each offering additional insights into the implications of exposure to AI chatbots."

Our findings offer insights into the complexity and context sensitivity of public sentiment toward AI. Brief exposure to an AI chatbot can reduce immediate fears about the system itself and, in some cases, extend to closely related domains. This suggests a practical pathway for reducing resistance to potentially beneficial technologies: even a short, interaction may be enough to begin shifting attitudes among individuals skeptical of AI, at least toward the specific system encountered. However, these effects do not reliably spill over into abstract or less tangible aspects of AI perception. This underscores the importance of domain-specific communication strategies and interaction design in shaping public understanding. Future research should explore how emotional engagement, personalized feedback, and repeated or varied forms of exposure may influence broader attitudes. Such efforts could inform strategies for both reducing unfounded fears and promoting critical, informed engagement with AI technologies.

6. Conclusion

This study highlights that brief interactions with an AI chatbot can meaningfully reduce system-specific fears and selectively extend to certain real-world domains. However, the limited spillover effects in more abstract contexts suggest that exposure alone may be insufficient to influence broader attitudes. Our findings reinforce the understanding that sentiments about AI are multidimensional, influenced by both the context of the technology and individual experiences. As AI systems increasingly integrate into daily life, fostering informed, realistic, and context-sensitive public perceptions will be crucial for promoting trust, encouraging responsible adoption, and enabling critical engagement with AI technologies.

CRediT authorship contribution statement

Eva-Madeleine Schmidt: Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Conceptualization. **Clara Bersch:** Software, Formal analysis, Data curation, Conceptualization. **Nils Köbis:** Writing – review & editing, Conceptualization. **Jean-François Bonnefon:** Supervision, Conceptualization. **Iyad Rahwan:** Supervision, Conceptualization. **Mengchen Dong:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

We have no known conflict of interest to disclose.

Acknowledgements

The authors acknowledge funding from grant ANR-19-P13A-0004, grant ANR-17-EURE-0010, and the research foundation TSE-Partnership. The funding sources had no other role other than financial support. The authors declare no other conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbah.2025.100223.

References

Alam, M., & Escobari, D. (2024). Projecting the sentiment shift towards AI: A Difference-in- difference investigation of individual's emotions in response to the emergence of ChatGPT. *AMCIS 2024 Proceedings, 1*.

Cave, S., & Dihal, K. (2019). Hopes and fears for intelligent machines in fiction and reality. *Nature Machine Intelligence, 1*(2), 74–78. https://doi.org/10.1038/s42256-019-0020-9

De Freitas, J., Agarwal, S., Schmitt, B., & Haslam, N. (2023). Psychological factors underlying attitudes toward AI tools. *Nature Human Behaviour, 7*(11), 1845–1854. https://doi.org/10.1038/s41562-023-01734-2

Dong, M., Bonnefon, J.-F., & Rahwan, I. (2024). Toward human-centered AI management: Methodological challenges and future directions. *Technovation, 131*, Article 102953. https://doi.org/10.1016/j.technovation.2024.102953

Dong, M., Conway, J., Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2023). A psychological model predicts fears about artificial intelligence across 20 countries and 6 domains of application. https://doi.org/10.31234/osf.io/pjvqt.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A. M. A., & Hatta, H. R. (2023). Analysing the role of ChatGPT in improving student productivity in higher education. *Journal of Education, 5*(4), Article 4. https://doi.org/10.31004/joe.v5i4.2563

Ge, X., Xu, C., Misaki, D., Markus, H. R., & Tsai, J. L. (2024). How culture shapes what people want from AI. https://doi.org/10.1145/3613904.3642660.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *The Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Haggadone, B. A., Banks, J., & Koban, K. (2021). Of robots and robotkind: Extending intergroup contact theory to social machines. *Communication Research Reports, 38*(3), 161–171. https://doi.org/10.1080/08824096.2021.1909551

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Kieslich, K., Lünich, M., & Marcinkowski, F. (2021). The threats of artificial intelligence scale (TAI). *International Journal of Social Robotics, 13*(7), 1563–1577. https://doi.org/10.1007/s12369-020-00734-w

Korkmaz, A., Aktürk, C., & Talan, T. (2023). Analyzing the user's sentiments of ChatGPT using Twitter data. *Iraqi Journal for Computer Science and Mathematics*, 202–214. https://doi.org/10.52866/ijcsm.2023.02.02.018

Liang, Y., & Lee, S. A. (2017). Fear of autonomous robots and artificial intelligence: Evidence from national representative data with probability sampling. *International Journal of Social Robotics, 9*(3), 379–384. https://doi.org/10.1007/s12369-017-0401-3

Liao, S. M. (2020). *Ethics of artificial intelligence*. Oxford University Press.

Longoni, C., Cian, L., & Kyung, E. J. (2023). Algorithmic transference: People overgeneralize failures of AI in the government. *Journal of Marketing Research, 60*(1), 170–188. https://doi.org/10.1177/00222437221110139

Lu, L., Zhu, Y., Yang, J., Yang, Y., Ye, J., Ai, S., & Zhou, Q. (2025). Healthcare professionals and the public sentiment analysis of ChatGPT in clinical practice. *Scientific Reports, 15*(1), 1223. https://doi.org/10.1038/s41598-024-84512-y

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments, 11*(1), 9. https://doi.org/10.1186/s40561-024-00295-9

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change, 175*, Article 121390. https://doi.org/10.1016/j.techfore.2021.121390

Manoli, A., Pauketat, J. V. T., & Anthis, J. R. (2025). The AI double standard: Humans judge all AIs for the actions of one. *Proc. ACM Hum.-Comput. Interact., 9*(2), CSCW185:1–CSCW185:24. https://doi.org/10.1145/3711083

Molina, M. D., & Sundar, S. S. (2022). When AI moderates online content: Effects of human collaboration and interactive transparency on user trust. *Journal of Computer-Mediated Communication, 27*(4), Article zmac010. https://doi.org/10.1093/jcmc/zmac010

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods, 54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Pettigrew, T. F. (2008). Future directions for intergroup contact theory and research. *International Journal of Intercultural Relations, 32*(3), 187–199. https://doi.org/10.1016/j.ijintrel.2007.12.002

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A.'Sandy', … Wellman, M. (2019). Machine behaviour. *Nature, 568*(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y

Roth, E. (2024). ChatGPT's weekly users have doubled in less than a year. *Verge*. https://www.theverge.com/2024/8/29/24231685/openai-chatgpt-200-million-weekly-users.

Shinners, L., Aggar, C., Stephens, A., & Grace, S. (2023). Healthcare professionals' experiences and perceptions of artificial intelligence in regional and rural health districts in Australia. *Australian Journal of Rural Health, 31*(6), 1203–1213. https://doi.org/10.1111/ajr.13045

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1260843

Starke, C., Ventura, A., Bersch, C., Cha, M., De Vreese, C., Doebler, P., Dong, M., Krämer, N., Leib, M., Peter, J., Schäfer, L., Soraperra, I., Szczuka, J., Tuchtfeld, E., Wald, R., & Köbis, N. (2024). Risks and protective measures for synthetic relationships. *Nature Human Behaviour, 8*(10), 1834–1836. https://doi.org/10.1038/s41562-024-02005-4

Yakura, H., Lopez-Lopez, E., Brinkmann, L., Serna, I., Gupta, P., & Rahwan, I. (2024). Empirical evidence of large language Model's influence on human spoken communication (No. arXiv:2409.01754). *arXiv*. https://doi.org/10.48550/arXiv.2409.01754

Yam, K. C., Tan, T., Jackson, J. C., Shariff, A., & Gray, K. (2023). Cultural differences in people's reactions and applications of robots, algorithms, and artificial intelligence. *Management and Organization Review, 19*(5), 859–875. https://doi.org/10.1017/mor.2023.21

Zhou, J., Liang, Z., Fang, Y., & Zhou, Z. (2024). Exploring public response to ChatGPT with sentiment analysis and knowledge mapping. *IEEE Access, 12*, 50504–50516. https://doi.org/10.1109/ACCESS.2024.3386362