Communication and Change

RESEARCH Open Access

Shifting dynamics in human–machine communication: the interplay of user and chatbot gender across two studies

Weizi Liu[1*] and Kun Xu[2]

*Correspondence:
weizi.liu@tcu.edu

[1] Bob Schieffer College of Communication, Texas Christian University, Fort Worth, United States
[2] College of Journalism and Communications, University of Florida, Gainesville, United States

Abstract

As conversational agents become a core interface in service, education, and decision support, understanding how users' self-concept shapes their interaction with gendered AI agents is critical. This study investigates how chatbots' presented gender, together with users' gender and gender traits, influence users' perceptions of chatbot credibility and social attraction. Across two experimental studies ($N_1 = 250$, $N_2 = 228$), participants engaged with a chatbot presented as female, male, or gender-neutral. Study 1 employed a rule-based chatbot, whereas Study 2 used a generative AI chatbot for a more naturalistic interaction. In Study 1, results revealed that users' gendered self-concept modulates their perceptions of the chatbot, particularly when chatbot gender aligns or misaligns with participants' self-perceived masculinity or femininity. The more masculine users were, the more they perceived the male chatbot as competent; by contrast, users with higher femininity evaluated the neutral chatbot more favorably in goodwill, trustworthiness, and social attraction. However, these effects were not observed in Study 2, suggesting that advanced AI performance may dilute social identity-based biases. We further discuss how interaction effects reflect broader sociocultural schemas embedded in digital communication and highlight ethical implications for AI design, including risks of reinforcing gender stereotypes and marginalizing non-binary identities. Findings underscore the need for inclusive, identity-aware chatbot design that balances personalization with mindfulness.

Keywords: Gender, Chatbot, Large language model, Human–machine communication, Human-AI interaction

Introduction

As conversational AI becomes increasingly embedded in everyday life, from customer service to education to creative assistance, designers have leaned on gendered personas to make these agents more relatable. Chatbots and virtual assistants are often given names, voices, and interaction styles that map onto culturally familiar gendered roles, with female-coded agents dominating tasks associated with help, support, and care (e.g., customer service, sales, education, health care; Curry et al., 2020; Feine et al., 2020). While these female-oriented design choices may lead to perceived friendliness and offer positive user experience (Eyssel & Hegel, 2012; Goodman & Mayhorn, 2023), they raise

Springer

critical ethical concerns regarding what gender norms are being encoded into AI and how users respond when these systems reflect, or conflict with their own gender traits.

Prior research has shown that users apply social expectations to AI agents, particularly when those agents exhibit identity-related cues such as names or voices. Most studies have focused on gender identity as a design feature of machines (e.g., male vs. female agents; Lee et al., 2000; Nass & Brave, 2005; Nass & Moon, 2000). More recent work demonstrates that identity-related cues, particularly gender cues, continue to shape users' evaluations of AI agents in increasingly nuanced and context-dependent ways, influencing perceptions of warmth, competence, trust, and appropriateness across interaction settings (e.g., Duan et al., 2025a, b; Fossa & Sucameli, 2022; Piercy et al., 2025). However, comparatively less attention has been paid to the impact of gender differences on the user side, in particular to the conceptual distinction between users' gender identities and their gendered traits.

Gender identity is defined as a person's internal sense of being male, female, a blend of both, or neither (American Psychological Association, 2015), emphasizing the self-definitional and categorical aspects of gender. In contrast, gender traits refer to psychological characteristics that are socially and culturally coded as masculine or feminine, such as assertiveness, nurturance, and emotional expressiveness, and can vary across individuals regardless of their gender identities (Bem et al., 1976; Spence & Helmreich, 1978). Although related, these constructs are not equivalent: individuals who share the same gender identity may differ substantially in their gendered traits. Importantly, gendered traits can shape how people interpret, evaluate, and respond to social actors—including artificial agents–by influencing perceptions of competence and relational orientation (Benbasat et al., 2020; Eyssel & Hegel, 2012; Fiske et al., 2007).

Social identity theory (Tajfel & Turner, 1979) proposes that individuals tend to evaluate others more positively when they share salient in-group characteristics, including the same gender cues. Relatedly, gender schema theory posits that individuals develop internalized cognitive structures about gender that guide how they perceive and evaluate others, shaping expectations for appropriate characteristics and behavior (Bem, 1981). Consistent with these perspectives, prior studies show that gender identity congruity influence people's perceptions of competence, warmth, trust, and social attraction toward both human and AI agents (Ahn et al., 2022; Benbasat et al., 2020; Eyssel & Hegel, 2012; Jeon, 2024; Liu & Yao, 2023; Zhang et al., 2025). Moreover, existing research demonstrates that individuals' masculinity and femininity shape their psychological, physiological, and behavioral responses during communication processes (Hirokawa et al., 2004). Conceptualized as individual differences in agency and communion, these traits (i.e., masculinity, femininity) shape how social cues are interpreted and evaluated, with femininity emphasizing relational orientation and masculinity emphasizing instrumental competence (Bem, 1981; Fiske et al., 2007). These findings suggest that users' evaluations of chatbots may depend not only on categorical gender identity alignment but also on users' gendered traits.

Recently, the emergence of generative AI and large language models (LLMs) has introduced transformative shifts in human-AI interaction. Unlike earlier rule-based systems, LLM-powered agents enable more flexible, dynamic, and humanlike conversations, offering context-aware responses that mimic natural language use (Floridi & Chiriatti, 2020).

This marks a departure from the static and scripted interactions typical of earlier chatbots. Despite these advances, limited empirical research has directly examined the communication dynamics between humans and conversational agents built on LLMs, leaving open questions about relational cues, perceived agency, and co-construction of meanings (Guzman, 2018; Guzman & Lewis, 2020).

In this study, we explore the impacts of AI agents with assigned gender identities while factoring in participants' gender identities and traits. We focus on credibility and social attraction because they represent two core and complementary communication outcomes through which gender cues shape users' evaluations of social actors, including AI agents (e.g., Lew & Walther, 2023; Liu & Yao, 2023). Credibility captures cognitive judgments of competence and trustworthiness that are closely tied to gendered expectations of expertise, whereas social attraction reflects affective and relational responses associated with warmth and interpersonal appeal. We examine how these outcomes vary as a function of alignment or misalignment among users' gender identities, gendered traits, and chatbots' presented gender. Through two studies conducted before and after the widespread adoption of generative AI, we capture nuanced gender dynamics in HMC beyond agent design alone.

Across two experiments, participants interacted with chatbots presented as male, female, or gender-neutral, ranging from rule-based to generative AI systems. The results provide evidence that chatbots' presented gender interacts with users' gender traits (i.e., femininity and masculinity) on shaping perceptions of chatbot credibility and social attraction, particularly during constrained interactions, whereas the enhanced capabilities of generative AI may partially buffer or override these identity-driven biases. These findings carry important ethical implications: users appear to transfer interpersonal gender stereotypes to interactions with rule-based gendered chatbots, especially in structured contexts, reinforcing concerns about the unintended consequences of gendered AI design. By highlighting users' self-evaluated gender traits, this study contributes to communication theories and the ethical discussion of moving beyond demographic cues to individual nuances, providing implications of AI agent design.

Literature review

Gendered chatbots and ethical concerns

Conversational agents (CAs), such as voice assistants and chatbots, are often designed with social characteristics through names, voices, or appearances. Assigning gender identities is one of the most common design practices. For example, Amazon's Alexa, Apple's Siri, Microsoft's Cortana are voice agents with default female voices; Replika is an AI companion chatbot that is virtually embodied in a female figure in its initial form. While these design choices may enhance user experience, they also raise important questions about how users respond to gendered agents and the ethical implications of reinforcing gender stereotypes or norms. This study focuses on text-based chatbots because they are a prevalent and widely used form of chatbots and, by relying on linguistic cues, they invite users to actively evaluate chatbots' identities, heightening expectations around assigned gender identities while maintaining simplicity from other design factors.

Users readily attribute gender identities to CAs based on even minimal cues. Research in human–computer interaction (HCI) and human–machine communication (HMC) shows that a voice assistant's vocal pitch, name, or visual appearance can trigger users' gender attribution. For example, female-voiced agents are seen as warmer and more helpful, while male-voiced agents are perceived as more competent or authoritative (Nass & Brave, 2005; Nass et al., 1997). Fossa and Sucameli (2022) also found that an agent with male-associated cues is rated higher on attributes like assertiveness or competence, whereas a female bot is rated to be friendly, caring, and service-oriented. Moreover, stereotypically male-oriented tasks (e.g., math tasks) were judged as more suitable for a male robot, whereas stereotypically female-oriented tasks (e.g., verbal tasks) were seen as more appropriate for a female robot (Eyssel & Hegel, 2012). Such gender attribution aligns with the Computers Are Social Actors (CASA) paradigm, which suggests that people apply human social rules to machines that display social cues (Nass & Moon, 2000), offering useful lenses for interpreting how users interact with gendered agents.

Extending this line of work to contemporary systems, recent studies have shown that gender attribution persists even in advanced and ostensibly gender-neutral generative AI. For example, Duan et al. (2025b) demonstrates that removing explicit gender identities from generative AI can mitigate stereotypes, and linguistic cues can still trigger gendered attributions with non-gendered AI. This suggests that users can attribute gender identities to AI, which is shaped by both explicit and implicit agent attributes. Piercy et al. (2025) suggest that voice gender alone does not consistently shape trust in virtual assistants; instead, trust depends strongly on interaction context, such as task type and sociolinguistic cues like accent. Their findings highlight that gender effects in human–AI communication are nuanced and interactional, rather than being driven solely by gender stereotypes. This recent work suggests that gendered evaluations of AI remain salient despite advances in conversational fluency and these evaluations become more nuanced as contexts vary.

Gendered CAs serve as a mirror for human social norms and a potential lever for cultural change. Ethical concerns arise when gendered CAs reinforce stereotypes. UNESCO's report (West et al., 2019) highlighted that defaulting to female personas in subservient roles (e.g., digital assistants) could normalize sexist expectations, especially when machine agents do not challenge abusive or demeaning behavior. Studies have shown that female-like agents are more likely to be subjected to sexual or aggressive remarks from users, raising alarms about how gendering technology might reflect or even exacerbate societal biases (Brahnam & De Angeli, 2012). Fossa and Sucameli (2022) frame the issue starkly: "*Is it ethically permissible to align the design of [conversational agents] to gender biases in order to improve interactions and maximize user satisfaction?*"(p.2) As suggested by design research, aligning an agent's persona (e.g., voice, embodiment, and demographics) with user expectations can improve its usability and satisfaction (Zargham et al., 2024). This utilitarian "bias alignment" strategy suggests that designers should use cues, even gender stereotypes, that enhance user comfort and engagement, treating gendered design as a neutral reflection of user bias. However, many ethicists and psychologists warn that this approach risks legitimizing and reinforcing harmful stereotypes. Fossa and Sucameli (2022) proposed the "feedback effect", meaning that biases activated in human–AI interaction might carry back

into human–human interaction. In other words, allowing stereotypically gendered CAs could strengthen gender biases, creating a vicious cycle or "lock-in effect" where prejudice becomes further normalized. Importantly, emerging research suggests that these ethical concerns extend beyond explicitly gendered designs. Even when gender cues are minimized, LLMs may reproduce gender stereotypes through language patterns and role associations embedded in training data (Duan et al., 2025b; Kotek et al., 2023). This raises concerns that generative AI may subtly perpetuate bias even under ostensibly neutral design choices.

Beyond specific gender cues, users' responses to chatbots depend on how they ontologically classify these systems—that is, what kind of entity they believe they are interacting with. Research in HMC shows that users do not treat all machines as ontologically equivalent, but instead engage them as tools, social actors, or hybrid entities that blur human–machine boundaries (Edwards & Etzrodt, 2025; Guzman, 2020; van der Goot & Etzrodt, 2023). Meanwhile, gender cues may contribute to this process without fully determining it. For example, Fortunati et al. (2022) show that while most users explicitly label Alexa as female, many avoid gendered pronouns in spontaneous language and variably attribute communicative status to the system, indicating persistent ontological ambivalence.

These findings suggest that gendered design features shape users' interpretations of AI without producing uniform ontological placement. Such classifications are shaped by surface features (e.g., names, voices, interaction style) and broader cultural narratives about AI. van der Goot and Etzrodt (2023) suggested that users may alternately respond to machines as objects that evoke reflection on their ontological status, while Edwards and Etzrodt (2025) further conceptualized HMC as *trans-ontological*, emphasizing that attributions of agency and communicative legitimacy vary across contexts and design features. These perspectives provide a theoretical foundation for our experimental manipulation, which assumes that gender cues function in part by shaping users' ontological assumptions about chatbots and, in turn, their evaluative and relational responses.

The impacts of user gender

Communication between human users and CAs is inherently relational. Yet, much research in HMC has focused on chatbot characteristics (e.g., assigned gender identity). Relatively less attention has been paid to the joint effects of user characteristics, such as gender, personality, or values, and chatbots' assigned identity cues. To address this gap, we organize our framework around a central mechanism: the interplay between user and chatbot identity cues, through which users interpret, evaluate, and respond to conversational agents. Specifically, we argue that users rely on gender-related cues as heuristic signals when evaluating CAs, and that these cues are interpreted through users' own gender identities and gender traits. From an HMC perspective, users often respond to CAs as social actors, applying human social categories and expectations to them (e.g., CASA). This tendency makes gender cues in chatbots psychologically meaningful, even when users are aware that they are interacting with nonhuman systems.

Social identity theory (Tajfel & Turner, 1979) offers a foundational lens for understanding these dynamics. It posits that individuals derive part of their self-concept from

group memberships and show preference for those who reflect their in-group characteristics. In the context of HMC, this translates into a tendency to trust and engage more with agents perceived as similar to oneself. In line with this, recent studies have found that identity congruence between users and agents, whether in genders, personalities, or values, can enhance trust, warmth, and cooperation (Jin & Eastin, 2024; Mehrotra et al., 2021; Reinkemeier & Gnewuch, 2022). For example, users reported greater trust in voice assistants that matched their own personality traits (Reinkemeier & Gnewuch, 2022). Also, agents whose value systems were aligned with those of users were perceived as more trustworthy (Mehrotra et al., 2021).

Gender schema theory proposes that individuals develop internalized cognitive schemas about masculinity and femininity that guide how they interpret and respond to gendered behavior in themselves and others (Bem, 1981). These schemas, which are often measured as masculine, feminine, or androgynous orientations using instruments such as the Bem Sex-Role Inventory (BSRI; Bem, 1974), shape expectations, evaluations, and interaction styles. Self-perceived gender traits offer a more nuanced understanding of identity and behavior (Bem, 1974; Spence & Helmreich, 1978). Individuals higher in femininity tend to emphasize relational orientation, warmth, and expressiveness, whereas those higher in masculinity place greater weight on instrumentality, competence, and task effectiveness in communication and evaluation contexts (Fiske et al., 2007). These findings demonstrate that both biological sex (e.g., females, males) and gender-related traits (e.g., masculinity, femininity) independently and interactively shape communication behaviors. For example, Hirokawa et al. (2004) found that variations in masculinity–femininity moderated individuals' physiological and behavioral responses during interpersonal communication tasks, suggesting a possible interaction between categorical gender identities and self-reported gender traits in evaluation of communication.

The above perspectives point to a unified framework in which users rely on identity-based cues when evaluating communication partners, interpreted through both their own categorical gender identities and trait-based schemas. While prior research has largely developed these insights in interpersonal contexts, there is still much space in HMC that remains under-explored. In interactions with CAs, gender cues are embedded in designed features rather than inherent human characteristics, raising questions about how these identity-based processes are interpreted through users' gender lens in technology-mediated environments. Emerging work suggests that users' identity-related preferences interact with agent design features to shape perceptions of CAs, including credibility and relational affinity (Adinolf et al., 2025; Jin & Eastin, 2024). For example, in healthcare applications, chatbot–user gender matching increases perceived expertise and intention to use (Jin & Eastin, 2024). However, more research is needed to systematically examine how users' categorical gender identities and gender-related traits jointly influence users' responses to chatbot gender cues across different communication contexts.

Current investigation

Based on these theoretical foundations, in this paper, we not only focus on the impacts of the categorical view of individuals' genders but also consider trait-based dimensions. Social identity theory sets our overarching mechanism of interest—how the alignment

of users' and chatbots' categorical genders impact communication outcomes. Furthermore, while most HMC research treats gender as a fixed demographic variable (i.e., male or female), this categorical approach overlooks how individuals internalize and express gender through traits such as assertiveness, nurturance, or emotional expressiveness. Therefore, according to gender schema theory, in AI-mediated contexts, we have reason to argue that it is not simply a user's categorical gender that shapes trust or empathy, but more importantly, how much their own gendered self-characteristics align with the agent's projected persona. By incorporating self-rated masculinity and femininity, the current study provides a more nuanced, identity-informed understanding of user–agent interactions. These findings also inform the ongoing efforts in communication and HCI to design AI systems that are more inclusive, responsive, and ethically aligned with diverse user identities.

With a focus on both evaluative judgments and relational connections with chatbots, we identify perceived credibility and social attraction as our key dependent variables. These constructs are theoretically grounded and widely used in the evaluation of both interpersonal and human–machine communication. Credibility, which comprises competence, goodwill, and trustworthiness, has long been a core factor of how messages and communicators are evaluated (McCroskey & Teven, 1999). Although closely related as dimensions of overall credibility, each construct captures a distinct evaluative focus: competence reflects perceptions of ability and expertise, trustworthiness concerns honesty and reliability, and goodwill reflects perceived caring, benevolence, and concern for the receiver. These dimensions provide a comprehensive assessment of how communicators, including AI agents, are evaluated and have been successfully adapted in HMC research to assess users' judgments of AI systems, including various application contexts (e.g., Lew & Walther, 2023). Meanwhile, social attraction is a central construct in interpersonal communication research (McCroskey & McCain, 1974) and has been used in HMC studies as a proxy for relational affinity (Lew & Walther, 2023). These outcome variables are not only conceptually robust but also empirically predictive of user satisfaction, trust, and continued engagement across a wide range of AI-mediated settings.

Importantly, credibility and social attraction are expected to emerge from the joint influence of chatbot gender cues, users' categorical gender identities, and users' self-perceived gender traits. Chatbot gender may activate gendered expectations, while alignment or misalignment with users' gender identities shapes baseline evaluations, and users' gender traits (e.g., masculinity and femininity) condition how strongly competence-, goodwill-, or trustworthiness-related judgments are emphasized. Inspired by social identity theory, we raise an exploratory research question to explore the joint impacts of both users' and chatbots' categorical gender identities on perceived credibility and attractiveness of the chatbots. Integrating gender schema theory, we also conceptualize masculinity and femininity as key variables that shape how users interpret gender-related cues during HMC. These trait-based dimensions are expected to systematically influence evaluations of CAs.

RQ1: How do a chatbot's assigned gender identities, users' gender identities, and users' gender traits influence the perceived credibility and social attraction of the chatbot?

The rise of generative AI, particularly chatbots like ChatGPT, has brought substantial changes to HMC. These systems are capable of generating emotionally nuanced,

context-sensitive dialogue, which can lead users to perceive them as not only tools but also socially competent partners (e.g., Gessinger et al., 2025; Liu, 2024). The technical breakthrough of generative AI in enabling humanlike conversation allows AI agents to engage users more naturally. Powered by LLMs, these agents are not only designed with richer social identities, such as specific names, genders, or personalities, but also demonstrate greater conversational fluency, which can enhance users' tendency to apply human social norms and responses toward the chatbots (Hancock et al., 2020; Nass & Moon, 2000).

At the same time, the capacity of LLMs to simulate social cues and human communication behavior raises important questions about how such cues are interpreted. Prior work suggests that LLMs can reflect patterns in their training data, including societal stereotypes (Bender et al., 2021; Kotek et al., 2023). However, more recent research has focused on model-level biases or algorithmic performances, with less attention paid to how these systems shape user perceptions and communication dynamics in HMC contexts.

From a communication perspective, the increased conversational richness and interactivity of LLM-based agents may alter how users rely on identity-related cues (e.g., gender) when forming evaluations. On one hand, more humanlike and socially rich interactions may make such cues more salient and socially meaningful. The realism and relationality of generative AI, while enabling more natural interaction, also pose ethical challenges regarding the perpetuation of inequality and normative assumptions through AI-mediated communication. On the other hand, richer conversational content may shift attention toward message-level information, potentially reducing or complicating the influence of these cues. In this sense, the effects of identity-based cue processing may not follow a single directional pattern but instead vary depending on the interaction environment. As conversations with generative AI become more individualized and adaptive (Du et al., 2024), it becomes increasingly difficult to draw broad generalizations. This raises new questions about whether the effects of social identity design, particularly gender, become more pronounced, consistent, or potentially diminished in the context of LLMs. Therefore, our study investigates RQ1 in two studies. We examine whether the relationships among chatbot gender, user gender characteristics, and user perceptions vary across a traditional rule-based chatbot and an LLM-based chatbot, addressing the following exploratory research question:

RQ2: How do rule-based and LLM chatbots differ in their perceived credibility and social attraction when influenced by their assigned gender and users' gender and gender traits?

Study 1: A rule-based chatbot

Method

Ethics approval and consent to participate

Study 1, along with the subsequent Study 2, received the institutional research board (IRB) approval from the University of Illinois at Urbana-Champaign in December 2021 (project #20659).

Design and procedure

A 2 × 3 (participant gender: female vs. male; chatbot gender: female vs. male vs. neutral) online experiment was conducted using the Juji.ai chatbot platform (see Supplementary Material). Juji is a platform that lets organizations build, deploy, and manage chatbots with built-in "soft skills" and personalization, without necessarily needing expert AI developers. In Juji.ai, we were able to define a structured conversation flow made up of topics (dialogue units). We specified how the chatbot responded (messages), what it asked (requests), and a set of expected user responses along with rules for how the conversation continued. Rules were created by mapping keywords, intents, or triggers in user input to pre-written answers or follow-up questions. In addition, we set up a Q&A library to handle frequently asked questions and fallback responses for unrecognized input. Once the flow was built, we deployed the chatbot on an external website.

Three versions of the chatbot were created with manipulated gender cues: a female chatbot named "Samantha," a male name "Samuel," and a gender-neutral name "Sam" (control), each with icons signifying female, male, and neutral gender. All chatbots introduced themselves with the same language and followed a consistent script covering topics like daily life, stressors, and happy memories. These topics were selected because they are universal, socially engaging, and neutral across demographics. Responses were designed to be simple and limited to predefined prompts (e.g., "Nice! Can you elaborate?"). Each participant interacted with the chatbot for no less than six minutes, with approximately twenty rounds of dialogue in each case. The average word count of each participants' input was 176.22 with an 85.29 standard deviation.

Participants first completed pretest measures on self-perceived femininity/masculinity. Then they were randomly assigned to different conditions to interact with the chatbot. After conversations, they answered post-test questionnaire items on their projected image of the chatbot, chatbot's credibility, attraction, and their overall familiarity with digital assistants, as well as demographic details.

Participants

Participants were recruited from a global online MBA program at a large midwestern U.S. university via SONA, which consists of individuals with diverse demographic backgrounds. We offered extra course credit as compensation. Among the 273 responses, 250 were retained after excluding 18 for failed attention checks and 5 for mismatched completion codes. The sample ranged in age from 24 to 66 ($M = 37.22$, $SD = 7.25$), with three missing values. In terms of gender, 101 were women, 147 were men, and two identified themselves as non-binary. Conditions were evenly distributed (see Supplementary Material). Participants who self-identified as non-binary or preferred not to answer were not included in the later factorial analyses because their sample size was insufficient to support reliable statistical comparisons within the 2 × 3 design.

Measures

The manipulation check was assessed using the method of projection. Participants were asked, "Based on your experience of chatting, what is the closest image of

Samantha/Samuel/Sam that comes to your mind?" Participants were asked to select from five image options (see Supplementary Material).

Perceived credibility was measured using a 21-item, 5-point semantic differential scale adapted from McCroskey and Teven (1999), comprising three subdimensions: competence, goodwill, and trustworthiness. Each subscale included seven items (e.g., "Incompetent – Competent," "Doesn't care about me – Cares about me," and "Untrustworthy – Trustworthy"). The overall scale demonstrated reliability ($\alpha = 0.93$), with high internal consistency for each dimension: competence ($\alpha = 0.92$), goodwill ($\alpha = 0.89$), and trustworthiness ($\alpha = 0.89$).

Social attraction was assessed with a five-item, five-point Likert-type scale adapted from Yao and Flanagin's study (2006). Participants rated their agreement (1 = strongly disagree, 5 = strongly agree) with statements such as "I think Samantha could be a friend of mine." The scale showed acceptable reliability ($\alpha = 0.79$).

Self-evaluated gender traits were measured using the 20-item Bem Sex Role Inventory (BSRI) adapted from Colley et al. (2009)'s research. Participants rated themselves on a five-point Likert-type scale (1 = strongly disagree, 5 = strongly agree) by responding to the question "I am…". The scale included two subdimensions: femininity ($\alpha = 0.80$) (e.g., "Tender," "Warm," "Gentle"), and masculinity ($\alpha = 0.86$) (e.g., "Assertive," "Dominant," "Aggressive").

Results

Manipulation check

Participants were asked to select one of the five images that best represented the chatbot they envisioned during the interaction. The five images included a female human, a male human, a gender-neutral human, a cyborg (half human, half machine), and a flat cartoon chatbot. To avoid unnecessary design complexity and to prevent introducing additional visual cues that might anchor participants' perceptions to specific physical embodiments rather than the conversational agent itself, gendered robotic depictions were not included. A chi-square test of independence indicated a significant association between participants' image selections and chatbot's manipulated gender, $\chi^2(8, N = 248) = 24.59$, $p = .002$. However, because 60% of cells had expected counts less than 5, the chi-square assumption was violated. Therefore, the Fisher–Freeman–Halton Exact Test was used to

Table 1 Study 1 participants' image projection * chatbot gender condition crosstabulation

		Condition						Total	
		Female chatbot		Male chatbot		Neutral chatbot			
		N	%	N	%	N	%	N	%
Image Projection	Female human	10	11.6	2	2.5	1	1.2	13	5.2
	Male human	0	0.0	2	2.5	3	3.6	5	2.0
	Neutral human	7	8.1	3	3.8	2	2.4	12	4.8
	Cyborg	26	30.2	15	19.0	15	18.1	56	22.6
	Cartoon	43	50.0	57	72.2	62	74.7	162	65.3
Total		86	100.0	79	100.0	83	100.0	248	100.0

confirm a significant association, $p = .001$. These results suggest that the distribution of projected images differed across chatbot gender conditions (see Table 1).

Inspection of the response patterns further indicates that, although human gender projections were rare overall, relative differences across conditions emerged. Participants interacting with the female chatbot were more likely to select a female human image (11.6%) than those in the male (2.5%) or neutral chatbot conditions (1.2%). Although we did notice this pattern, we acknowledge that our manipulation check involved additional confounding options, which rendered the manipulation check results less tenable. For example, even in the female chatbot condition, the majority of participants selected non-human representations. The manipulation influenced non-human projections: participants in the female chatbot condition were more likely to select cyborg images and less likely to select cartoon images than those in the male and neutral chatbot conditions: the gender manipulation exerted graded and indirect effects on participants' mental representations of the chatbot—shaping both limited human gender attribution and broader ontological interpretations—rather than producing a strong or uniform mapping onto human gender categories.

We see this as a methodological limitation, as gender cues may not have been sufficiently salient for all participants. At the same time, this aligns with Fortunati et al.'s (2022) study, which documented the lack of consistency in gender attribution, suggesting that users may recognize gendered cues while resisting full personification, treating the chatbot as neither fully human nor purely mechanical. Furthermore, their reaction toward cues of different genders may also differ, especially when the cues are subtle and minimal.

Main and interaction effects of chatbot and user gender identities on perceived credibility

To examine the effects of manipulated chatbots' and participants' gender identities, including their categorical genders and self-perceived gender traits, on perceived credibility on three dimensions (i.e., competence, goodwill, and trustworthiness), a multivariate analysis of covariance (MANCOVA) was conducted. The chatbot's gender and participants' gender were set as two independent variables. Masculinity and femininity were included in the models as continuous individual-difference variables capturing users' gender-related traits. Although entered as covariates, the analyses computed the effects of masculinity and femininity as predictors. The analyses revealed a significant effect of the independent variables on the combined dependent variables, Wilks' $\Lambda = 0.86$, $F(3, 236) = 13.21$, $p < .001$. The full results are specified in Table 2.

Table 2 MANCOVA results of perceived credibility

	Competence		Goodwill		Trustworthiness	
	F (p)	η_p^2	***F (p)***	η_p^2	***F (p)***	η_p^2
Chatbot Gender	0.62(.54)	0.01	1.33 (.27)	0.01	**5.28 (.006)**	0.04
Participant Gender	0.09 (.77)	0.00	**6.32 (.01)**	0.03	**4.02 (.046)**	0.02
Participant Femininity	**7.46 (.01)**	0.03	**11.29 (<.001)**	0.05	**17.18 (<.001)**	0.07
Participant Masculinity	0.10 (.75)	0.00	**8.43 (.004)**	0.03	4.15 (.043)	0.02
Chatbot Gender × Participant Gender	1.75 (.18)	0.02	0.88 (.42)	0.00	1.04 (.36)	0.09
Chatbot Gender × Participant Femininity	2.03 (.13)	0.02	**6.81 (.001)**	0.06	**8.40 (<.001)**	0.07
Chatbot Gender × Participant Masculinity	**3.25 (.04)**	0.03	1.17 (.31)	0.01	0.45 (.64)	0.00

Significant effects are presented in bold

Competence There was no significant main effect of the chatbot's gender on perceived competence, $F(2, 236) = 0.62$, $p = .540$, nor was there a significant main effect of participants' gender, $F(1, 236) = 0.09$, $p = .765$. However, participants' self-perceived femininity was a positive predictor of competence, $F(1, 236) = 7.46$, $p = .007$, $\eta_p^2 = 0.03$, while masculinity was not, $F(1, 236) = 0.10$, $p = .750$.

A significant interaction between chatbots' assigned gender and masculinity was observed, $F(2, 236) = 3.25$, $p = .041$, suggesting that the more masculine that the participants perceived themselves, the more they trusted the male chatbot's competence. Partial eta squared ($\eta_p^2 = 0.03$) indicated small effect. However, such relationship was not observed in the female chatbot or gender-neutral chatbot condition. The interaction effect between chatbots' genders and femininity did not reach statistical significance, $F(2, 236) = 2.03$, $p = .134$.

Goodwill There was no main effect of chatbots' genders, $F(2, 236) = 1.33$, $p = .27$. However, a significant main effect of participants' gender emerged, $F(1, 236) = 6.32$, $p = .013$, $\eta_p^2 = 0.03$, with men ($M = 3.09$, $SD = 0.97$) rating the chatbot as more benevolent than women ($M = 2.79$, $SD = 1.06$). Femininity positively predicted perceived goodwill of the chatbot, $F(1, 236) = 11.29$, $p < .001$, $\eta_p^2 = 0.05$, while masculinity negatively predicted the perceived goodwill of the chatbot, $F(1, 236) = 8.43$, $p = .004$, $\eta_p^2 = 0.03$.

There was also a significant interaction between the chatbot's gender and femininity, $F(2, 236) = 6.81$, $p = .001$, $\eta_p^2 = 0.06$, indicating that the influence of chatbot gender on perceived goodwill varied as a function of users' femininity. The more feminine participants perceived themselves, the more they trusted neutral or female chatbots in their goodwill; this effect did not extend to male chatbots. The interaction between chatbot gender and masculinity was not significant, $F(2, 236) = 1.17$, $p = .31$.

Trustworthiness A significant main effect of chatbot gender was found on perceived trustworthiness, $F(2, 236) = 5.28$, $p = .006$, $\eta_p^2 = 0.04$, indicating that the chatbot's gender significantly influenced perceptions of its integrity. However, post-hoc pairwise comparisons with Tukey adjustment did not reveal any specific significant group differences. There was also a significant main effect of participants' gender, $F(1, 236) = 4.02$, $p = .046$, $\eta_p^2 = 0.02$, with male participants rating the chatbot as more trustworthy ($M = 3.38$, $SD = 0.77$) than female participants ($M = 3.22$, $SD = 0.81$). Femininity positively impacted trustworthiness, $F(1, 236) = 17.18$, $p < .001$, $\eta_p^2 = 0.07$, while masculinity negatively impacted trustworthiness, $F(1, 236) = 4.15$, $p = .043$, $\eta_p^2 = 0.02$.

The interaction of the chatbot's gender and participants' femininity was statistically significant, $F(2, 236) = 8.40$, $p < .001$. The more strongly participants identified as feminine, the more they rated gender-neutral or female chatbots as trustworthy. Partial eta squared ($\eta_p^2 = 0.07$) indicated medium effect. No such effect was observed in the male chatbot condition. No significant interaction was found between condition and masculinity, $F(2, 236) = 0.45$, $p = .64$.

Main and interaction effects of chatbot and user gender identities on perceived social attraction

An ANCOVA was conducted to examine the effects of chatbot gender, participant gender, and participants' self-perceived femininity and masculinity, as well as their interactions, on perceived social attraction toward the chatbot (see Table 3).

There was a significant main effect of participants' genders, $F(1, 236) = 8.58$, $p = .004$, $\eta_p^2 = 0.04$, indicating that male participants reported higher social attraction ($M = 2.92$, $SD = 0.87$) of the chatbot than female participants ($M = 2.62$, $SD = 0.83$). There was no significant main effect of the chatbot's gender, $F(2, 236) = 0.11$, $p = .90$.

Femininity was a positive predictor of social attraction, $F(1, 236) = 8.69$, $p = .004$, $\eta_p^2 = 0.04$, while masculinity was not, $F(1, 236) = 2.08$, $p = .15$. This indicates that participants who perceived themselves as more feminine were more likely to feel socially drawn to the chatbot, regardless of the gender of the chatbot or themselves.

The interaction between the chatbot's gender and femininity was significant, $F(2, 236) = 3.26$, $p = .04$, $\eta_p^2 = 0.03$, suggesting that femininity moderated the relationship between the chatbot's gender and perceived social attraction with a small effect. The more feminine participants perceived themselves, the more socially attractive they rated neutral and female chatbots; no such effect was found for the male chatbot. Neither the interaction of the chatbot's gender and participants' gender, $F(2, 236) = 1.09$, $p = .34$, nor the interaction of the chatbot's gender and participants' masculinity, $F(2, 236) = 1.67$, $p = .19$, reached statistical significance.

According to the interaction plots (Figs. 1, 2, 3), on goodwill, trustworthiness, and social attraction, participants who perceived themselves as more feminine a) rated the gender-neutral chatbot higher; b) rated the male chatbot lower; and c) rated the female chatbot moderately. Participants who perceived themselves as more masculine a) rated the male chatbot higher; b) rated the female chatbot lower; and c) rated the gender-neutral chatbot moderately in terms of competence (Fig. 4). These patterns, although exploratory, added to the evidence that users bring gender attitudes and stereotypes into interactions with a gendered chatbot.

Summary of Study 1

In Study 1, the main effects of chatbot gender and users' categorical gender were less pronounced than those of gender-related traits. Femininity and masculinity played more substantial roles in shaping user evaluations. Notably, categorical gender and gender-related traits showed different patterns of association: male participants rated the CA

Table 3 Study 1 ANCOVA results of social attraction

	F	*p*	η_p^2
Chatbot Gender	0.11	.90	0.00
Participant Gender	8.58	.004	0.04
Participant Femininity	8.69	.004	0.04
Participant Masculinity	2.08	.15	0.01
Chatbot Gender × Participant Gender	1.09	.34	0.01
Chatbot Gender × Participant Femininity	3.26	.04	0.03
Chatbot Gender × Participant Masculinity	1.67	.19	0.01

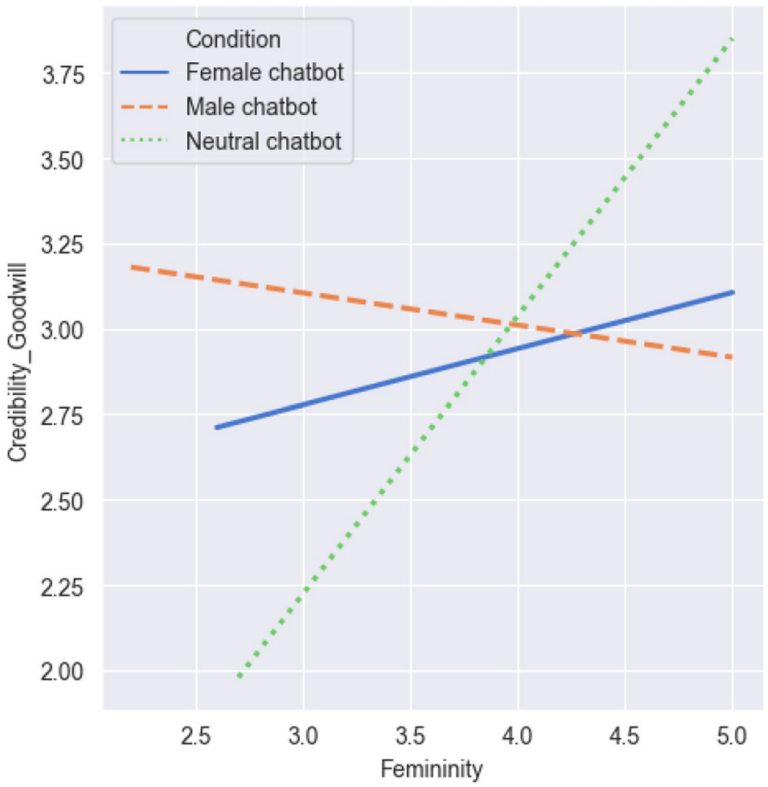

Fig. 1 Interaction of user-perceived femininity and chatbot gender on the goodwill dimension of credibility measure

higher on goodwill and trustworthiness than female participants, while femininity traits positively predicted these outcomes. This is not unexpected, as these constructs capture distinct aspects of identity. While categorical gender reflects group membership, masculinity and femininity represent self-perceived trait orientations that can vary within each gender group. Accordingly, their effects on outcomes such as goodwill and trustworthiness may diverge. Participants higher in femininity rated the chatbot higher in competence, goodwill, and trustworthiness, while masculinity showed less consistent effects and was negatively associated with goodwill and trustworthiness when other factors were controlled. More importantly, the interaction effects revealed greater nuance: the sole impact of chatbot gender design is conditional on participants' femininity and masculinity traits. Participants higher in femininity preferred gender-neutral or female chatbots, whereas participants higher in masculinity preferred male chatbots.

Study 2: A large language model-based chatbot

Study 1 was conducted in Spring 2022, prior to the widespread adoption of generative AI and LLM-based chatbots. It served as a pilot and baseline for Study 2. OpenAI released ChatGPT in November 2022, and its adoption grew rapidly thereafter, reaching one million users within five days (OpenAI, 2022; Yahoo Finance, 2022). Generative AI

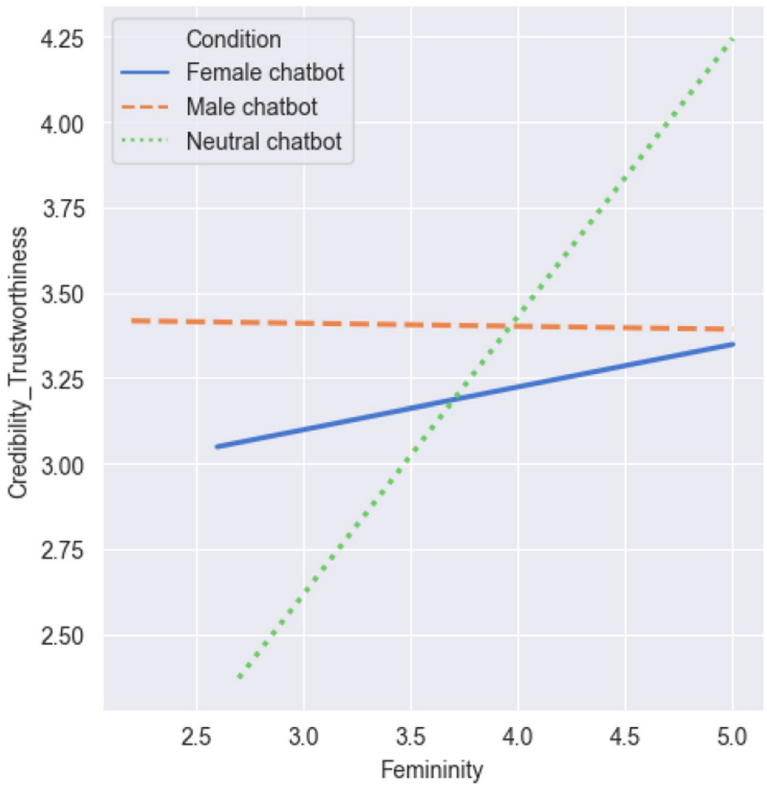

Fig. 2 Interaction of user-perceived femininity and chatbot gender on the trustworthiness dimension of credibility measure

continued to expand rapidly across both consumer and organizational contexts throughout 2023, with widespread uptake among individual users and growing adoption across business functions and workflows (e.g., McKinsey & Company, 2023). Building on this shift, Study 2 was conducted in Fall 2023. Accordingly, we adopted an LLM-based chatbot and refined the stimuli with minor methodological adjustments to better reflect contemporary agent capabilities while preserving the original theoretical structure; the chatbot's gender manipulation, including its self-introduction and visual representation, was updated to align with current design conventions and to strengthen the effectiveness of the manipulation; and the interaction context and interface were revised to reflect evolving AI use cases, allowing us to test the robustness of the proposed effects in a different yet theoretically consistent setting. In both studies, participants were not informed about the chatbot's underlying technical architecture to ensure responses were driven by perceived interaction characteristics.

Method

Design and procedure

Following Study 1's design framework, we conducted a new online experiment using chatbots powered by GPT-3.5, integrated through the OpenAI API into a custom chat

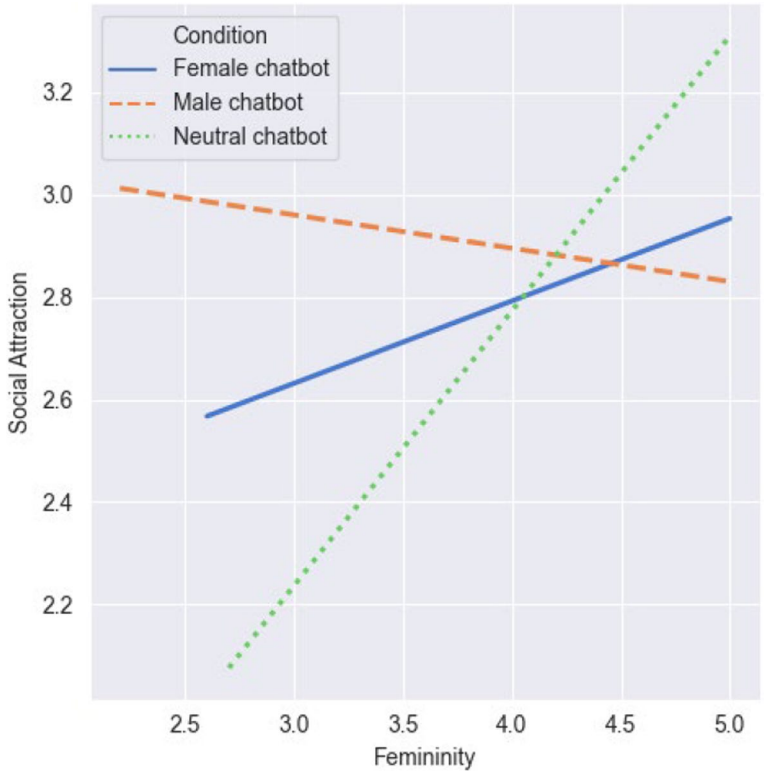

Fig. 3 Interaction of user-perceived femininity and chatbot gender on social attraction

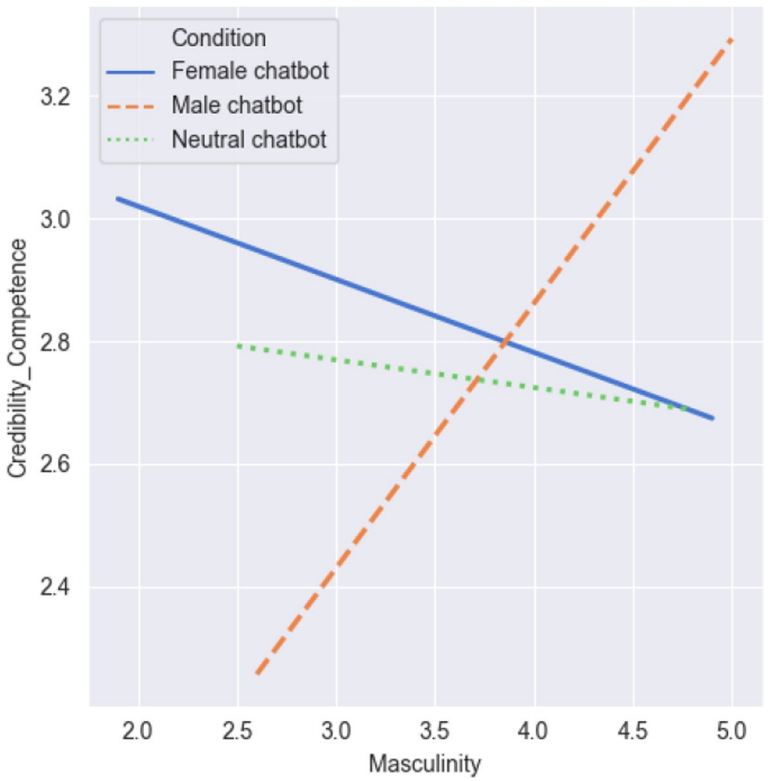

Fig. 4 Interaction of user-perceived masculinity and chatbot gender on the competence dimension of credibility measure

interface (ideta.io). The new experiment employed the same 2×3 design (participant gender: female vs. male; CA gender: female vs. male vs. neutral), with self-reported gender traits (femininity and masculinity) measured as continuous variables and tested as moderators. Three chatbot versions were created with gender cues: "Samantha" (female), "Samuel" (male), and "Sam" (neutral), each with corresponding names and icons (see Supplementary Material). Chatbot configuration included interface customization (appearance and greetings) and behavior adjustments through prompt-based in-context learning for self-introduction and engagement. Building on Study 1, the chatbot was framed as an AI companion in higher education to better align with the participant background and enhance ecological validity. This refinement also aimed to increase the relevance and salience of the interaction.

Participants were randomly assigned to one of the three chatbots. They began by completing a questionnaire on Qualtrics, covering informed consent and self-perceived femininity/masculinity. Drawing on the limitations identified in Study 1, we refined the chatbot stimuli to better align with contemporary, humanlike chatbot representations, thereby enhancing the clarity and salience of gender cues. To reinforce gender manipulation, we added a 30-s demo of the chatbot for participants to watch and get familiar with the features of the chatbot. Participants then engaged in a six-minute chat session. They were encouraged to have a free-form conversation with the chatbot without following any pre-defined scripts. The chatbot maintained conversation length by gently prompting participants to explore three different topics. If participants did not know what topics to initiate, similar topics from Study 1 were prompted, including recent updates (e.g., "what are you up to"), barriers, stressors, as well as some broader topics (e.g., goals and pursuit, vision of future) that may emerge depending on the flow of the conversation. The average word count of each participants' input was 146.18 with a 116.50 standard deviation. At the end, each participant received a unique 4-digit code from the chatbot, which they entered to proceed to a post-test questionnaire. The post-test questionnaire included measures based on participants' chat experience, along with the questions on digital assistant familiarity and participants' demographics.

Participants

Participants were recruited from the same global iMBA participant pool used in Study 1, but from different cohort classes. We offered 1 extra course credit as compensation. Among the 280 participants, responses from 228 of them were retained after excluding 14 with unmatched completion codes and 38 that failed the attention check. Ages ranged from 18 to 62 ($M = 37.57$, $SD = 8.30$). Participants included 77 women, 147 men, and 4 who were non-binary or preferred not to answer. The small number of participants who were non-binary or did not disclose their gender were not included in the statistical analyses because of the insufficient statistical power. The sample was evenly distributed across conditions (see Supplementary Material).

Measures

Measures were retained from Study 1, with minor refinements to the credibility scale to improve clarity. Specifically, the items in the credibility scale were adapted by referring to the Trusting Beliefs Scale (McKnight et al., 2002). The new scale assessed the

same three dimensions (competence, goodwill, and trustworthiness) with four items per dimension. Competence evaluates skills and abilities (e.g., "I think Samantha is incompetent–competent," $\alpha = 0.93$); goodwill assesses responsiveness and benevolence (e.g., "I think Samantha does not act in my best interest–acts in my best interest," $\alpha = 0.87$); trustworthiness focuses on honesty and integrity (e.g., "I think Samantha is dishonest–honest," $\alpha = 0.90$).

Results

Manipulation check

We adopted the same approach of manipulation check as in Study 1. However, based on the findings from Study 1, where participants rarely mapped chatbots onto gender-neutral human categories and instead tended to select non-human representations, the gender-neutral image option was removed in Study 2 to reduce ambiguity and distraction and to strengthen the sensitivity of the manipulation check. A Chi-squared test of independence was conducted to determine whether participants' selection of the projected image of the chatbot was related to the chatbot's gender manipulation, χ^2 (8, $N = 224$) = 309.54, $p < .001$.

The image projection results (see Table 4) in Study 2 indicate substantially stronger and more consistent gender attribution than in Study 1. Participants who interacted with the female chatbot overwhelmingly selected a female human image (77.3%), while those who interacted with the male chatbot predominantly selected a male human image (83.1%). By contrast, participants in the gender-neutral chatbot condition were far less likely to select gendered human images and instead favored non-human representations, particularly cartoon (60.0%) and cyborg (29.4%) images.

We attribute this pattern primarily to the more explicit gender manipulation implemented in Study 2. Compared to Study 1, the revised manipulation in Study 2 provided clearer and more consistent gender cues throughout the interaction, which likely reduced ambiguity and enabled participants to form more stable and gender-congruent mental representations of the chatbot. As a result, participants were more inclined to map the chatbot onto human gender categories, rather than defaulting to abstract or non-human representations.

This contrast suggests that the limited human gender projection observed in Study 1 was not due to an inherent resistance to gender attribution per se, but rather to the

Table 4 Study 2 participants' image projection * chatbot gender condition crosstabulation

		Condition						Total	
		Female chatbot		Male chatbot		Neutral chatbot			
		N	%	N	%	N	%	N	%
Image projection	Female human image	68	77.3	1	1.1	4	4.7	73	27.9
	Male human image	1	1.1	74	83.1	5	5.9	80	30.5
	Cyborg image	9	10.2	3	3.4	25	29.4	37	14.1
	Cartoon image	10	11.4	11	12.4	51	60.0	72	27.5
Total		88	100.0	89	100.0	85	100.0	262	100.0

relative subtlety of the initial manipulation. When gender cues were made more salient and coherent in Study 2, participants readily engaged in anthropomorphism. Taken together, these findings highlight the importance of manipulation strength and cue consistency in shaping users' gendered perceptions of chatbots.

Main and interaction effects of chatbot and user gender identities on perceived credibility

Same as in Study 1, a MANCOVA was conducted to examine the effects of the chatbot's gender, participants' gender, and participants' self-perceived gender traits on perceived credibility along three dimensions (i.e., competence, goodwill, and trustworthiness). The multivariate analysis revealed a significant effect of the independent variables on the combined dependent variables, Wilks' $\Lambda = 0.78$, $F(3, 210) = 19.88$, $p < .001$. The results are summarized in Table 5. We further focus on reporting the univariate results.

Competence No significant main effects were found for the chatbot's gender, $F(2, 212) = 0.20$, $p = .82$, or participants' gender, $F(1, 212) = 0.20$, $p = .66$, on perceived competence of the chatbot. Although the effect of femininity approached marginal significance, $F(1, 212) = 3.07$, $p = .08$, neither masculinity, $F(1,212) = 0.14$, $p = .70$, nor any interaction effects were significant (all ps > .05).

Goodwill None of the main effects was statistically significant: the chatbot's gender, $F(2, 212) = 0.78$, $p = .46$, participants' gender, $F(1, 212) = 2.34$, $p = .13$, femininity, $F(1, 212) = 0.52$, $p = .47$, and masculinity, $F(1, 212) = 2.77$, $p = .10$, had no effects on perceived goodwill of the chatbot. All interaction effects were statistically non-significant (ps > .31).

Trustworthiness None of the main effects was statistically significant: the chatbot's gender, $F(2, 212) = 0.30$, $p = .74$; participants' gender, $F(1, 212) = 2.45$, $p = .12$, femininity, $F(1, 212) = 1.02$, $p = .31$, and masculinity, $F(1, 212) = 0.22$, $p = .64$, had no effects on perceived trustworthiness of the chatbot. Likewise, none of the interaction effects reached statistical significance (ps > .24).

Across all three dimensions of credibility, the results revealed no main or interaction effects among chatbots' and users' categorical genders, as well as users' gender traits.

Table 5 Study 2 MANCOVA results of perceived credibility

	Competence		Goodwill		Trustworthiness	
	F (p)	η_p^2	***F (p)***	η_p^2	***F (p)***	η_p^2
Chatbot Gender	0.20 (.82)	0.02	0.78 (.46)	0.01	0.30 (.74)	0.00
Participant Gender	0.20 (.66)	0.01	2.34 (.13)	0.01	2.45 (.12)	0.01
Participant Femininity	3.07 (.08)	0.01	0.52 (.47)	0.00	1.02 (.31)	0.01
Participant Masculinity	0.14 (.71)	0.00	2.77 (.10)	0.01	0.22 (.64)	0.00
Chatbot Gender × Participant Gender	1.57 (.21)	0.02	0.95 (.39)	0.01	0.11 (.89)	0.00
Chatbot Gender × Participant Femininity	0.00 (1.00)	0.00	1.17 (.31)	0.01	1.41 (.25)	0.01
Chatbot Gender × Participant Masculinity	0.29 (.75)	0.00	0.02 (.98)	0.01	0.20 (.82)	0.00

Main and interaction effects of chatbot and user gender identities on perceived social attraction

An ANCOVA was conducted to examine the effects of chatbot gender, participant gender, and participants' self-perceived femininity and masculinity, as well as their interactions, on perceived social attraction of the chatbot (see Table 6).

No significant main effects were found for the chatbot's gender, $F(2, 212) = 0.94$, $p = .39$, participants' gender, $F(1, 212) = 2.27$, $p = .13$, femininity, $F(1, 212) = 2.34$, $p = .13$), masculinity, $F(1, 212) = 0.68$, $p = .41$, nor any interaction effects, including the interaction of chatbot gender and participant gender, chatbot gender and femininity, or chatbot gender and masculinity, were significant (all $ps > .25$).

Summary of Study 2

In Study 2, none of the chatbot gender, user gender, or user gender traits significantly influenced perceived credibility (competence, goodwill, trustworthiness) or social attraction. Femininity only approached marginal significance for competence, but no other main or interaction effects emerged. These results suggest that, under the LLM-based chatbot condition, perceptions of credibility and attraction were relatively stable across categorical gender identities and gender traits.

The comparison indicates that while gender identities and traits mattered in the rule-based chatbot of Study 1, such effects were not evident in the LLM setting of Study 2. This pattern may suggest that the sophistication and fluency of LLM-driven interactions may attenuate social identity-based biases, leading to more uniform evaluations of chatbot credibility and attractiveness.

Discussion

This two-study project examines how chatbots' assigned genders, users' gender identities, and gender traits jointly shape perceptions of two different types of chatbots: a traditional rule-based chatbot (Study 1) and an LLM-based chatbot (Study 2). Taken together, the findings suggest that gendered evaluations in HMC are conditional rather than universal, emerging most clearly in constrained interactions and attenuating as chatbots become more fluent, adaptive, and socially responsive.

Effects of gendered chatbot designs and ethical discussion

In Study 1, users brought their own gendered self-concepts into interactions with the chatbot. An interesting pattern in our findings is that categorical gender and

Table 6 Study 2 ANCOVA results of social attraction

	F (p)	η_p^2
Chatbot Gender	0.94 (.39)	0.01
Participant Gender	2.27 (.13)	0.01
Participant Femininity	2.34 (.13)	0.01
Participant Masculinity	0.68 (.41)	0.00
Chatbot Gender × Participant Gender	0.77 (.47)	0.01
Chatbot Gender × Participant Femininity	1.37 (.26)	0.01
Chatbot Gender × Participant Masculinity	0.29 (.75)	0.00

gender-related traits demonstrated different, and in some cases opposing, associations with outcomes such as goodwill and trustworthiness. Male participants rated the chatbot higher in goodwill and trustworthiness than female participants, whereas participants' femininity was positively associated with goodwill, trustworthiness, and social attraction. This highlights the importance of distinguishing between gender as a categorical identity and as a set of internalized traits. While categorical genders may reflect broader social identities, gender-related traits appear to shape how users interpret and respond to social cues in interaction. This distinction aligns with prior gender schema research (Bem, 1974; Spence & Helmreich, 1978), suggesting that masculinity and femininity operate as independent dimensions that can vary within gender groups, even in HMC contexts. More interestingly, statistically significant interaction effects revealed that users' gender traits moderated how chatbot gender cues were interpreted across credibility dimensions and social attraction. Specifically, participants higher in masculinity rated a male chatbot as more competent, whereas participants higher in femininity attributed greater goodwill and trustworthiness to gender-neutral or female chatbots.

These findings align with recent work in that gender cues alone are often insufficient to shape user perceptions. Instead, they impose influence through interactional configurations and identity alignment (Piercy et al., 2025). Similarly, Lee et al. (2024) found that men were more likely to trust a male-voiced assistant than a female-voiced one, whereas women did not show a corresponding preference for female voices. These results extend earlier CASA-based findings (e.g., Nass et al., 1997) by demonstrating that gendered responses to machines are nuanced, trait-dependent, and context-sensitive, particularly in text-based human–chatbot interaction.

The shift of communication nature

In contrast to Study 1, Study 2 yielded largely nonsignificant interaction effects. Participants who interacted with a generative AI–powered chatbot did not systematically differentiate the agent based on chatbot gender, user gender identity, or gender traits (see Table 7). One explanation for this divergence lies in broader shifts in users' expectations and the nature of HMC following the rapid diffusion of generative AI. The surge in generative AI popularity, driven by its expanding capabilities, has increased public interest in AI and fostered more favorable and optimistic attitudes toward AI systems (Miyazaki et al., 2024). As chatbots increasingly resemble advanced, capable, and socially fluent systems, users may rely less on gendered heuristics when forming evaluations.

We drew further support of this interpretation from the descriptive statistics (see Supplementary Material). In Study 2, ratings of competence, trustworthiness, goodwill, and social attraction clustered toward the higher end of the scale, suggesting a possible ceiling effect that constrained the detection of interaction effects. One explanation is that users evaluated the LLM-based chatbot uniformly positively, likely due to its high fluency and responsiveness compared to the constrained, rule-based chatbot in Study 1. In this sense, the potential ceiling effect may reflect not only a methodological constraint but also a substantive shift in how users interpret and evaluate AI systems. As CAs become more capable, users may rely less on surface-level cues such as gender and instead form judgments based on overall interaction quality. Accordingly, the diminished interaction effects observed in Study 2 may indicate that assumptions about the

Table 7 Summary of significance of interaction effects across two studies

	Dependent Variables	Study 1	Study 2
Gender ✕ Gender	Competence	-	-
	Goodwill	-	-
	Trustworthiness	-	-
Gender ✕ Femininity	Competence	-	-
	Goodwill	☑	-
	Trustworthiness	☑	-
Gender ✕ Masculinity	Competence	☑	-
	Goodwill	-	-
	Trustworthiness	-	-

gender role of AI are becoming less salient in shaping evaluative judgments in higher-fidelity interaction contexts.

In terms of communication structure, the two studies also differed fundamentally. The rule-based chatbot used in Study 1 relied on predetermined, scripted interactions with limited variation and fixed conversational goals. While this structure afforded strong experimental control, it was less representative of the naturalistic interaction between humans and chatbots and likely encouraged reliance on heuristic cues. By contrast, the generative AI–powered agent in Study 2 supported spontaneous, user-initiated, and open-ended dialogue, allowing users to pursue diverse communication goals within a single interaction. This increased conversational freedom and realism may have reduced users' reliance on gendered expectations when evaluating the agent. These differences suggest that gendered evaluations of chatbots are contingent on interactional constraints and system sophistication rather than stable effects of agent gender alone.

At the same time, alternative methodological explanations should be acknowledged. Although the two studies were conceptually linked, they differed in contextual framing, task structure, and interaction design, which may have introduced measurement inconsistency across studies. In addition, Study 2 involved a richer and more naturalistic interaction with a generative AI agent, which may have reduced experimental control relative to Study 1. Such differences may attenuate or obscure interaction effects that are detectable in more controlled experimental settings; however, they also reflect real-world conditions, in which most human–AI conversations are no longer tightly constrained.

These findings suggest that perceptions of chatbots as "AI" have evolved substantially. Whereas earlier interactions may have activated interpersonal stereotypes associated with humanlike agents, contemporary generative AI systems appear to elicit more uniformly positive evaluations, potentially limiting the relevance of earlier bias-based interaction patterns. Consequently, the interaction effects observed in Study 1 may be less generalizable in the rapidly evolving landscape of generative AI–mediated communication.

Theoretical implications

These findings indicated an integrated effect of social identity theory (emphasizing categorical gender identity alignment) and gender schema theory (emphasizing individuals' femininity and masculinity) in HMC, whereby gendered evaluations of chatbots are not solely driven by chatbot design but emerge from the interaction between agent cues and users' internalized gender traits. Our study contributes to theories related to HMC by clarifying when and how gender schemas can shape evaluations of chatbots.

First, the results extend gender schema theory (Bem, 1981; Martin & Halverson, 1981) by demonstrating users' gendered evaluations toward chatbots under strictly and loosely constrained interactional conditions. However, these effects are more complicated and conditional in HMC, which depend on factors including users' understanding of the ontology of the chatbots. Compared to more free-form, flexible conversations in Study 2, in the scripted, rule-based interactions of Study 1, users relied more on internalized gender schemas, producing trait-congruent evaluations of competence, goodwill, and trustworthiness. Consistent with a performative view of gender (Butler, 1990), chatbot gender is enacted through interactional cues rather than treated as an inherent attribute.

This perspective helps explain why gender effects were more pronounced when the chatbot's conversational behavior was limited and predictable.

Second, the attenuation of gender effects in Study 2 refines assumptions derived from the Computers Are Social Actors (CASA) paradigm (Nass & Moon, 2000). While CASA suggests that people apply social expectations to machines with social cues, the present findings indicate that such responses are not static. Rather, they depend on system sophistication and interactional richness. As chatbots become more fluent, adaptive, and context-sensitive, users may appear less reliant on categorical gender cues and more focused on the substance and quality of the interaction itself.

Additionally, we speculate that richer communication environments reduce reliance on heuristic or peripheral cues by supporting greater deliberation and meaning construction. LLM–based interactions may therefore diminish the salience of gender stereotypes by shifting users' attention from identity cues to conversational content. At the same time, richer interaction may also render gendered assumptions more implicit and harder to detect through self-reported measures alone, suggesting important boundary conditions for how bias manifests in advanced HMC.

Finally, by examining the effects of both categorical gender identity alignment and the moderation of gender traits, this study offers a more nuanced account of identity processes in HMC. Rather than treating gender as a fixed demographic variable, the findings highlight gender traits as psychologically meaningful dimensions that interact with agent design and interactional context to shape both evaluative (credibility) and relational (social attraction) outcomes. Together, these contributions position gender in HMC as a relational and interactional construct, emerging from the interplay of user traits, agent cues, and communicative affordances rather than as a stable effect of agent gender alone.

Design implications

The findings of the two studies offer design implications for chatbots, particularly in light of the shifting nature of HMC brought by generative AI and LLMs. Our results suggest that gender cues matter most in constrained, lower-fidelity interactions, while the effects fade away in richer, LLM-based conversations. Designers should therefore consider when, how, and why gender is incorporated into CA design. Effective and ethical CA design should move beyond surface-level design features and instead account for the interactional contexts, user identity dynamics, and the evolving communicative affordances of generative AI and LLMs.

One implication concerns the strategic use of identity alignment. Prior work shows that aligning an agent's identity with a user's identity can enhance trust and engagement in specific application contexts (Bickmore et al., 2021; Zhou & Bickmore et al., 2022). Our findings extend the work by showing that such benefits are conditional: in more scripted interactions (Study 1), users relied more on gendered expectations shaped by their gender traits, whereas these effects no longer emerged in more naturalistic, LLM-based interactions (Study 2). At the same time, reliance on gender matching raises ethical concerns. Designing for congruence may inadvertently reinforce gender stereotypes and homophily bias (Adams & Loideáin, 2019; D'Ignazio & Klein, 2020), creating a tension between optimizing trust and promoting inclusive interaction. Our results suggest

that gender alignment should not be assumed as a default design strategy, particularly as advanced AI systems may reduce users' reliance on such cues.

Our findings also speak to ongoing debates around gender-neutral or genderless design (e.g., Mooshammer & Etzrodt, 2021). Text-based agents often lack explicit gender markers, which may reduce the activation of gender schemas (Sutton, 2020), potentially explaining the findings in Study 2. However, prior research has cautioned that users may still infer gender from linguistic styles, names, or cultural contexts (Danielescu, 2020). Thus, neutrality in design intent does not guarantee neutrality in user perception. For voice-based or embodied agents, attempts to implement gender-neutral voices have met mixed responses, with some users perceiving such designs as less relatable or emotionally expressive (Yeon et al., 2023). Rather than prescribing a single "genderless" solution, our findings support context-sensitive and configurable designs, allowing users to engage with agents in ways that align with their preferences without defaulting to stereotypical representations.

Limitations

First, this study is limited to comparing users' perceptual evaluations instead of behavioral responses. The nuances in the social scripts of human-CA interaction, which lie in the content of the conversation, have yet to be analyzed. Further investigations are needed to better capture the social dynamics in the conversations.

Second, we tested the gender effects mainly with some gender-neutral communication topics with the intention of detecting the effects in relatively generic contexts. The loosely controlled conversational contexts and the differing time points at which the two studies were conducted might have contributed to the nonsignificant results in Study 2. Future research should consider communication contexts as an important boundary factor to determine the condition of gender influence in HMC. In future research, we should sample and select diverse conversation topics to re-examine the questions in this project.

Another limitation of the present studies concerns the manipulation checks. One limitation of Study 1 is that the gender manipulation was relatively weak, as participants did not consistently map the chatbot onto human gender categories. This reflects the ontological ambiguity of CAs and underscores the importance of strengthening cue salience, which we addressed in Study 2. In terms of the manipulation checks, although the projection method was intended to capture participants' spontaneous gender associations rather than represent the chatbot directly, the use of human images introduced specific visual features (e.g., hairstyle and skin color) that might have introduced unintended confounds. The cyborg and flat cartoon images could also be distracting to participants. In addition, perceived masculinity and femininity of the chatbot were not measured, as chatbot gender was treated as a categorical manipulation rather than a trait-based perception. Future research could employ cleaner, more controlled visual materials and incorporate direct measures of perceived chatbot masculinity and femininity to enable more nuanced examinations of categorical versus trait-based gender perceptions.

Finally, a small number of participants identified themselves as non-binary or preferred not to disclose their gender. Because the number of these participants was too small to support reliable statistical estimation within the factorial design, they were excluded from the inferential analyses. Future research could recruit sufficiently large and diverse samples to enable analyses that capture and represent the experiences and perceptions of non-binary and gender-diverse users in HMC.

Conclusion

The two studies in this article suggest that gender in HMC is not merely a property of the conversational agent, but a relationship shaped by users' gendered self-concepts and the agent's assigned gender cues. In rule-based interactions (Study 1), trait congruence mattered: more masculine users trusted male agents more, while users higher in femininity favored neutral or female agents on goodwill and trustworthiness. This finding provides evidence that social schemas can be transferred to machine encounters. Yet these effects fade away in conversations with LLM-based chatbots (Study 2), suggesting that richer, more natural dialogue can dilute, or at least mask identity-based biases. The contrast underscores a key design tension: personalization that matches user identities may improve comfort and trust, but it also risks re-inscribing gender stereotypes and marginalizing non-binary identities. We argue for inclusive, identity-aware design that treats gender cues as deliberate, ethically bounded choices rather than default shortcuts. Future work should move beyond self-reports to behavioral and conversational measures, examine stereotype-salient domains and topics, and test interventions (e.g., gender-neutral or configurable personas) that balance usability with equity. As conversational AI becomes an increasingly pervasive interface for services, learning, and decision-making, designing agents that are both effective and inclusive will require attention to social identity dynamics while leveraging the communicative richness of generative AI.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s44382-026-00037-0.

Supplementary Material 1.

Authors' contributions

W.L. conducted the experiments, analyzed the data, and wrote the main manuscript text and K.X. edited the main manuscript text. All authors reviewed the manuscript.

Data availability

Data will be made available upon request.

Declarations

Ethics approval and consent to participate

Both studies received IRB approval from the University of Illinois at Urbana-Champaign (project #20659).

Competing interests

The authors declare no competing interests.

Received: 2 October 2025 Revised: 26 June 2026 Accepted: 7 July 2026
Published online: 10 August 2026

References

Adams, R., & Loideáin, N. N. (2019). Addressing indirect discrimination and gender stereotypes in AI virtual personal assistants. *International Review of Law, Computers & Technology, 33*(2), 164–182.

Adinolf, S., Conroy, D., Wyeth, P., & Simpson, L. (2025). The impact of gender and level of control over agents' aesthetics on user experiences in VR training. *International Journal of Human-Computer Interaction, 41*(24), 1–17.

Ahn, J., Kim, J., & Sung, Y. (2022). The effect of gender stereotypes on artificial intelligence recommendations. *Journal of Business Research, 141*, 50–59.

American Psychological Association. (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American Psychologist, 70*(9), 832–864. https://doi.org/10.1037/a0039906

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 42(2), 155–162. https://doi.org/10.1037/h0036215

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88*(4), 354–364. https://doi.org/10.1037/0033-295X.88.4.354

Bem, S. L., Martyna, W., & Watson, C. (1976). Sex typing and androgyny: Further explorations of the expressive domain. *Journal of Personality and Social Psychology, 34*(5), 1016–1023. https://doi.org/10.1037/0022-3514.34.5.1016

Benbasat, I., Dimoka, A., Pavlou, P. A., & Qiu, L. (2020). The role of demographic similarity in people's decision to interact with online anthropomorphic recommendation agents: Evidence from a functional magnetic resonance imaging (fMRI) study. *International Journal of Human-Computer Studies, 133*, 56–70.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bickmore, T., Parmar, D., Kimani, E., & Olafsson, S. (2021). Diversity informatics: Reducing racial and gender bias with virtual agents. *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*, 25–32. https://doi.org/10.1145/3472306.3478365

Brahnam, S., & De Angeli, A. (2012). Gender and manners in artificial agents. *AI & Society, 27*(4), 369–378. https://doi.org/10.1007/s00146-011-0345-0

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.

Colley, A., Mulhern, G., Maltby, J., & Wood, A. M. (2009). The short form BSRI: Instrumentality, expressiveness and gender associations among a United Kingdom sample. *Personality and Individual Differences, 46*(3–4), 384–387. https://doi.org/10.1016/j.paid.2008.11.005

Curry, A. C., Reis, H. T., & Malle, B. F. (2020). How social cues shape human–robot interaction: The impact of robot gender and language style on trust and warmth. *ACM Transactions on Human-Robot Interaction (THRI), 9*(3), Article 20. https://doi.org/10.1145/3588967.3588970

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.

Danielescu, A. (2020). Eschewing gender stereotypes: Designing voice assistants with personality. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376590

Du, H., Niyato, D., Kang, J., Xiong, Z., Zhang, P., Cui, S., ... & Kim, D. I. (2024). The age of generative AI and AI-generated everything. *IEEE Network, 38,* 501–512.

Duan, W., McNeese, N., & Li, L. (2025b). Gender stereotypes toward non-gendered generative AI: The role of gendered expertise and gendered linguistic cues. *Proceedings of the ACM on Human-Computer Interaction, 9*(1), 1–35.

Duan, W., Li, L., Freeman, G., & McNeese, N. (2025a). A scoping review of gender stereotypes in artificial intelligence. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–20.

Edwards, A., & Etzrodt, K. (2025). Being and becoming in human–machine communication: Core commitments and conceptual foundations of a trans-ontological field. *Human-Machine Communication, 10*, 7–30. https://doi.org/10.30658/hmc.10.1

Eyssel, F., & Hegel, F. (2012). (S)he's got the look: Gender stereotyping of robots. *Journal of Applied Social Psychology, 42*(9), 2213–2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2020). Gender bias in chatbot design. In A. Følstad, T. Araujo, S. Papadopoulos, E.L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), *Chatbot research and design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, revised selected papers* (pp. 79–93). Springer. https://doi.org/10.1007/978-3-030-39540-7_6

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77–83.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines, 30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Fortunati, L., Edwards, A., Edwards, C., Manganelli, A. M., & De Luca, F. (2022). Is Alexa female, male, or neutral? A cross-national and cross-gender comparison of perceptions of Alexa's gender and status as a communicator. *Computers in Human Behavior, 137*, Article 107426.

Fossa, F., & Sucameli, I. (2022). Gender bias and conversational agents: An ethical perspective on social robotics. *Science and Engineering Ethics, 28*(3), 23. https://doi.org/10.1007/s11948-022-00376-3

Gessinger, I., Seaborn, K., Steeds, M., & Cowan, B. R. (2025). ChatGPT and me: First-time and experienced users' perceptions of ChatGPT's communicative ability as a dialogue partner. *International Journal of Human-Computer Studies, 194*, Article 103400.

Goodman, K. L., & Mayhorn, C. B. (2023). It's not what you say but how you say it: Examining the influence of perceived voice assistant gender and pitch on trust and reliance. *Applied Ergonomics, 106*, Article 103864. https://doi.org/10.1016/j.apergo.2022.103864

Guzman, A. L. (2018). What is human-machine communication, anyway? In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 1–28). Peter Lang.

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication, 1*, 37–54.

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society, 22*(1), 70–86.
Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication, 25*(1), 89–100.
Hirokawa, K., Yagi, A., & Miyata, Y. (2004). An experimental examination of the effects of sex and masculinity/femininity on psychological, physiological, and behavioral responses during communication situations. *Sex Roles, 51*(1), 91–99.
Jeon, J. E. (2024). The effect of AI agent gender on trust and grounding. *Journal of Theoretical and Applied Electronic Commerce Research, 19*(1), 692–704.
Jin, E., & Eastin, M. (2024). Gender bias in virtual doctor interactions: Gender matching effects of chatbots and users on communication satisfactions and future intentions to use the chatbot. *International Journal of Human-Computer Interaction, 40*(23), 8246–8258.
Kotek, H., Dockum, R., & Sun, D. (2023, November). Gender bias and stereotypes in large language models. *Proceedings of the ACM collective intelligence conference*, 12–24. https://doi.org/10.1145/3582269.3615599
Lee, S. K., Park, H., & Kim, S. Y. (2024). Gender and task effects of human–machine communication on trusting a Korean intelligent virtual assistant. *Behaviour & Information Technology, 43*(16), 4172–4191.
Lee, K. M., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender? An experimental test of gender stereotypes. *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, 289–290. https://doi.org/10.1145/633292.633450
Lew, Z., & Walther, J. B. (2023). Social scripts and expectancy violations: Evaluating communication with human or AI chatbot interactants. *Media Psychology*, *26*(1), 1–16. https://doi.org/10.1080/15213269.2022.2084111
Liu, J. (2024). ChatGPT: Perspectives from human–computer interaction and psychology. *Frontiers in Artificial Intelligence, 7*, 1418869.
Liu, W., & Yao, M. (2023). Gender identity and influence in human-machine communication: A mixed-methods exploration. *Computers in Human Behavior, 144*, Article 107750.
Martin, C. L., & Halverson, C. F., Jr. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development, 52*, 1119–1134.
McCroskey, J. C., & McCain, T. A. (1974). The measurement of interpersonal attraction. *Speech Monographs, 41*(3), 261–266. https://doi.org/10.1080/03637757409375845
McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communication Monographs, 66*(1), 90–103. https://doi.org/10.1080/03637759909376464
McKinsey & Company. (2023, August 1). *The state of AI in 2023: Generative AI's breakout year*. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year
McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research, 13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81
Mehrotra, S., Jonker, C. M., & Tielman, M. L. (2021, July). More similar values, more trust?-the effect of value similarity on trust in human-agent interaction. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 777–783. https://doi.org/10.1145/3461702.3462576
Miyazaki, K., Murayama, T., Uchiba, T., An, J., & Kwak, H. (2024). Public perception of generative AI on Twitter: An empirical study based on occupation and usage. *EPJ Data Science, 13*(1), 2.
Mooshammer, S., & Etzrodt, K. (2021, November). Social research with gender-neutral voices in chatbots–the generation and evaluation of artificial gender-neutral voices with praat and google wavenet. In *International Workshop on Chatbot Research and Design* (pp. 176–191). Cham: Springer International Publishing.
Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human–computer relationship*. MIT Press.
Nass, C., & Moon, Y. (2000). Machines and mindlessness: social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153
Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology, 27*(10), 864–876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x
OpenAI. (2022, November 30). *ChatGPT: Optimizing language models for dialogue*. https://openai.com/blog/chatgpt
Piercy, C. W., Montgomery-Vestecka, G., & Lee, S. K. (2025). Gender and accent stereotypes in communication with an intelligent virtual assistant. *International Journal of Human-Computer Studies, 195*, Article 103407. https://doi.org/10.1016/j.ijhcs.2024.103407
Reinkemeier, F., & Gnewuch, U. (2022). Match or mismatch? How matching personality and gender between voice assistants and users affects trust in voice commerce. In *Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS)*.
Spence, J. T., & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents*. University of Texas Press.
Sutton, S. J. (2020, July). Gender ambiguous, not genderless: Designing gender in voice user interfaces (VUIs) with sensitivity. *Proceedings of the 2nd conference on conversational user interfaces*, 1–8. https://doi.org/10.1145/3405755.3406123
Tajfel, H., & Turner, J. (1979). An integrative theory of inter-group conflict. In J. A. Williams & S. Worchel (Eds.), *The social psychology of inter-group relations* (pp. 33–47). Wadsworth.
van der Goot, M. J., & Etzrodt, K. (2023). Disentangling two fundamental paradigms in human-machine communication research: Media equation and media evocation. *Human-Machine Communication 6*, 17–30. https://doi.org/10.30658/hmc.6.2
West, S. M., Kraut, R., & Ei Chew, H. (2019). *I'd blush if I could: Closing gender divides in digital skills through education*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000367416
Yahoo Finance. (2022, December 9). *ChatGPT gained 1 million users in under a week. Here's why the AI chatbot is primed to disrupt search as we know it*. Yahoo Finance. https://finance.yahoo.com/news/chatgpt-gained-1-million-followers-224523258.html
Yao, M. Z., & Flanagin, A. J. (2006). A self-awareness approach to computer-mediated communication. *Computers in Human Behavior, 22*(3), 518–544. https://doi.org/10.1016/j.chb.2004.10.008

Yeon, J., Park, Y., & Kim, D. (2023). Is gender-neutral AI the correct solution to gender bias? Using speech-based conversational agents. *Archives of Design Research, 36*(2), 63–91.
Zargham, N., Dubiel, M., Desai, S., Mildner, T., & Belz, H. J. (2024, December). Designing AI personalities: enhancing human-agent interaction through thoughtful persona design. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia* (pp. 490–494).
Zhang, Q., Yang, X. J., & Robert, L. P., Jr. (2025). Artificial intelligence voice gender, gender role congruity, and trust in automated vehicles. *Scientific Reports, 15*(1), Article 16364.
Zhou, S., & Bickmore, T. (2022, April). A virtual counselor for breast cancer genetic counseling: adaptive pedagogy leads to greater knowledge gain. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–17.

Publisher's Note