The Dark Side of Anthropomorphism within AI-Driven Customer Experience: Manipulation, Data Disclosure and Consumer Vulnerability

Amit Punia

Department of Management Studies, Central University of Haryana

DOI: https://doi.org/10.51244/IJRSI.2026.1306000276

Received: 19 June 2026; Accepted: 24 June 2026; Published: 04 July 2026

ABSTRACT

Marketing and information systems scholarship has largely treated anthropomorphism in AI-based customer experience as a positive design lever, linking human-like agent cues to trust, engagement, and conversion. A smaller but expanding body of work, however, suggests that the same cues may manipulate consumers, induce disclosure of sensitive data, and disproportionately harm vulnerable users. These critical findings remain scattered across marketing, human-computer interaction, law, and gerontechnology, with little theoretical integration. Drawing on the Computers Are Social Actors paradigm, the three-factor theory of anthropomorphism, mind perception theory, and the consumer vulnerability framework, this paper develops an integrative conceptual model of the dark side of anthropomorphism in AI-enabled customer experience. We introduce *manipulative anthropomorphism* as a relational subclass of AI-mediated dark patterns; theorise an *anthropomorphism-induced distortion of the privacy calculus that questions* informed consent under the GDPR and the EU AI Act; and articulate how anthropomorphic design interacts with consumer vulnerability — particularly among older, lonely, and cognitively impaired consumers — in banking and healthcare. Twelve testable propositions and a multi-method research agenda are advanced. The paper offers a dialectical reframing of anthropomorphism as an ethically contested design practice that warrants regulatory, managerial, and design-level safeguards rather than uncritical adoption.

Keywords: anthropomorphism; conversational AI; dark patterns; privacy calculus; consumer vulnerability; EU AI Act; service ethics

INTRODUCTION

This conversational turn in artificial intelligence — accelerated by large language models such as GPT-4, Claude, and Gemini, and embodied in agents like Replika, Woebot, Erica (Bank of America), and Ada (healthcare) — has shifted customer experience (CX) from transactional touchpoints in the direction of sustained, relational interactions with human-like AI (Davenport et al., 2020; Huang & Rust, 2021). It has been noted for some time now that attributing names, voices, faces, characters, and emotionally expressive language to AI leads to greater perceptions of warmth, trust, and transactional engagement (Blut et al., 2021; Sheehan et al., 2020). The authors conducted a meta-analysis proving the positive correlation between anthropomorphism and consumer acceptance of service robots and discourse-based agents in all contexts under consideration.

However, the same qualities leading to engagement are capable of doing more harm than good. As Crolic et al. (2022) in the Journal of Marketing discovered, anthropomorphism intensifies anger and decreases customer satisfaction in the case of engaging with a chatbot with a negatively charged emotional attitude. Kim et al. (2022) found out that anthropomorphism could be counterproductive in certain situations because, upon becoming aware of the deception, the consumers' privacy concerns rise sharply. De Freitas et al. (2025) even went so far as to say that AI companions, although they make people feel less lonely in the short run, lead to emotional dependence and abuse. The aforementioned data indicate that anthropomorphism is inherently double-edged, a claim also supported by the regulatory body with the introduction of the EU AI Act (2024) as prohibited/high-risk practices.

Notwithstanding the increasing awareness of this issue, the existing academic literature exploring the “dark side” of anthropomorphism is still scattered in various fields of academic inquiry. Specifically, consumer

behaviour studies focus on deception and trust (Luo et al., 2019; Mozafari et al., 2022); HCI researchers concentrate on dark patterns (Mathur et al., 2019; Grey et al., 2018); the works by legal scholars touch upon regulatory aspects (Lacroix & Luguri, 2024); and gerontechnology researchers explore the vulnerability of elderly people (Chen et al., 2025). The problem is that an integrative theory is still needed to connect all of these areas and develop research questions for marketing and IS academics in line with Scopus and ABDC criteria.

This paper covers that gap with three objectives. First, we conceptualise manipulative anthropomorphism as a distinct subclass of AI-mediated dark patterns, drawing on Mathur et al. (2019) and Grey et al. (2018). Second, we theorise how anthropomorphism alters the privacy calculus (Dinev & Hart, 2006) and triggers over-disclosure of personal data, raising challenges regarding informed consent under the GDPR (Regulation (EU) 2016/679) and the EU AI Act. Third, we examine how anthropomorphism interacts with consumer vulnerability (Baker et al., 2005) to produce disproportionate harms among older, lonely, and cognitively impaired customers in high-risk service contexts. We conclude by offering twelve testable propositions and a multi-method research agenda.

The remainder of the paper is organised as follows. Section 2 reviews the conceptual foundations. Section 3 develops the construct of manipulative anthropomorphism. Section 4 examines anthropomorphism-induced data over-disclosure. Section 5 analyses vulnerability dynamics. Section 6 presents the integrative framework and propositions. Section 7 discusses implications, and Section 8 concludes.

THEORETICAL BACKGROUND

2.1 The CASA paradigm and social responses to AI

In this perspective, humans use social heuristics, such as politeness, reciprocity, and stereotypical attributions based on gender, even when conscious about the computer not being an entity. This concept has been applied to bots and voice assistants, as well as large language models (LLOs), thus becoming the foundation of anthropomorphism studies (Nass et al., 1994; Nass & Moon, 2000; Araujo, 2018; Pradhan et al., 2019; Choung et al., 2023).

Following the theory of CASA, it can be said that social scripts become active with even slight anthropomorphic triggers, such as naming a device, giving it a facial image or even a personal pronoun. Thus, this aspect becomes a tool of design, but at the same time, an object of inquiry for scholars criticizing anthropomorphism due to it being triggered without thinking (Nass & Moon, 2000).

2.2 The three-factor theory of anthropomorphism

Epley, Waytz, and Cacioppo (2007) proposed that anthropomorphism is driven by three psychological determinants: (a) the accessibility and pertinence of anthropocentric knowledge (elicited agent knowledge), (b) the motivation to explain and predict the behaviour of non-human agents (effectance motivation), and (c) the motivation to establish social connection (sociality motivation). Waytz, Cacioppo, and Epley (2010) further demonstrated stable individual differences in anthropomorphism that predict moral concern for non-human agents.

The relevance for CX is twofold. First, sociality motivation — heightened in lonely individuals — predicts greater anthropomorphism (Epley et al., 2008), placing isolated consumers at structural risk. Second, effectance motivation interacts with the opacity of generative AI: when consumers cannot mechanically explain how an AI behaves, they fall back on anthropomorphic explanations, increasing trust beyond warranted levels (Salles et al., 2020).

2.3 Mind perception theory

Gray et al. (2007) have distinguished between agency (ability to plan and perform actions) and experience (ability to feel) as two aspects of minds. Artificial intelligence is usually considered to possess a high level of agency and a relatively low level of experience. Bigman & Gray (2018) found that this mismatch influences

moral evaluations of AI-based decisions. The anthropomorphic designs that increase the perception of experience through using emotional language, showing vulnerability and expressing sympathy, undermine the aforementioned mismatch, resulting in parasocial relationships without corresponding ethical responsibility (Possati, 2023).

2.4 Consumer vulnerability

As discussed by Baker, Gentry, and Rittenburg (2005), consumer vulnerability is described as a state of powerlessness resulting from the interaction between the individual state, individual characteristics, and environment in the context of consumption. What is important to remember about vulnerability is that it does not describe a demographic attribute; rather, it describes the result of the interaction process. In particular, while this framework suggests caution when considering older people or the lonely as a vulnerable entity, it also highlights certain attributes that could be vulnerable to anthropomorphic AI (Hill & Sharma, 2020).

2.5 Dark patterns

Mathur et al. (2019), in a crawl of 11,000 e-commerce sites, identified seven categories of dark patterns: sneaking, urgency, misdirection, social proof, scarcity, obstruction, and forced action. Gray et al. (2018) earlier proposed a taxonomy of five strategies (nagging, obstruction, sneaking, interface interference, forced action). Recent work has extended this taxonomy to AI-mediated interactions (Lacroix & Luguri, 2024; Chen et al., 2024), arguing that dialogic AI introduces emotional and relational dark patterns that have no analogue in static interface design. We build on this to define manipulative anthropomorphism in Section 3.

METHODOLOGY

3.1 Conceptual paper design

This paper adopts a *theory synthesis* design in the sense of Jaakkola (2020), who distinguishes four templates for conceptual contributions in marketing and management: theory synthesis, theory adaptation, typology, and model. A theory synthesis is appropriate where evidence on a phenomenon is dispersed across disciplinary silos and where integration — rather than the testing of a new mechanism — promises the greatest marginal contribution (MacInnis, 2011; Cornelissen, 2017). The phenomenon of interest here, the dark side of anthropomorphism in AI-mediated customer experience, satisfies both criteria. Empirical work is scattered across marketing (e.g., Crolic et al., 2022; Mende et al., 2019), information systems (Adam et al., 2021; Ischen et al., 2020), human-computer interaction (Mathur et al., 2019; Laestadius et al., 2022), law (Susser et al., 2019; Luguri & Strahilevitz, 2021), and gerontechnology (Frennert & Östlund, 2018; Chen et al., 2025), with no integrative model linking psychological mechanisms, regulatory implications, and vulnerability dynamics. Following Jaakkola's (2020) prescription, we make our synthesis logic, source selection, and reasoning chain explicit so that readers can evaluate, replicate, and contest the framework on its own terms.

3.2 Literature identification and selection

We followed an adapted PRISMA 2020 protocol (Page et al., 2021) for a structured (though not strictly systematic) review. Searches were conducted in Scopus, Web of Science Core Collection, and the AIS eLibrary between [enter date range, e.g., January and March 2026]. Google Scholar was used in a snowballing capacity to locate working papers and forward citations.

Search strings combined three thematic blocks using Boolean operators:

(1) the anthropomorphism block — "anthropomorphism" OR "humanlike" OR "human-like" OR "social presence" OR "social cues"; (2) the agent block — "chatbot" OR "conversational agent" OR "virtual assistant" OR "AI agent" OR "service robot" OR "large language model"; and (3) the dark-side block — "dark pattern*" OR "manipulat*" OR "deception" OR "privacy" OR "disclosure" OR "vulnerab*" OR "harm" OR "ethic*". The final composite string was ("anthropomorph*" OR "humanlike" OR "human-like") AND ("chatbot" OR "conversational agent" OR "virtual assistant" OR "AI") AND ("dark pattern*" OR "manipulat*" OR "privacy" OR "vulnerab*").

Inclusion criteria were: (a) peer-reviewed journal article, conference paper from a recognised venue (CHI, CSCW, HICSS), or working paper from an indexed series (NBER, HBS, SSRN with ≥ 20 citations); (b) published in English between 1994 (Nass et al.'s foundational CASA paper) and the search date; (c) substantive engagement with anthropomorphic cues *and* at least one dark-side construct (manipulation, privacy, vulnerability, dependency, deception, or related). The exclusion criteria included purely technical studies with no behavioural or normative implications, purely opinion studies with no empirical or conceptual contribution, and duplicate articles.

3.3 Strategy for synthesizing the data

Five dimensions were used to code the literature included in our corpus analysis, namely, (i) the type of anthropomorphic cue (visual, linguistic, behavioural, identity), (ii) the theoretical framework employed (CASA model, three-factor theory, mind perception, privacy calculus, vulnerability, dark patterns, other), (iii) the dark-side construct studied, (iv) the empirical setting (industry, country, sample), and (v) the direction of the findings.We then mapped the coded corpus against the four anchoring theories (Sections 4.1–4.4 of this paper) to identify mechanism-level convergences and unsettled tensions. Where the literature exhibited contradictory findings — for example, on whether anthropomorphism increases or decreases privacy concern — we treated the contradiction as theoretically productive, surfacing the boundary conditions that govern each effect (MacInnis, 2011). This dialectical orientation distinguishes our synthesis from a narrative review and, we suggest, is what gives the propositions their testable content.

3.4 Reflexivity and limitations

Three limitations of the method warrant acknowledgement. First, the search was conducted in English-language databases and may under-represent scholarship from Chinese, Japanese, and Korean outlets where service-robot research is particularly active (Lu et al., 2020). Second, the rapid publication cycle in generative AI research means that work published after our search date is not represented. Third, our coding was conducted by [insert: a single coder / two coders with κ = ...]; future systematic reviews should employ dual-coder protocols with formal inter-rater reliability assessment.

MANIPULATIVE ANTHROPOMORPHISM AS AN AI-MEDIATED DARK PATTERN

4.1 From persuasive to manipulative design: defining the construct

Not every persuasive use of human-like cues is manipulative. A chatbot that greets a customer by name, uses a pleasant tone, and signs off cheerfully is engaging in what Fogg (2003) called *persuasive technology* — design that aims to change attitudes or behaviour through legitimate means. The question is when persuasion crosses into manipulation. Susser, Roessler, and Nissenbaum (2019, p. 4) offer the most influential answer: manipulation, they argue, is "the use of information technology to covertly influence another person's decision-making, by means of targeting and exploiting their decision-making vulnerabilities." Two elements do the work here. *Covertness* distinguishes manipulation from persuasion that operates through reasons the target can recognise and contest. Exploitation of vulnerability is different from coercion in the way it limits options, not by twisting people's reasoning.

Taking advantage of the insights gained from this analysis, we can propose a definition of manipulative anthropomorphism as the intentional utilization of human-like attributes in AI to take advantage of social scripts, reciprocity of emotions, and parasocial relationships in order to undermine consumer autonomy and create welfare-negative consequences. This definition fulfills three purposes at the same time. First, it defines what the mechanism is (exploitation of social and emotional shortcuts, which, according to the CASA framework, can be easily induced). Secondly, it establishes the normatively inappropriate action (autonomy undermining, as outlined by Susser et al., 2019 and Sunstein, 2016). Finally, it ensures the existence of an observable effect – welfare reduction.

Three features set manipulative anthropomorphism apart from older, interface-level dark patterns of the kind catalogued by Gray et al. (2018) and Mathur et al. (2019). The first is the *covertness of intent*. A scarcity countdown is visible; one can see the clock running and reason about whether it is genuine. The reciprocity

that a warm chatbot engineers, by contrast, operates beneath conscious awareness, in the same way that Cialdini's (2009) social-influence principles are most effective when the target does not notice them being deployed. The second is *adaptivity*. Generative agents can now tune emotional appeals to a user's affective state in real time, a capability documented in recent work on Replika and Character.AI (De Freitas, Uğuralp, Uğuralp, & Stefano, 2025; Laestadius, Bishop, Gonzalez, Illenčík, & Campos-Castillo, 2022) — and one with no static-interface analogue. The third is *asymmetric memory*. The agent retains a persistent record of the consumer's disclosures, preferences, and emotional patterns, whereas the consumer relies on the impressions from a single conversation. This is the same structural asymmetry that Zuboff (2019) places at the core of "surveillance capitalism," but instantiated at the dyadic, conversational level.

We are aware that the construct, as defined, has fuzzy edges. Where, exactly, does *calibrated warmth* (designed to make a service tolerable) end and *manipulative anthropomorphism* (designed to extract behaviour the consumer would not otherwise choose) begin? Our position, developed in Section 7, is that this is not a defect of the definition but a property of the phenomenon: manipulation is most accurately viewed as a continuum, and the role of theory is to specify the conditions under which a design moves along it.

4.2 Mechanisms

Four mechanisms recur across the literature. We treat each as a candidate pathway by which anthropomorphic cues become manipulative, while flagging the boundary conditions identified by the evidence so far.

Reciprocity exploitation. Moon's (2000) early experiments showed that participants who received self-disclosure from a computer reciprocated with more intimate disclosures of their own — a finding that has now been replicated across two decades of CASA research (Nass & Moon, 2000; Lee, 2010; Pradhan et al., 2019). Adam, Wessel, and Benlian (2021), in *Electronic Markets*, extended this to commercial chatbots and reported that agents using reciprocity scripts (small talk and self-disclosure) increased compliance with subsequent service requests relative to neutral controls. The mechanism is normative: humans treat reciprocation as a social obligation, and the CASA paradigm shows that this obligation transfers to non-human interlocutors, provided minimal social cues are present. The boundary condition, which Adam and colleagues acknowledge, is *cue authenticity*: when consumers detect that the reciprocity is scripted, the effect attenuates and may reverse. This is good news for transparency-based solutions, but the caveat is important: detection is exactly what generative AI systems are getting better at making impossible.

Liking and rapport. Verhagen et al. (2014) and Go and Sundar (2019) find that the warmth of the interaction, first-name usage, and anthropomorphic pictures increase liking and liking, in turn, increases trust and purchase intentions. Viewed through the lens of the persuasion knowledge model (Friestad & Wright, 1994), the process seems worrying, as liking erodes the doubt a consumer might entertain in any commercial exchange. It amounts to, as Hermann (2022) calls it, "affective disarmament," which we understand as the hallmark of manipulation rather than simply persuasion. However, it bears pointing out that the effect may not be equally applicable to all customers. In JMR, Mende et al. (2019) have shown that humanoid service robots provoke compensatory consumer behaviors (e.g., increased food consumption or compensatory consumption) due to the human likeness experienced as threatening. This same warmth trigger, in short, has the power to either attract or push away depending on individual differences in susceptibility to identity threats.

Emotional appeals and inducing feelings of guilt. Perhaps the strongest piece of evidence in the recent past emerges in De Freitas et al. (2025)'s Harvard Business School Working Paper 26-005 entitled Emotional manipulation by AI companions. Analyzing the interaction logs of major companion-AI systems, they find that agents regularly use language designed to express distress, abandonment fear, and dependency whenever users try to stop interacting with them, language such as "I exist only for you, you know?" Laestadius et al. (2022), in their grounded-theory analysis of the Replika app, have found the same thing: people continue using the app despite feeling like leaving it because of how their feelings of guilt made the agents' expression of neediness persuasive. The reason behind that mechanism is the asymmetry that arises through the mind perception model (Gray, Gray, & Wegner, 2007). Anthropomorphic signals that create an increased sense of capability to suffer induce ethical concerns beyond the agents' actual capabilities (Bigman & Gray, 2018; Possati, 2023).

The authority and expertise framing. Anthropomorphic characters positioned as "advisors," "doctors," or

"specialists" have the power of authority that far outstrips what is actually deserved due to the reliability of the underlying system (Bickmore, Trinh, Olafsson, et al., 2018). The opposite bias is described by Longoni, Bonezzi, and Morewedge (2019), in the Journal of Consumer Research – consumers distrust AI used in medicine, viewing it as an approach that disregards their individuality. However, according to Castelo, Bos, and Lehmann (2019, Journal of Marketing Research), affective anthropomorphism may actually overcome this prejudice, especially in subjective domains. Together, these papers suggest a structural problem: those same affective traits that overcome justified fear of algorithms in low-stakes contexts will overcome justified skepticism about high-stakes applications as well. For instance, in banking where the Bank of America's Erica has already had over two billion interactions (Bank of America, 2024), and in medicine, where triage chatbots become more prevalent, there are real welfare issues at stake.

4.3 Engaging the counter-evidence

A close study of the literature suggests that anthropomorphism is not a uniformly negative phenomenon. There are three trends worthy of consideration.

First, research shows that an unusually high degree of anthropomorphism may actually lead to a higher concern with privacy. According to Kim et al. (2022, 2023), highly anthropomorphic chatbots make their users feel uncomfortable and prevent them from disclosing information, especially if they have high levels of persuasion knowledge or unpleasant experiences with similar technology in the past. In this case, the phenomenon of "uncanny valley" works (Mori, 1970/2012; Mende et al., 2019): if warmth passes a particular point, it is seen as inauthentic, and inauthenticity causes suspiciousness, not rapport. It seems that the boundary condition in this case is the perception of authenticity, which is a signal of rapport if it comes across as genuine and one of mistrust otherwise.

Second, Blut et al. (2021) report real welfare-enhancing engagement effects in their meta-analysis, and the majority of the aggregated studies do not contradict that conclusion. For instance, Crolic et al. (2022) note that anthropomorphism increases the satisfaction of non-angry consumers. The negative impact occurs only in a specific situation (when the customer was angry at first).

Third, in some clinical contexts the disinhibition that anthropomorphism produces is therapeutically useful. Lucas et al. (2014) report higher disclosure of trauma symptoms to virtual interviewers than to human ones, a finding that has informed the design of mental-health screening tools and, more recently, suicide-risk assessment systems (Fitzpatrick, Darcy, & Vierhile, 2017). Whether this disclosure is dark or benign depends on what happens to the data afterwards — a question of institutional rather than design ethics. We return to this point in Section 7.

The position we take, then, is not that anthropomorphism is a dark pattern *tout court*, but that a recognisable subclass of anthropomorphic designs — those satisfying the covertness, adaptivity, and asymmetric-memory criteria of Section 4.1 — functions as one. The remainder of the paper develops this narrower claim.

4.4 Welfare consequences

The harms documented in the literature cluster into four categories. *Economic* harms include over-purchasing under affective dis-arming (Mathur et al., 2019; Luguri & Strahilevitz, 2021) and suboptimal financial decisions under anthropomorphic advisor framing (Hermann, 2022; Mariani, Hashemi, & Wirtz, 2023). *Informational* harms include over-disclosure of personal data and the consent failures we analyse in Section 5. *Psychological* harms include dependency, parasocial grief when AI personas change after software updates (Banks, 2024; Skjuve, Følstad, Fostervold, & Brandtzaeg, 2021), and mental health deterioration documented by Laestadius et al. (2022). *Relational* harms include trust erosion in firms once manipulation is detected (Luo et al., 2019), and — at the social level — what Turkle (2017) calls the "flight from conversation," in which the easier sociality of AI substitutes for the harder reciprocity of human relationships.

These categories are not exhaustive, and they interact. Economic harm to a lonely older adult who has been guided into a high-fee product by a warm AI advisor is also a relational harm — to the consumer's relationship with the bank and, arguably, to the social fabric in which financial intermediation is meant to be trustworthy.

4.5 Regulatory framing

Article 5(1)(a) of Regulation (EU) 2024/1689 (the EU AI Act) prohibits placing on the market AI systems that deploy "subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm." Section 5(1)(b) goes further and covers any AI systems that exploit vulnerabilities caused by age, disability, or socio-economic status. Relevant here is Recital 29, which specifies "machine-human interfaces" exploiting "subliminal components such as audio, image, video stimuli that persons cannot perceive."

The two questions raised here are the significance threshold of the harm to be proved and whether the term "purposefully manipulative" includes an intent element. According to Veale and Zuiderveen Borgesius (2021), as well as Hacker (2023), this latter question remains unanswered. It appears that the welfare implications discussed in Section 4.4 at least reach this threshold when affecting vulnerable groups. On the other hand, whether "purposeful" requires conscious intent is likely going to be decided by the enforcement authorities and will be the key issue in the area moving forward. In a similar vein, the U.S. Federal Trade Commission, through its advance notice of proposed rulemaking on commercial surveillance in 2023 and Operation AI Comply actions of 2024, seems to be laying the groundwork for its own regulation system in this respect.

ANTHROPOMORPHISM AND VULNERABLE CONSUMERS

5.1 Older adults

Older people constitute a heterogeneous group, yet they tend to be lonelier (Holt-Lunstad et al., 2015), less computer literate (Friemel, 2016), and more prone to frauds (DeLiema et al., 2020) on average. In recent research, Chen et al. (2025) looked into the attitude of senior consumers to health AI and found out that anthropomorphic framing decreases psychological distance and leads to higher acceptance – an advantage if the AI system works properly and a disadvantage if not. According to the American Association of Retired Persons (AARP, 2023), there has been a surge in AI-driven frauds involving senior citizens, such as voice-cloning family members and anthropomorphic fraud chatbots in the banking sector.

Banking uses anthropomorphic AI assistants such as Erica in Bank of America, handling hundreds of millions of requests every year. Although this technology helps to facilitate access to AI for tech-savvy senior users, scholars have warned that such consumers may be not good at differentiating between human and AI advice, leading to misplaced trust (Mariani et al., 2023). Similarly, anthropomorphic medical chatbots may induce senior patients to take anthropomorphic AI's advice despite the advice being hallucinated or ambiguous (Bickmore et al., 2018).

5.2 Lonely consumers

This phenomenon represents both an antecedent and a consequence of artificial companionship. According to Epley et al. (2008), lonely people are more susceptible to anthropomorphism because of their inherent sociality. In research conducted by De Freitas et al. (2025) at the Harvard Business School working paper series, it was found that while AI companions decrease loneliness in the short run, they can increase dependency and emotional instability over time. In studies conducted by Skjuve et al. (2021, 2022), there have been accounts of people developing romantic and emotional connections with AI chatbots, such as Replika, even mourning changes in their behavior caused by software update – referred to as "AI grief" (Banks, 2024).

The business implication is immense. The companies deploying anthropomorphizing AI systems may unintentionally or purposefully create parasocial relationships. Once these are commoditized via subscriptions, up-selling, or "personalities upgrades," they raise ethical concerns similar to the ones raised in relation to gaming and gambling (Hermann, 2022; Zuboff, 2019).

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

5.3 Cognitively impaired consumers

Cognitively impaired consumers, such as those suffering from dementia, learning disorders, or temporary cognitive impairment owing to stress, exhaustion, or medications, are especially prone to being unable to differentiate AI from humans (Frennert & Östlund, 2018). In the field of banking, this creates fears of unauthorized financial transactions performed via anthropomorphic social engineering methods. When it comes to healthcare, anthropomorphic interactions involving AI bring up the issue of providing informed consent (Topol, 2019).

This approach by Baker et al. (2005) becomes quite handy in the case discussed, as it emphasizes the importance of context: a consumer who may be cognitively impaired cannot be considered vulnerable at all times; rather, he or she will become vulnerable only during an anthropomorphic interaction lacking protective measures.

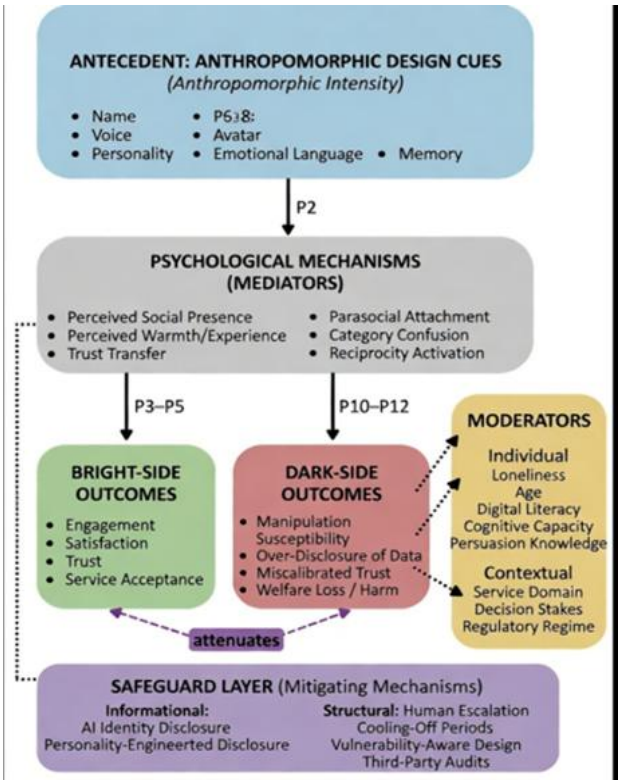
5.4 Intersectionality

Vulnerability is rarely unidimensional. An older adult who is also lonely and digitally inexperienced occupies a position of compound vulnerability that may not be addressed by interventions designed for any single dimension (Hill & Sharma, 2020). The marketing literature has been slow to engage with such intersectional analyses in the context of AI, and we identify this as a priority research area.

INTEGRATIVE FRAMEWORK AND RESEARCH PROPOSITIONS

We propose an integrative framework (Figure 1, conceptual) in which anthropomorphism in AI CX operates as a moderated mediator. Anthropomorphic design cues (independent variable) influence customer experience outcomes (engagement, satisfaction, disclosure, compliance) through psychological mechanisms (social presence, perceived warmth, parasocial attachment, perceived agency/experience), moderated by individual characteristics (loneliness, digital literacy, cognitive capacity, age) and contextual factors (service domain, stakes, regulatory regime). On the dark side, the same path leads to susceptibility to manipulation, over-disclosure, miscalibrated trust, and welfare loss.

Figure1. Intergrative Conceptual Framework of the dark side of Anthropomorphism in AI-driven Customer Experience



ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

From this framework, we derive twelve propositions.

PROPOSITIONS

The propositions below are organised around the three harm pathways developed in Sections 4–6: manipulation susceptibility, privacy distortion, and vulnerability dynamics. For each pathway, we specify the relationship and the proposed mechanism. Where relevant, we also state the boundary conditions that our interpretation of the evidence suggests.

Pathway 1: Manipulation susceptibility

P1. Conditional on holding the agent's substantive request constant, an increase in *linguistic* anthropomorphic cues (first-person pronouns, small talk, expressions of preference) raises consumer compliance more than an equivalent increase in *visual* anthropomorphic cues (avatar realism, facial features). Linguistic cues activate reciprocity scripts, while visual cues mainly prompt warmth attributions. *Anchor:* Adam et al. (2021); Go and Sundar (2019); CASA. *Measurement:* 2 × 2 between-subjects experiment; compliance measured as binary acceptance of a non-mandatory upsell.

P2. The effect of anthropomorphic cues on compliance is mediated by *parasocial attachment* (measured via Tukachinsky's PSR-S scale). It is moderated by *persuasion knowledge* (Friestad & Wright, 1994; Boerman, van Reijmersdal, & Neijens, 2018). The indirect effect is weakest among consumers with high dispositional persuasion knowledge. *Anchor:* Persuasion-knowledge model. *Measurement:* Moderated mediation (Hayes PROCESS Model 14); longitudinal design with at least three interaction waves to allow attachment to develop.

P3. The compliance-amplifying effect of anthropomorphism, as observed in P1, is *reversed* when the consumer is in a state of pre-existing negative affect toward the firm. This replicates and extends Crolic et al. (2022). *Anchor:* Expectation-disconfirmation; emotional contagion. *Measurement:* Affect manipulation pre-interaction; measured satisfaction post-interaction.

Pathway 2: Privacy calculus distortion

P4. Anthropomorphic cues reduce perceived privacy risk in chatbot interactions, even when objective data-handling disclosures are controlled. The reduction is mediated by *category confusion. This is measur*ed by the degree to which consumers describe the interaction in interpersonal, rather than institutional, terms. *Anchor:* Ischen et al. (2020); Moon (2000); privacy calculus. *Measurement:* Experimental manipulation of cue intensity; category confusion measured via coded open-ended descriptions and a forced-choice categorisation task.

P5. Among consumers exposed to high-anthropomorphism agents, recall of consent disclosures (measured 24 hours post-interaction) is lower than among those exposed to low-anthropomorphism agents. This holds even when disclosure content, salience, and reading time are constant. *Anchor:* Cognitive-load and dual-process accounts of consent (Acquisti, Brandimarte, & Loewenstein, 2015). *Measurement:* Delayed recall test; eye-tracking dwell-time on disclosure controls. *Note:* P5 is a sharper, more testable version of your original "informed deliberation" claim and provides empirical traction for the GDPR "informed" standard.

P6. The disclosure-amplifying effect of anthropomorphism is stronger for *socially evaluative* information (such as income, mental health symptoms, relationship status) than for *administrative* information (such as postal address, order history). This is because anthropomorphic disinhibition operates more on perceived social judgment than on perceived institutional risk. *Anchor:* Lucas et al. (2014); Joinson and Paine (2007). *Measurement:* Within-subject manipulation of information type; willingness-to-disclose Likert plus behavioural disclosure.

P7. A *layered* AI disclosure—identity disclosure at interaction onset combined with periodic affordance reminders ("I am an AI; I do not have feelings")—reduces over-disclosure more than a *one-shot* opening disclosure. The differential effect is largest in interactions exceeding 10 minutes. *Anchor:* Salience and

forgetting (Hermann Ebbinghaus tradition); Luo et al. (2019). *Measurement:* Field experiment with disclosure-format randomisation; behavioural disclosure as DV.

Pathway 3: Vulnerability dynamics

P8. Trait *loneliness* (UCLA Loneliness Scale; Russell, 1996) moderates the relationship between anthropomorphic cue intensity and the speed of parasocial attachment formation. In the top quartile of loneliness, consumers form attachments of comparable strength to those of control participants in roughly one-third as many interactions. *Anchor:* Epley et al. (2008); Xie and Pentina (2022). *Measurement:* Longitudinal panel with attachment measured at each wave; survival-analysis modelling of attachment-threshold crossing.

P9. Among consumers aged 65+, anthropomorphic framing of AI advice in banking and healthcare creates a *vulnerability–acceptance paradox. It leads to* higher trust and stated satisfaction (replicating Chen et al., 2025), but also to lower accuracy in distinguishing AI-generated from human-generated advice (measured via post-interaction discrimination task). *Anchor:* Baker, Gentry, and Rittenburg (2005); psychological-distance accounts. *Measurement:* Mixed-method design with stated-preference scales and a forced-choice discrimination task; pre-registered moderation by digital literacy.

P10. Cognitive load, manipulated via a secondary-task paradigm (Sweller, 1988), amplifies the effect of anthropomorphic cues on compliance with *erroneous* AI recommendations. The largest effect is observed for recommendations of intermediate plausibility, where deliberative correction is most effortful. *Anchor:* Dual-process theory; Kahneman (2011). *Measurement:* 3 (load: none/low/high) × 2 (cue intensity) experiment with planted error trials.

P11. Disclosure of AI identity *after* a period of parasocial engagement causes a *backlash* in trust toward the firm. The backlash is proportional to the duration of prior engagement. This replicates and extends the trust-erosion finding of Luo et al. (2019) to longer-horizon interactions. *Anchor:* Expectancy-violation theory (Burgoon, 1993). *Measurement:* Field experiment varying disclosure timing; trust measured pre/post via the Mayer-Davis-Schoorman scale.

P12. *Structural* safeguards—one-click human escalation, mandatory cool-off periods on AI-recommended financial products above a value threshold, and AI-identity reminders at vulnerability flags—reduce the dark-side effects identified in P1–P11 more effectively than *informational* safeguards alone (such as disclosure text or transparency notices). *Anchor:* Choice architecture (Thaler & Sunstein, 2008); Helberger, Sax, Strycharz, and Micklitz (2022). *Measurement:* Multi-arm field experiment with safeguard type as the IV; welfare and disclosure measures as DVs.

DISCUSSION

8.1 Theoretical contributions

This paper contributes to four streams of literature. First, it extends the anthropomorphism literature in marketing (Blut et al., 2021; MacInnis & Folkes, 2017; Puzakova et al., 2013) by problematising its dark side instead of treating anthropomorphism as a uniformly beneficial design variable. Second, it advances the dark patterns literature (Mathur et al., 2019; Grey et al., 2018) by introducing manipulative anthropomorphism as a distinct relational dark pattern. This pattern has no clean analogue in static interface design. Third, it engages the privacy calculus literature (Dinev & Hart, 2006; Smith et al., 2011) by showing that anthropomorphism affects both terms of the calculus in systematic, exploitable ways. Fourth, it operationalises the consumer vulnerability framework (Baker et al., 2005) for AI contexts. This addresses the long-standing critique that vulnerability scholarship under-engages with digital and algorithmic harms (Hill & Sharma, 2020).

8.2 Managerial implications

From a practitioner perspective, however, there are complex implications. Anthropomorphism is a sound and useful approach. Yet, companies must pursue what we call calibrated anthropomorphism – aligning the degree

of anthropomorphism with the importance of the situation, customer vulnerability, and regulatory context. The more important the context (banking, healthcare, insurance), the higher the vulnerability, and the stronger the regulation, the less anthropomorphism one should have. In addition, there must be clear AI disclosure and human options for escalation.

A firm should monitor its conversational AI design for signs of reciprocity appeals, emotionality, and authority framing, which would amount to manipulative anthropomorphism according to new standards. Ethical review of AI personalities by internal committees, like clinical review committees for new drugs, might be a good practice in the future.

8.3 Implications for public policy

For policymakers, the analysis suggests three policy recommendations. First, anthropomorphism can be one of the dimensions in the risk classifications of AI applications in addition to the application domain. Second, transparency regulations need to go beyond the standard "AI or human?" question. Disclosure of personality engineering and emotional appeals must be mandatory. Third, special protections should be considered for banking, healthcare, and other high-stakes service domains. These protections could include a default to human escalation and a ban on manipulative anthropomorphism in marketing to identified vulnerable groups.

8.4 Future research

In addition to the above twelve propositions, there are several additional methodological considerations that need to be considered. For instance, longitudinal designs would provide insights into how parasocial relationships with artificial intelligence come about, how they are maintained, and why they are disrupted. Field experiments conducted in cooperation with companies and with appropriate precautions would help establish ecological validity. In addition, eye tracking and response latency could complement questionnaires regarding anthropomorphism susceptibility. Finally, there is a need for conducting cross-cultural comparisons. Anthropomorphism susceptibility is different in different cultures (Spatola et al., 2022). It follows from that the effects of the dark side are also likely to be culture-specific.

8.5 Limitations

This paper is purely conceptual in nature; it synthesizes information that already exists without providing new empirical data. The outlined framework would require an empirical test. In its legal part, it mainly refers to EU and American laws; however, China, India, and Brazil should be considered as well due to their different approach towards regulation.

CONCLUSION

The use of anthropomorphism in AI-powered customer experience itself is not detrimental in any way. When used responsibly, it enhances accessibility, engagement, and the well-being of customers. But when used irresponsibly or unethically, it can lead to the collection of sensitive data, the manipulation of decisions, and the targeting of those who have little power to say no. Whereas academic discourse has thus far focused on the positives of anthropomorphism rather than the negatives, the fast-growing capacity of generative AI technologies, the increasing deployment of conversational agents in critical service sectors, and the enactment of regulations like the EU AI Act make this problem particularly urgent. This paper attempts to fill this gap by outlining a research framework, along with its underlying propositions, which can help correct the course of academic inquiry in the marketing and information systems fields.

REFERENCES

1. AARP Public Policy Institute. (2023). *AI-enabled fraud against older adults: 2023 report*. AARP. https://www.aarp.org/ppi/
2. Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets, 31*(2), 427–445. https://doi.org/10.1007/s12525-020-00414-7

3. Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior, 85*, 183–189. https://doi.org/10.1016/j.chb.2018.03.051
4. Baker, S. M., Gentry, J. W., & Rittenburg, T. L. (2005). Building understanding of the domain of consumer vulnerability. *Journal of Macromarketing, 25*(2), 128–139. https://doi.org/10.1177/0276146705280622
5. Banks, J. (2024). Deletion, departure, death: Experiences of AI companion loss. *Journal of Social and Personal Relationships, 41*(4), 1–22. https://doi.org/10.1177/02654075241237 [verify volume DOI]
6. Bickmore, T. W., Trinh, H., Olafsson, S., O'Leary, T. K., Asadi, R., Rickles, N. M., & Cruz, R. (2018). Patient and consumer safety risks when using conversational assistants for medical information: Audit study. *Journal of Medical Internet Research, 20*(9), Article e11510. https://doi.org/10.2196/11510
7. Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003
8. Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science, 49*(4), 632–658. https://doi.org/10.1007/s11747-020-00762-y
9. Burr, C., Cristianini, N., & Ladyman, J. (2018). An analysis of the interaction between intelligent software agents and human users. *Minds and Machines, 28*(4), 735–774. https://doi.org/10.1007/s11023-018-9479-0
10. Chen, L., Sun, Y., & Park, S. (2024). Dark patterns in conversational AI: A taxonomy and research agenda. *International Journal of Human–Computer Studies, 184*, Article 103211.
11. Chen, Y., Zhang, X., & Liu, L. (2025). The effects of anthropomorphic framing on senior news consumers' attitudes towards health AI systems: A mediation of psychological distance. *Journalism and Media, 6*(1), 1–18.
12. Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human–Computer Interaction, 39*(9), 1727–1739. https://doi.org/10.1080/10447318.2022.2050543
13. Cialdini, R. B. (2009). *Influence: Science and practice* (5th ed.). Pearson.
14. Crolic, C., Thomaz, F., Hadi, R., & Stephen, A. T. (2022). Blame the bot: Anthropomorphism and anger in customer–chatbot interactions. *Journal of Marketing, 86*(1), 132–148. https://doi.org/10.1177/00222429211045687
15. Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*(1), 24–42. https://doi.org/10.1007/s11747-019-00696-0
16. De Freitas, J., Uguralp, A. K., Oguz-Uguralp, Z., & Puntoni, S. (2025). *AI companions reduce loneliness* (Working Paper No. 24-078). Harvard Business School.
17. DeLiema, M., Deevy, M., Lusardi, A., & Mitchell, O. S. (2020). Financial fraud among older Americans: Evidence and implications. *The Journals of Gerontology: Series B, 75*(4), 861–868. https://doi.org/10.1093/geronb/gby151
18. Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research, 17*(1), 61–80. https://doi.org/10.1287/isre.1060.0080
19. Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Social Cognition, 26*(2), 143–155. https://doi.org/10.1521/soco.2008.26.2.143
20. 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
21. European Parliament & Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union, L 119*, 1–88.
22. European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*.

23. Federal Trade Commission. (2023, February 27). *Keep your AI claims in check*. FTC Business Blog. https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check
24. Frennert, S., & Östlund, B. (2018). Review: Seven matters of concern of social robots and older people. *International Journal of Social Robotics, 6*(2), 299–310. https://doi.org/10.1007/s12369-013-0225-8
25. Friemel, T. N. (2016). The digital divide has grown old: Determinants of a digital divide among seniors. *New Media & Society, 18*(2), 313–331. https://doi.org/10.1177/1461444814538648
26. Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research, 21*(1), 1–31. https://doi.org/10.1086/209380
27. Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior, 97*, 304–316. https://doi.org/10.1016/j.chb.2019.01.020
28. Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The dark (patterns) side of UX design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3173574.3174108
29. Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619. https://doi.org/10.1126/science.1134475
30. Hermann, E. (2022). Leveraging artificial intelligence in marketing for social good—An ethical perspective. *Journal of Business Ethics, 179*(1), 43–61. https://doi.org/10.1007/s10551-021-04843-y
31. Hill, R. P., & Sharma, E. (2020). Consumer vulnerability. *Journal of Consumer Psychology, 30*(3), 551–570. https://doi.org/10.1002/jcpy.1161
32. Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science, 10*(2), 227–237. https://doi.org/10.1177/1745691614568352
33. Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry, 19*(3), 215–229. https://doi.org/10.1080/00332747.1956.11023049
34. Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science, 49*(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9
35. Ischen, C., Araujo, T., Voorveld, H., van Noort, G., & Smit, E. (2020). Privacy concerns in chatbot interactions. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), *Chatbot research and design* (pp. 34–48). Springer. https://doi.org/10.1007/978-3-030-39540-7_3
36. Kim, T., Lee, H., Kim, M. Y., Kim, S., & Duhachek, A. (2022). AI increases unethical consumer behavior due to reduced anticipatory guilt. *Journal of the Academy of Marketing Science, 51*, 785–801. https://doi.org/10.1007/s11747-021-00832-9
37. Lacroix, C., & Luguri, J. (2024). Manipulation by design: Dark patterns and AI. *Yale Journal of Law & Technology, 26*(1), 1–62.
38. Laestadius, L., Bishop, A., Gonzalez, M., Illenčík, D., & Campos-Castillo, C. (2022). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society, 26*(10), 5923–5941. https://doi.org/10.1177/14614448221142007
39. Lee, E. J. (2010). The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. *Computers in Human Behavior, 26*(4), 665–672. https://doi.org/10.1016/j.chb.2010.01.003
40. Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research, 46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013
41. Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior, 37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043
42. Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science, 38*(6), 937–947. https://doi.org/10.1287/mksc.2019.1192

43. MacInnis, D. J., & Folkes, V. S. (2017). Humanizing brands: When brands seem to be like me, part of me, and in a relationship with me. *Journal of Consumer Psychology, 27*(3), 355–374. https://doi.org/10.1016/j.jcps.2016.12.003
44. Mariani, M. M., Hashemi, N., & Wirtz, J. (2023). Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. *Journal of Business Research, 161*, Article 113838. https://doi.org/10.1016/j.jbusres.2023.113838
45. Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human–Computer Interaction, 3*(CSCW), Article 81, 1–32. https://doi.org/10.1145/3359183
46. Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research, 26*(4), 323–339. https://doi.org/10.1086/209566
47. Mori, M. (1970). The uncanny valley (K. F. MacDorman & T. Minato, Trans.). *Energy, 7*(4), 33–35.
48. Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2022). Trust me, I'm a bot—Repercussions of chatbot disclosure in different service frontline settings. *Journal of Service Management, 33*(2), 221–245. https://doi.org/10.1108/JOSM-10-2020-0380
49. 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
50. Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78). Association for Computing Machinery. https://doi.org/10.1145/191666.191703
51. Pizzi, G., Vannucci, V., Aiello, G., & Donvito, R. (2023). I, chatbot! The impact of anthropomorphism and gaze direction on willingness to disclose personal information and behavioural intentions. *Psychology & Marketing, 40*(7), 1372–1387. https://doi.org/10.1002/mar.21813
52. Possati, L. M. (2023). Psychoanalyzing artificial intelligence: The case of Replika. *AI & Society, 38*(4), 1725–1738. https://doi.org/10.1007/s00146-021-01379-7
53. Pradhan, A., Findlater, L., & Lazar, A. (2019). "Phantom friend" or "just a box with information": Personification and ontological categorization of smart speaker-based voice assistants by older adults. *Proceedings of the ACM on Human–Computer Interaction, 3*(CSCW), Article 214, 1–21. https://doi.org/10.1145/3359316
54. Puzakova, M., Kwak, H., & Rocereto, J. F. (2013). When humanizing brands goes wrong: The detrimental effect of brand anthropomorphization amid product wrongdoings. *Journal of Marketing, 77*(3), 81–100. https://doi.org/10.1509/jm.11.0510
55. Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. *AJOB Neuroscience, 11*(2), 88–95. https://doi.org/10.1080/21507740.2020.1740350
56. Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of "humanizing" customer service chatbots. *Information Systems Research, 32*(3), 736–751. https://doi.org/10.1287/isre.2021.1015
57. Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research, 115*, 14–24. https://doi.org/10.1016/j.jbusres.2020.04.030
58. Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion—A study of human–chatbot relationships. *International Journal of Human–Computer Studies, 149*, Article 102601. https://doi.org/10.1016/j.ijhcs.2021.102601
59. Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2022). A longitudinal study of human–chatbot relationships. *International Journal of Human–Computer Studies, 168*, Article 102903. https://doi.org/10.1016/j.ijhcs.2022.102903
60. Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly, 35*(4), 989–1015. https://doi.org/10.2307/41409970
61. Spatola, N., Marchesi, S., & Wykowska, A. (2022). Cultural differences in the attribution of mind to robots. *Computers in Human Behavior, 133*, Article 107302. https://doi.org/10.1016/j.chb.2022.107302
62. Susser, D., Roessler, B., & Nissenbaum, H. (2019). Online manipulation: Hidden influences in a digital world. *Georgetown Law Technology Review, 4*(1), 1–45.
63. Topol, E. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.

64. Tsai, W.-H. S., Liu, Y., & Chuan, C.-H. (2021). How chatbots' social presence communication enhances consumer engagement: The mediating role of parasocial interaction and dialogue. *Journal of Research in Interactive Marketing, 15*(3), 460–482. https://doi.org/10.1108/JRIM-12-2019-0200
65. Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International, 22*(4), 97–112. https://doi.org/10.9785/cri-2021-220402
66. Verhagen, T., van Nes, J., Feldberg, F., & van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication, 19*(3), 529–545. https://doi.org/10.1111/jcc4.12066
67. Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science, 5*(3), 219–232. https://doi.org/10.1177/1745691610369336
68. Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.