Communication Theory, 2024, **00**, 1–14
https://doi.org/10.1093/ct/qtae016
Original Article

International Communication Association
OXFORD

Visioning a two-level human–machine communication framework: initiating conversations between explainable AI and communication

Kun Xu[1], **and Jingyuan Shi**[2,*]

[1]College of Journalism and Communications, University of Florida, Gainesville, FL, USA
[2]Department of Interactive Media, Hong Kong Baptist University, Hong Kong SAR, China
*Corresponding author: Jingyuan Shi. Email: jolieshi@hkbu.edu.hk

Abstract

Amid mounting interest in artificial intelligence (AI) technology, communication scholars have sought to understand humans' perceptions of and attitudes toward AI's predictions, recommendations, and decisions. Meanwhile, scholars in the nascent but growing field of explainable AI (XAI) have aimed to clarify AI's operational mechanisms and make them interpretable, visible, and transparent. In this conceptual article, we suggest that a conversation between human–machine communication (HMC) and XAI is advantageous and necessary. Following the introduction of these two areas, we demonstrate how research on XAI can inform the HMC scholarship regarding the human-in-the-loop approach and the message production explainability. Next, we expound upon how communication scholars' focuses on message sources, receivers, features, and effects can reciprocally benefit XAI research. At its core, this article proposes a two-level HMC framework and posits that bridging the two fields can guide future AI research and development.

Keywords: artificial intelligence, human–machine communication, explainable AI, human-in-the-loop approach, human–AI interaction

Communication researchers have been increasingly intrigued by the opportunities and challenges brought by artificial intelligence (AI) technology. Ranging from Twitter's withdrawal of AI bot Tay due to its assimilation of online racist and misogynistic comments (Vincent, 2016) and Facebook's discovery of its own AI engines' deviation from predesigned conversation scripts (Bradley, 2017), to scholars' attempts to combine manual coding and generative pretrained transformers (GPT) in qualitative analyses (Xiao et al., 2023) and the use of large language models (LLMs) to produce software agents that mimic human behaviors (Park et al., 2023), understanding both social implications and psychological effects of AI has become a major focus of communication research. Although human–machine communication (HMC), a rising and expanding field in communication, has fostered an in-depth understanding of the relationships between humans and machines that serve as digital interlocutors (e.g., chatbots, robots, AI agents) (Guzman et al., 2023), the rapid advance and updating of AI technology still leaves users limited time and resources to fully understand AI's working mechanisms and its impact. While communication researchers have growingly stressed the importance of *how AI can be communicative* (Gunkel, 2012; Guzman & Lewis, 2020), the question of *how AI can be communicated* remains relatively understudied. Against this backdrop, scholars in computer science, engineering, and information science have started a new research realm, explainable AI (XAI), to understand the factors that render AI systems comprehensible to humans (Liao et al., 2020). By making AI's decision-making process transparent and interpretable, XAI seeks to build up users' trust and understanding in AI (Ehsan et al., 2021).

Because XAI is still in its infancy, limited communication research has referred to this field to investigate the effects of unpacking the backstage mechanisms of AI. Yet, as users' demand for understanding AI's internal working grows, communication research could benefit from some of the inquiries in XAI research. For example, whereas much HMC research has investigated the similarities and differences between human performances and machine performances in various contexts (e.g., content moderation, news reporting, persuasion, decision-making) along the dimensions of perceived trustworthiness, competency, social presence, attractiveness, and the potential to gain compliance or behavioral conformity (e.g., Edwards et al., 2019; Spence et al., 2014; Spence et al., 2019a; Xu et al., 2020), calls for a more systematic understanding of the specific features or characteristics of AI that lead to its homogeneity with or distinction from human–human communication are growing (Liu et al., 2023). Hence, unpacking the black box of algorithms, as what XAI researchers are focusing on, could benefit HMC research, as it involves elucidating the generation of AI's recommendations, predictions, and decisions and understanding how humans are engaged in multiple stages of the algorithmic decision-making processes. Through such explorations, XAI can proffer new theories and concepts to enrich our understanding of AI in future HMC research.

Meanwhile, communication research, especially its recent development in HMC, can enlarge the scope of XAI research. HMC research, along with other directions, including computer-mediated communication (CMC), persuasion, and information processing is abundant with findings about message exchange, message design, and message effects. These aspects could foster XAI researchers' efforts to incorporate human-centric explanations, articulate data sources, and use different techniques to enhance the visibility and interpretability of AI systems. As Lai

Received: 22 October 2023. **Revised:** 9 July 2024. **Accepted:** 14 July 2024

et al. (2023) indicated, "current XAI paradigms deal only with the reasoning process and concern little with the communication process" (p. 357). Designing and promoting human comprehensible explanations can thus enable AI users to make informed and empowered decisions (Wolf & Ringland, 2020). Therefore, like the benefits of introducing XAI to HMC research, communication scholars can make pivotal contributions to XAI in that research on explanations can directly benefit from the communication scholarship on the message exchange process between users and machines.

Taking these two emerging fields together, this article suggests that XAI and communication research can mutually benefit each other, such that concepts and theories of XAI can lead to more in-depth exploration of the factors underlying human understanding and acceptance of AI. Meanwhile, communication literature on meaning-making and message design can provide a fruitful starting point to further XAI research.

This article unfolds with three major sections. In the first section, we respectively introduce the rise of XAI and HMC and demonstrate why XAI and HMC could mutually benefit each other. In the second section, we list six directions where the insights from XAI and HMC can enrich each other. These directions include how the human-in-the-loop approach and the message production explainability can be useful to future HMC research. They also include how communication research, including HMC research, on sources, messages, receivers, and effects can guide future XAI research. In the final section, we conclude by proposing a two-level HMC framework, which afford theoretical, ethical, and practical implications for future research at the intersection of communication and XAI.

Explainable AI

AI is defined as "a science and a set of computational technologies that are inspired by—but typically operate quite differently from—the ways people use their nervous systems and bodies to sense, learn, reason, and take action" (Stone et al., 2016, p. 4). XAI refers to "the class of systems that provide visibility into how an AI system makes decisions and predictions and executes its actions" (Rai, 2020, p. 138). XAI can be applied in multiple areas. Financial investors can rely on explanations to understand what factors an algorithm includes as key predictors in its recommendations (Xu et al., 2019). Physicians can review explanations to understand what pathological features AI has used to generate its diagnosis (Xu et al., 2019).

The development of XAI serves multiple purposes. First, considering that AI may produce biased or discriminating results, explanations about AI's working mechanisms enable researchers to understand the generation of the results and prevent AI from making more errors. Second, as XAI makes part of AI's internal architecture visible and interpretable, researchers can know how to adjust or fine-tune AI models, which can improve AI's prediction accuracy. Third, when AI learns new strategies or solutions, explanations can serve as a tool to discover new knowledge (Adadi & Berrada, 2018). Overall, XAI has been regarded as a new research program that converts the black box of algorithmic decision-making into a glass box (Rai, 2020).

To enhance transparency and provide interpretable reasoning, XAI researchers have broadly explored two dimensions of explanations: scoop-based approaches and model-based approaches (Adadi & Berrada, 2018; Liao et al., 2020). Scoop-based approaches include global interpretability and local interpretability. Global interpretability refers to explanations of the logic applied to all models in AI systems and the entire reasoning leading to possible outcomes. By contrast, local interpretability refers to the explanations of a specific decision or prediction made by AI models.

The other dimension, model-based approaches, consists of model-specific interpretability and model-agnostic interpretability. The former reflects the explanations of the constraints added to the structure and learning mechanisms of the AI models. The latter refers to the approach that separates models from explanations and offers post hoc explanations to elucidate AI's predictions (Adadi & Berrada, 2018). For example, model-specific interpretability can explain the specific model structures (e.g., decision tree) or training algorithms (e.g., convex optimization), whereas model-agnostic interpretability can explain AI-generated outcomes using visualizations, examples, or metaphorical expressions (Xu et al., 2019). While model-specific interpretability is often used to provide explanations for AI experts, including computer scientists, engineers, and program developers, model-agnostic interpretability seeks to use human comprehensible language to meet non-AI experts' needs for understanding AI (Ehsan et al., 2021). In this article, we limit the scope of XAI literature to model-agnostic interpretability, which mostly falls in communication scholars' interests, as model-agnostic interpretability is brought to bear on lay people's perceptions and understanding of AI's internal workings.

What merits note here is that model-agnostic approaches can be either global or local. For example, researchers can use heatmaps or choropleths to illustrate how an AI-based medium recommends restaurants based on all users' check-in behavior. They may also use the same visual messages to explain how AI predicts a particular restaurant based on an individual's browsing records or check-in history.

Centering on the model-agnostic interpretability, XAI researchers have used various methods to provide explanations. A taxonomy of methods listed in Liao et al.'s (2021) article suggests that researchers can focus on explaining the input of data (e.g., data source and data labeling), the output of data (e.g., meanings of AI's predictions and verification of those predictions), and the performance of AI systems (e.g., the reliability and the limitations of the algorithms). XAI scholars have also adopted explanation techniques revolving around *how* AI makes predictions, *why* or *why not* certain predictions are presented, and *what if* model features change.

While XAI is burgeoning and exploring more techniques to enhance users' understanding of AI, its focus on explanations, which can be understood as a message exchange process between senders and receivers, would naturally benefit from communication research. For example, Miller (2019) argued that explanations about AI are socially constructed and selective. An explainer must understand an explainee's mental model before delivering the messages. Malle (2006) also mentioned that a qualified explainer must not only engage in gathering evidence for explanations but also learning to communicate explanations. Thus, much communication literature on message features and message exchange will have implications for the explanation dynamics explored in XAI.

In addition, Liao et al. (2020) pointed out that closing the gap between understanding the sophisticated algorithmic architecture and creating easy-to-understand explanations entails drawing on disciplines outside of XAI because it requires an understanding of human perception and cognition. Here, communication literature could also inform future XAI research on individuals' psychological processing of explanation-related messages and on the impact of message design on users' attitudinal and behavioral change.

Human–machine communication

Above we have introduced XAI and demonstrated how future XAI can be open to communication literature because of its needs for understanding the uses and effects of explanations. Below we introduce the development of HMC research and demonstrate how HMC, as a rising communication field to understand human–AI relationships, can benefit from XAI works.

HMC is defined as "the creation of meaning among humans and machines" (Guzman, 2018, p. 1). As a growing field of communication that focuses on machines as communicative subjects (Guzman, 2018), it involves human communication with a variety of digital interlocutors, including embodied social robots and virtual/augmented agents in either real or mixed environments (Edwards & Edwards, 2017). HMC emerged as scholars noticed a paradigm shift in human relationships with computer technologies, in which computers not only assume the roles of communication channels but also serve as digital interlocutors. In other words, individuals not only communicate *through* technologies, but also *with* technologies (Gunkel, 2012; Guzman, 2018).

Research on HMC can broadly be categorized into three interrelated threads. One thread discusses the ontological, moral, and cultural implications of machines, including the increasingly blurry boundaries between humans and machines (Guzman & Lewis, 2020), moral and legal expectations for robots (Gunkel, 2023), and the cultures and linguistic norms arising from interactions with machines, such as voice assistants and algorithms (Fortunati & Edwards, 2020).

Another thread of HMC research focuses on revisiting and extending classic concepts and theories, such as anthropomorphism (Kühne & Peter, 2023) and the Computers Are Social Actors (CASA) paradigm (Nass et al., 1994; Nass & Moon, 2000). For example, Gambino et al. (2020) suggested that as emerging technologies evolve, scholars should switch focuses from human–human scripts to human–media scripts. Drawing on evolutionary psychology, Lombard and Xu (2021) proposed updating the CASA paradigm based on a hierarchy of social cues that impose different effects on individuals' social responses to machines. More recently, van der Goot and Etzrodt (2023) distinguished media equation from media evocation and proposed using mixed methods to understand two interrelated paradigms: machines *are* social actors and machines *as* social actors.

The third thread examines individuals' direct interaction with technology interfaces and theorizes the effects of affordances on individuals' heuristic processing and actions (Sundar & Chen, 2023). For example, Sundar (2008) proposed the machine heuristic and suggested that if a machine interface presents machine characteristics (e.g., machines as sources), then individuals will likely attribute features such as randomness, objectivity, and fairness to its performance. Recently, Sundar (2020) proposed human–AI interaction (HAII)-theory of interactive media effects (TIME) as an extension of the TIME and theorized that AI-based media can influence users' cognitive heuristics by presenting attributes that either indicate AI's performances or trigger users' engagement with media, such as adjusting privacy settings, managing news feeds, and providing feedback for algorithms (Molina & Sundar, 2022).

Overall, HMC has been rapidly evolving. On the one hand, it inherits the intellectual discussion about machines across fields like human–computer interaction, human–robot interaction, information science, and sociology. It renews the perusal of topics related to machines, ranging from Suchman's (2007) conceptualizations of planned versus situated actions and Latour's (1992) actor-network theory to Turkle's (2012) theorization of children-Furby relationships and Weiser's (1991) ubiquitous computing. On the other hand, it is rooted in communication research and develops research programs about how individuals see machines as communicators, how individuals maintain relationships with machines, and essentially, the meaning-making process through exchange of messages between humans and machines (Guzman et al., 2023).

Like how XAI research may benefit from the communication scholarship on understanding messages, a few factors can portray the possibility and necessity to incorporate XAI works into future HMC research. First, the boundaries of HMC are still expanding. Guzman et al. (2023) cautioned against cutting off the trajectories of the HMC development before the full potential of HMC research is explored. Aligned with Guzman et al.'s (2023) perspective, technologies like LLM and diffusion models used in generative AI have raised users' curiosity about how and why AI generates recommendations, decisions, or creative content. The evolution of HMC can thus benefit from XAI literature in that the explanations about AI's decision-making process, including the quality of training data, model transparency, and algorithmic design can be viewed as potential factors that affect human–machine relationships.

Second, when outlining a research agenda for future HMC work, Guzman and Lewis (2020) called for research on both the functional dimensions and the relational dimensions of AI. Incorporating XAI literature can extend HMC research in both directions, as XAI can not only investigate how users justify AI's decisions and accept AI's recommendations but also create new relationships between humans and machines, as explanations about AI performances may illustrate the multi-phased and multi-layered human participation in algorithmic recommendations, which complicates research on users' perceptions of AI and its backstage mechanisms.

Third, HMC research has dominantly focused on individuals' direct interactions with technologies. Examples include users' evaluation and responses to AI-authored artworks or news reports (Spence et al., 2019a; Xu et al., 2020), users' psychological responses to chatbots and robots (Edwards et al., 2019; Lee & Liang, 2018; Westerman et al., 2019; Xu et al., 2024), and users' subjective experience of interacting with mobile voice assistants (Fortunati et al., 2022; Guzman, 2019;). Comparatively little research has examined how users perceive and respond to machines while simultaneously receiving and reacting to explanations about machines' output. If the direct interactions between humans and machines can be conceived as first-level HMC, which focuses on users' responses to the technical, social, and cultural dimensions of machines, then receiving, evaluating, and even interacting with the explanations of how machines work can be viewed as the second-level interaction in the broad HMC framework, especially considering that machines' performances can be explained from multiple angles related to human participation and human knowledge involvement in data annotation, model selection, and outcome verification, all of which might influence users' perceptions of and attitudes toward AI's recommendations. In other words, as an extension of the current HMC scholarship, XAI opens discussion for a two-level HMC framework in which users' responses can be conceptualized as outcomes of the interactions between their reactions to machine interfaces (i.e., first level) and their understanding of machines' working mechanisms (i.e., second level).

Integration directions: insights from XAI

Building on the benefits and feasibility of the integration between XAI and HMC, the subsections below will introduce six directions for future integration of XAI and HMC. We suggest that the human-in-the-loop approach and the dimension of message production explainability could enrich future HMC work. Meanwhile, knowledge about message sources, receivers, features, and effects allows HMC and other communication perspectives to segue into future XAI research.

Human-in-the-Loop approach

XAI researchers have applied the human-in-the-loop approach to understand the role of human elements in developing, intervening in, and verifying AI-made decisions. According to Deng et al. (2020), integrating human knowledge into machine learning can reduce the data requirement, increase the reliability of AI's predictions, and boost the precision of machine learning. The strength of the human-in-the-loop approach is that, while machine learning demands large data sets, humans can learn patterns from relatively small samples (Holzinger, 2016). Thus, when large data sets are not easily accessible (e.g., data sets about rare diseases), human experience and knowledge can provide direct instructions for and insights into data training.

Deng et al. (2020) suggested that two types of knowledge can be integrated into machine learning: general knowledge and domain knowledge. Whereas integrating general knowledge involves using knowledge about statistics, computer science, and calculation to enhance the performance of AI systems, integrating domain knowledge involves incorporating human experience with specific subjects into AI systems. For example, when developing algorithms for AlphaGo, a computer program that defeats professional human players, researchers incorporated professional Go players' domain knowledge into machine learning to increase AlphaGo's probability of finding the best strategy. In another example, when diagnosing diseases, doctors' expertise can be included in the algorithm-training process to reduce misclassification and enhance decision-making efficiency (Holzinger, 2016).

Research on HMC can benefit from the human-in-the-loop approach because the approach explicates the role of human elements in AI's black box decision-making process. In past works, although much HMC research has examined how users evaluate AI-generated content versus human-generated content, the backstage AI systems has not been fully taken into consideration. As some examples, HMC research has investigated automated journalism and suggested that machine-authored news was considered less credible and newsworthy than human-authored news (Waddell, 2018). Spence et al. (2019a) found that a Twitter bot that reported weather news was perceived as similar in news quality but less socially attractive than a human meteorologist. Xu et al. (2020) found that users' evaluations of AI-generated and human-generated paintings did not significantly differ, which challenged researchers to reconsider the meanings of human creativity.

Similar patterns have emerged when it comes to AI's decision-making. Molina and Sundar (2022) found that users tend to trust AI's decisions as much as humans' decisions in content moderation. However, users' perceptions of humans' and AI's decision-making may depend on specific communication tasks and contexts. Users perceived AI-made decisions in mechanical tasks (e.g., work scheduling) as equally fair and trustworthy as human-made ones but they questioned AI-made decisions in human tasks (e.g., hiring, work evaluation) (Lee, 2018).

Despite aforementioned HMC research on comparing human performances and machine performances, few communication studies have examined individuals' responses when they are exposed to the human elements in AI's internal working mechanisms (Liu, 2021; Spence et al., 2019b). As XAI scholars have mentioned that involving human knowledge in algorithms and making AI's internal architecture transparent can enhance users' trust (Deng et al., 2020; Ehsan et al., 2021), we suggest at least three forms of explanations about the human participation in AI's decision-making are ripe for future HMC scholarship: data annotation, outcome verification, and model selection.

Data annotation

In the data training phase, explanations about the role of human knowledge in determining and annotating training data may substantially affect users' perceptions of AI. For example, allowing AI to autonomously learn from digital traces without the involvement of human knowledge may cause the training data to be biased and useless. Posts about hate speech by Twitter's bot Tay have aptly captured this situation. Similarly, facial recognition technology may also learn from a non-representative sample of face images when no human knowledge is incorporated, which can lead to errors in various settings such as job candidate filtering and criminal detection. Nevertheless, if human experts are kept in the loop of data scraping and data labeling, then the representativeness and the quality of the training data could be checked and potentially improved. Thus, it should be safe to postulate that when AI users receive explanations about whether and how humans participate in sourcing data and labeling data, their perceptions of AI's performances would be affected.

Outcome verification

Based on the human-in-the-loop approach, testing explanations about whether and how AI-made predictions have been verified and/or modified by humans is also vital. In supervised learning, if training data are nonrepresentative, it would be risky for users to adopt AI's final recommendations directly. Perceptions of AI's decisions could change, however, if users know how human verification has acted as a guardrail to ensure the reliability of the results. Letting users know that humans have rewarded or punished AI-made decisions as part of the reinforcement learning process may also increase the perceived competency and perceived trustworthiness of AI during HMC.

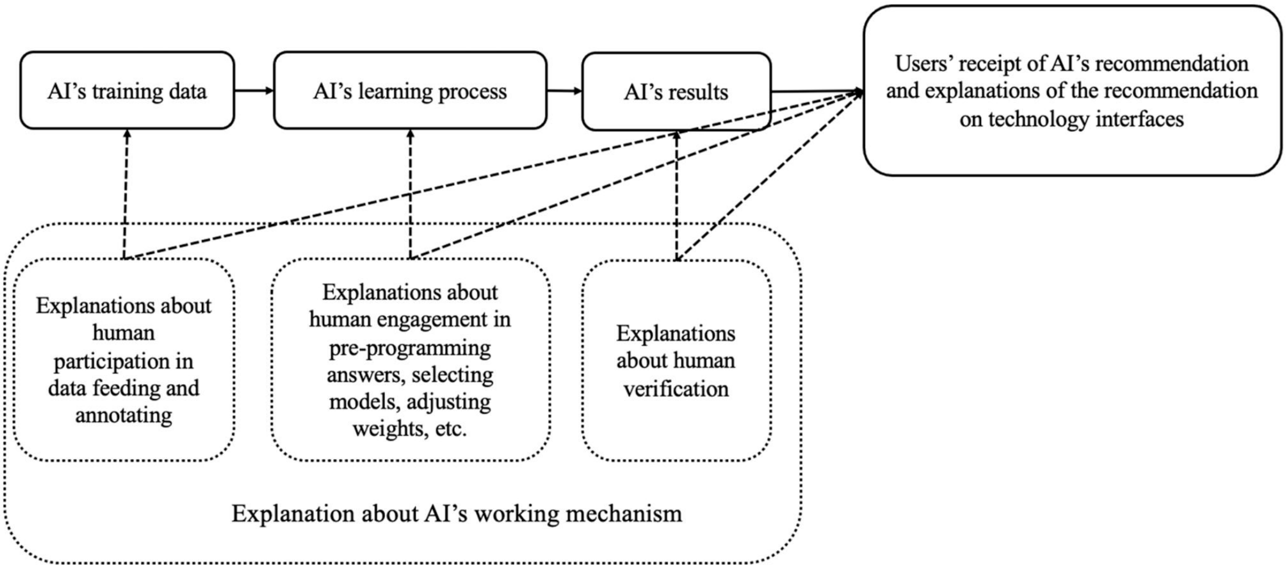
AI's training data
AI's learning process
AI's results
Users' receipt of AI's recommendation and explanations of the recommendation on technology interfaces
Explanations about human participation in data feeding and annotating
Explanations about human engagement in pre-programming answers, selecting models, adjusting weights, etc.
Explanations about human verification
Explanation about AI's working mechanism

Figure 1. Explanations of human participation in different AI production stages.

Model selection

Another type of explanations that communication scholars might explore through incorporating XAI research could concentrate on how humans are engaged in selecting or adjusting AI's models. Chatbots, for example, can be designed with rule-based pattern-matching approaches (e.g., using predefined, human-made answers to respond to users) or machine learning techniques, including natural language processing (e.g., using computing systems to collect and comprehend human language), artificial neural networks (e.g., using systems to compute vector representations, feeding them as features into the neural network, and producing responses), and LLMs (e.g., using massive amounts of text to predict the relationships between words) (see Adamopoulou & Moussiades, 2020). If users receive explanations about how human knowledge is involved in selecting or adjusting AI's models, then users may develop different perceptions and attitudes toward these chatbots. In the past, HMC research on these explanations about AI's working mechanisms remains fragmented, except that Liu (2021) found that AI systems using human-made rules triggered higher social presence than AI systems using machine-learned rules, which reduced users' uncertainty and increased their trust in AI.

Examining the elements of human participation in AI's decision-making processes can test and potentially expand the scope of some HMC theories. For instance, HMC research has applied the machine heuristic, derived from the MAIN model (modality, agency, interactivity, and navigability; Sundar, 2008), to understand individuals' cognitive processing of machines when the technology interfaces demonstrate machine agency cues (e.g., AI as an author or content generator). Future HMC research could reap the benefits from bridging the human-in-the-loop approach and machine heuristic and testing, for example, how machine agency, along with the explanations about human participation in the backstage of AI's model selection, affects users' cognition. Such combination may complicate individuals' attribution of machine features (e.g., objectiveness, accuracy) to AI. It could also raise questions about how individuals process both the machine agency cues on machine interfaces and the human participation cues in machines' internal architecture at the same time, which could in turn challenge, extend, or improve the predictive power of the MAIN model or its derivative frameworks such as HAII-TIME (Sundar, 2020).

Given that explanations are fundamentally messages delivered to AI users, we shall also add that beyond data annotation, outcome verification, and model selection, even some broad explanations about the human participation in AI's decision-making process may affect users' evaluation. For example, considering that non-AI experts may not have the cognitive load, motivation, or ability to process how humans are involved in different phases of AI's decision-making, these users may simply prefer to know *whether* or *how much* human knowledge has intervened AI's content generation process. Therefore, simply informing users of the degree of human knowledge input or human supervision in AI can also serve as a type of message, which could imply different perceptual or social consequences (Gil de Zúñiga et al., 2024). Other broad descriptions about human knowledge involvement (e.g., long-term vs. short-term human intervention, experts versus amateurs' participation) may also sway users' attitudes and potentially induce behavioral change. We demonstrate the potential of engaging the human-in-the-loop approach in HMC research in Figure 1.

Message production explainability

While adding the human-in-the-loop approach to HMC work could be one of the means of engaging XAI literature in communication scholarship, another contribution could lie in the theoretical expansion of the masspersonal model of communication (O'Sullivan & Carr, 2018). O'Sullivan and Carr (2018) proposed that media technology can be categorized along two perceptual dimensions—message accessibility and message personalization—that constitute a Cartesian coordinate system. The first quadrant features low message accessibility and high message personalization and thus refers to private and personal communication (e.g., phone calls, direct messages on social media platforms). The second quadrant features high message accessibility and low message personalization and thus refers to public and impersonal communication (e.g., podcasts, TV news). The third quadrant features high message accessibility and high message personalization and thus refers to the masspersonal nature of communication (e.g., radio call-ins, Facebook comments). Although O'Sullivan and Carr (2018) did not explicate the fourth quadrant, which features low message accessibility and low message personalization, some innovations of AI technology can be situated in this fourth quadrant. For example, humans' interactions with home-based voice assistants can be considered a private conversation given their low accessibility. Meanwhile, some of these voice assistants' messages can be impersonal, only providing the same predefined answers across users, devices, and contexts (e.g., when Alexa or Google Voice responds to users' inquiries about weather and time).

When O'Sullivan and Carr (2018) proposed their quadrant-based model, XAI research was not fully fledged. Today, however, lay users' exposure to some technical terms and concepts (e.g., machine learning, deep learning, generative AI) has become prevalent. Even so, terms like "Bayesian classifier," "decision tree," or "recurrent neural network" may not be intuitive or comprehensible to users, which could lead to various folk theories for understanding how AI works (Liao & Tyson, 2021). Meanwhile, these folk theories indeed reflect users' desire to reduce uncertainty and comprehend the black box of AI. To that end, we suggest that a third dimension, *message production explainability*, could be added to O'Sullivan and Carr's (2018) two-dimensional model. Derived from the emphasis of XAI literature on explanations, this dimension reflects the degree to which users perceive technology's message production process as interpretable, transparent, and visible.

We name the proposed three-dimensional model the M-PAPE model, which stands for message personalization, accessibility, and production explainability (see Figure 2). Theorizing emerging technology along these three dimensions is consistent with the variable-based approach proposed by Nass and Mason (1990), which indicates that technologies should be viewed as a combination of variables with different values. Using this variable-centered approach can enable researchers to understand how each technology represents different values on these dimensions and how technologies evolve in these features over time (Nass & Mason, 1990). Thus, adding the dimension of message production explainability can be seen as adding a new variable to understand technology features, which allows the M-PAPE model to more precisely capture users' perceptions of a range of AI technologies. Below, we use the examples of chatbots and augmented reality (AR) technologies to illustrate the value of using this updated model in future HMC research.

Chatbots may feature high message personalization when they provide tailored responses to individuals' needs and preferences on various kinds of websites. Meanwhile, they may also be considered impersonal when they use standardized identical messages (e.g., expressing gratitude for users and asking users to rate their service). At the level of message production, chatbots may differ in their algorithms. Those with pattern-matching algorithms feature higher human knowledge involvement in message production because pattern-matching algorithms rely on humans' preprogrammed, keyword-based responses to address users' inquiries (Liu, 2021). However, chatbots with unsupervised learning may feature comparatively lower human knowledge involvement in their algorithms. At the level of message accessibility, whereas chatbots used for private conversations have low accessibility (e.g., Replika), chatbots on social media can be highly accessible, as their posts and replies are public to users. Based on those characteristics, HMC research may draw on the M-PAPE model and manipulate the degree of perceived message personalization, message production explainability, and message accessibility to understand users' perceptions and interactions with different types of chatbots.

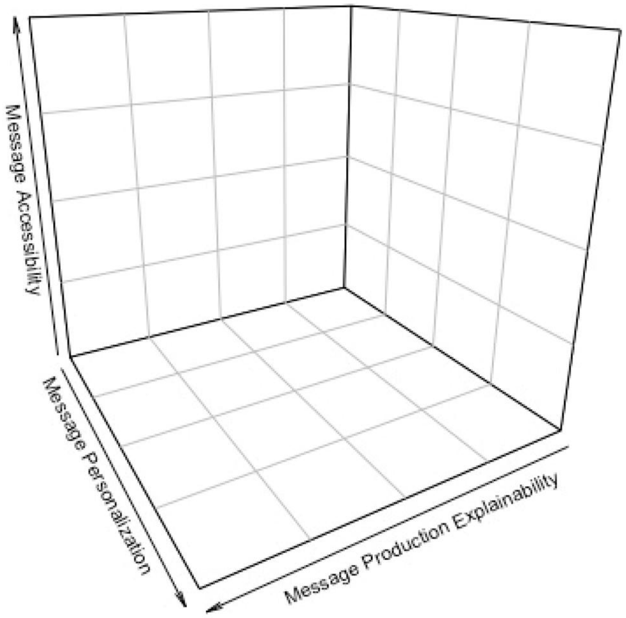

Figure 2. The Message Personalization, Accessibility, and Production Explainability (M-PAPE) Model.

Another example is in mobile AR technology (e.g., Pokémon Go, Apple Vision Pro), which uses location-based information to activate virtual overlays on users' physical surroundings (Liao & Humphreys, 2015). Depending on how explanations are framed, users may believe that the algorithms producing location-based AR content are self-learned, automatic, and location-sensitive. However, some users may also perceive that the AR messages are not produced by AI but by human artists or engineers. Thus, location-based AR messages may be mapped at different positions at the level of message production explainability based on how explanations are provided for users as well as how users interpret AR technology's working systems. Meanwhile, mobile AR messages can be regarded as having low accessibility if they are posted only within a small group of users in AR environments (e.g., Meta's Horizon workspace). In other cases, however, AR messages can be considered as having high accessibility if they are designed to be available in public spaces (e.g., the AR explanations of collections in museums and galleries). Furthermore, mobile AR technology can deliver impersonal messages if the messages are preprogrammed (i.e., low message personalization), or deliver personalized messages if the messages are shared in social settings in real time (e.g., AR Layar Tweets). Therefore, to understand and compare various forms of mobile AR technology, researchers may refer to the M-PAPE model to locate technological features on each dimension and test how AR technology differs from other technologies or how variations in perceived AR messages affect users' perceptions.

We envision that just like chatbots and AR, most AI technologies can be mapped onto the M-PAPE model. While research to date has focused only on the original two-dimensional model (e.g., Shi & Dai, 2022, 2023), users' frequent exposure to today's AI technology necessitates more research on users' perception of AI's message production process. Thus, future HMC research may draw from XAI literature to explore this dimension of message production explainability to understand users' perceptions of and engagement with AI technology.

Integration directions: insights from HMC and broad communication perspectives

Above we have demonstrated how XAI research can benefit future HMC research. In this part, we describe how HMC research, rooted in communication scholarship, could make contributions to the field of XAI. In the past, XAI literature has primarily focused on the content of the explanations, including describing *how* and *why/why not* AI makes certain decisions (Liao et al., 2021). Attention to classic communication concepts like sources, receivers, and messages rooted in Shannon and Weaver's (1964) model of communication has been scant. Indeed, these concepts have made important contributions to our understanding of HMC, such that Lee and Liang (2018) argued that in persuasion contexts, machines (i.e., senders/receivers) can use persuasive strategies (as messages) to gain compliance (i.e., effect) from users (i.e., senders/receivers). When defining persuasive AI, Dehnert and Mongeau (2022) also indicated that a communicative AI can generate, modify, or augment messages that are delivered to human receivers. In both cases, message sources, features, receivers, and effects serve as key concepts in understanding the HMC processes. Given that XAI seeks to deliver explanations to render AI's working mechanisms transparent, interpretable, and credible, we elaborate on how communication research, especially HMC research, on message sources, features, receivers, and effects can extend the current directions and expand the boundary of XAI.

Explanation sources

One important integration direction lies in understanding the sources of the explanations. Although XAI literature has indicated that the source of the training data could be a focus of explanation (Liao et al., 2021), communication scholars may enrich this area of research by providing knowledge about the multiple layers of explanation sources that could be involved in HMC. Specifically, sources in XAI may be categorized into the *providers of explanations*, the *source of training data*, and the (perceived) *source of communication partners* who deliver explanations to receivers.

More specifically, *providers of explanations* refer to the humans or machines that produce and/or offer explanations based on their expertise, authority, or learning results. For example, in HMC, explanations about algorithmic recommendations could be traced to human experts who originally designed or developed the algorithms. It is noteworthy that the providers of explanations may not necessarily be humans. At times, explanations can be provided by machines alone (e.g., when GPT is prompted to explain its answers) or by a hybrid of both humans and machines (e.g., when human explainers combine machine-generated explanations with their own knowledge).

Explaining the sources may also involve elaborating on the *source of training data* in supervised or deep learning models. By *source of training data*, we mean those who obtain and annotate the training data as well as how training data is sampled. The source of training data has received increasing attention, as scholars have found that problematic data labeling can generate biased algorithmic recommendations. Although Chen and Sundar (2023) found that showing AI users the quality of data labeling increased users' trust in AI, research on the source of training data remains limited. Moreover, given that the work of scoring, labeling, and classifying input data has often been crowdsourced to underpaid digital labor (Tubaro et al., 2020), explaining the characteristics of the digital labor, including the cultural backgrounds, demographic characteristics, and socioeconomic conditions of the digital labor behind AI's mechanisms would affect users' trust in AI. Thus, it is reasonable to expect that during HMC, explanations about the source of training data will play a role in users' evaluation for AI's performances.

The *source of the communication partner* refers to the participant who serves as the (perceived) communicator in delivering the explanations to users. Theorizing the source of the communication partner also involves multiple layers, such that users may believe that they are communicating with the AI per se (e.g., chatbot or virtual agent), the AI system (e.g., the algorithms), the engineers, the company that owns the AI system, or even the brand of the system. Perceptions of these different sources may lead to different cognitive and behavioral outcomes among users. Sometimes the perceived source of communication partner could overlap with the providers of explanations.

The reason to list all these possible sources that users may become aware of during HMC is to demonstrate that although XAI scholars have indicated the need to explain the source of the data, HMC perspectives reveal that sources can indeed involve many more layers and meanings, which can help clarify or expand the focuses of future XAI research. For example, in studying the source of the communication partner, Guzman (2018) interviewed participants about their experiences with voice-based mobile virtual assistants. She found that some participants perceived the source of the voice as an assistant in the mobile phone whereas others perceived the source of the voice as the phone per se. Ischen et al. (2020) also found that the perceived source of the communication partner (e.g., chatbots, websites) could be different from the source of AI's recommendations (e.g., expert-made recommendations, algorithmic recommendations).

A more comprehensive framework that has been applied to understand the perceived *source of the communication partner* in HMC is the source orientation model (Solomon & Wash, 2014). Imagine when a user communicates with an AI-based virtual assistant, the user interface presents explanations about why the virtual assistant makes certain recommendations. In this process, users may perceive the explanations as if the virtual assistant offered them directly. Users may also believe that the technology company or computer scientists who created the assistant provided the explanations. To understand how users target their communication partners, Solomon and Wash (2014) proposed the source orientation model, which suggests that with the increase of perceived source distance, users may sequentially orient to an application/software of the computer, the computer itself, other users, programmers, and organizations. The source proximity determines users' default orientation target. For example, Sundar and Nass (2000) suggested that during HMC, users naturally and socially orient to the computer agents first instead of the remote programmers of the agents.

The source orientation model further suggests that a few factors can change users' orientation process and reorient users' attention to different targets. When users accomplish tasks or goals with minimal effort, their orientation is likely to stay the same as their initial target (Solomon & Wash, 2014). When failure occurs during users' interaction with the perceived communication partner, they are likely to attribute the failure to other distant sources (e.g., programmers, organizations). For example, when a computer automatically updates its operating system and it takes longer than expected, users are likely to experience reorientation and blame the programmers for not creating a faster and smoother updating experience or even the brand of the computer for its frequent system updating requirement. Thus, the source orientation model may be informative for future XAI research when XAI scholars need to cast light on multiple sources involved in AI's working mechanisms.

Overall, HMC literature on source orientation is one example of how communication scholarship can add to XAI research. It indicates that using communication frameworks to theorize the sources of explanations offers a comprehensive approach to investigating the explanation sources in XAI.

Explanation messages

Given that explanations can be considered as messages that are designed to make AI systems more credible and interpretable, another direction in which communication research may inform XAI literature is elucidating how different message features can shape users' beliefs or attitudes toward an AI system or its recommendations.

To understand the effects of explanation messages, past HMC research, along with theories from CMC, AI-mediated communication (AI-MC), and persuasion, has shed light on the influence of message features. Within CMC theories, the social information processing theory suggests that online users can accrue impressions and advance relationships with technologies to a level that is expected in interpersonal communication (Walther et al., 2015). Despite the lack of visual cues, individuals can decode alternative combinations of cues that facilitate communication, which include communication styles, timing of message exchange, response delays, and the framing of messages (Walther & Parks, 2002). As an extension of social information processing theory, the hyper-personal model of communication suggests that impressions of others and relational states may even exceed what can be expected in offline communication, as communication receivers may form an exaggerated impression of others based on the cues they decode in the communication processes (Walther et al., 2015). One study that has focused on the effects of message features and bridged social information processing theory and HMC indicated that designing typos in a chatbot's messages negatively impacted the perceived attraction of the chatbot (Westerman et al., 2019). In a similar vein, Lew and Walther (2023) examined chatbots' response speed and conversational contingency and indicated that both factors had main effects on perceived trustworthiness. Although these studies were not directly conducted in XAI contexts, they may suggest that receivers' attitudes towards AI could hinge upon the cues embedded in messages, whether these explanation messages are generated by humans or machines.

A more systematic theorization of cues has inferred that as the boundaries between CMC and HMC become ambiguous, messages that involve the variations of cues like misspellings, memes, emojis, or punctuations can substantially affect individuals' understanding of AI (Xu & Liao, 2020). In our context, the messages may not only exist at the level of direct interaction between humans and machines, but also function as explanations communicated to AI users.

Beyond cues, explanation messages may be designed to reflect communication persuasion strategies, including using message sequences and frames (O'Keefe, 2015). Applying these persuasion strategies in HMC, Lee and Liang (2019) examined the foot-in-the-door effect and found that a robot that started with a small request then progressed to a large request was more likely to gain compliance from participants than one that directly began with the large request. Additionally, researchers may incorporate gain-frames or loss-frames into explanation messages. Although these two types of frames promote an identical recommended behavior, the emphases on the expected gains or the consequences of not accepting the recommendation could lead to divergent results, depending on contextual factors and individuals' cognitive processing (O'Keefe & Jensen, 2006). Although XAI scholars have used *what if* strategies (e.g., what would occur if the model changes to a different one) in explaining AI-made decisions, communication research may be well suited to make contributions here as it can impart to XAI how gain- versus loss-framed explanations exert disparate effects.

Future XAI research might also address the impact of different language styles. Various HMC studies have explored such impacts and the factors driving them. For example, Hancock et al. (2020) suggested that in future AI-MC, AI's overly positive language could lead to users' adoption of positive language, which over the long term, may shape our language norms and expectations for AI. When testing the CASA paradigm, Nass et al. (1995) found that participants who identified themselves as dominant perceived a computer that used dominant language (e.g., confident and assertive language) as more attractive and convincing than one that used compliant language (e.g., hesitant and indefinite language). Along the same line of CASA research, Xu (2020) found that for those who spent less daily time on mobile devices, mobile voice assistants' anthropomorphic language (i.e., casual, self-referential) strengthened users' intention of conformity, whereas non-anthropomorphic language (i.e., formal, non-self-referential) increased the intention of conformity for those who spent more daily time on mobile devices. Just as how different language styles set the stage for persuasive effects in HMC, they may also be employed in explanation messages to affect users' reactions to AI.

Overall, past HMC works jointly with other communication perspectives like CMC and persuasion may coalesce into future focuses of XAI research. Beyond message sequences, frames, and language styles, many more communication theories (e.g., communication accommodation theory) could be further applied to understand the potential of explanation messages. Although research combining XAI and communication is still limited, communication research revolving around messages could be fertile ground for understanding the various outcomes of explanations.

Explanation receivers

XAI scholars have acknowledged that the effects of explanations vary across individuals (Ehsan et al., 2021). Thus, they have used human-centered XAI to understand individuals' needs and to ease their interactions with AI systems. For example, XAI scholars have found that users with low-to-medium AI literacy reported a lower need for explanations than users with high AI literacy (Kim et al., 2023). In addition, those with rich experience using AI preferred concept-based explanations with detailed coefficients and equations. By contrast, users with limited AI experience felt overwhelmed and confused by seeing those numbers and instead reported a stronger preference for visual-oriented explanations (Kim et al., 2023).

Communication scholars may again contribute to XAI based on the ample literature on message receivers. Past HMC research has indicated that one's personality, anthropocentrism tendency, critical thinking, and cultural values can all affect their social responses to emerging technologies (Lombard & Xu, 2021). For instance, in the context of information classification, Molina and Sundar (2022) found that individuals who had more fear of AI were more likely to experience the negative machine heuristic (i.e., perceiving machines as less fair and less objective than humans in making judgments), while users who had more distrust in interpersonal communication were more likely to experience the positive machine heuristic, as they regarded machines as more accurate in classifying information than humans.

In addition to fear of AI, message receivers' motivation for processing explanations may also matter. The elaboration likelihood model (ELM) suggests that individuals' motivation is one of the factors determining whether individuals will carefully scrutinize messages or use mental shortcuts to digest information (Petty & Cacioppo, 1984). For instance, Liang et al. (2013) found that, compared to a high motivation state, individuals in a low motivation state were more likely to comply with a computer agent even if the agent used sham reasons, meaning that seeing the word "because" was sufficient to elicit those users' compliance, whether the reasons were legitimate or not.

ELM research has further indicated that issue involvement may affect users' motivation for systematic processing. Issue involvement refers to the extent to which an issue relates to one's goals or outcomes (Petty & Cacioppo, 1984). This concept could also be adopted in XAI. For instance, compared with an AI that explains its diagnoses of users' health status, users may have far less issue involvement with an AI that explains the recommendation of a banner advertisement. Considering that not all AI-made decisions are highly relevant to individual explanation receivers, it is here where XAI research could be more human-centered than domain-centered, allowing communication research to weigh in on how receivers' personal differences, including personalities, motivations, AI literacy, issue involvement, and other characteristics serve as important references in developing user-friendly explanations in HMC.

Explanation effects

Although XAI seeks to use explanations to increase users' trust and understanding, providing explanations does not always lead to users' trust. One study revealed that, among users with expertise in AI, simply providing an explanation induced caution, not trust, especially when the explanations did not match experts' expectations (Kloker et al., 2022). Moreover, Poursabzi-Sangdeh et al. (2021) found that, compared with a black box model, a transparent model undermines rather than improves users' ability in detecting and correcting the model's mistakes, partly due to the information overload induced by the transparent model. Ananny and Crawford (2018) also inquired into the limitations of treating transparency as an ideal solution, arguing that transparency can sometimes evoke individuals' privacy concerns and inhibit honest conversation. The effects of transparency depend on what information is explained and how clear, relevant, and precise the explanation is.

Despite the goal of using explanations to enhance users' trust, the abovementioned perspectives demonstrate a more complicated picture. Thus, future XAI research could home in on not only the expected effects but also unintended or undesired consequences. To understand unintended consequences, communication literature would be useful, for research on uses and gratifications (Katz et al., 1974; Palmgreen & Rayburn, 1985) has extensively documented the discrepancies between gratifications sought and gratifications obtained, which have been applied to describe how individuals leverage the benefits and the privacy concerns of using AI technologies such as voice assistants (Xu et al., 2022). Moreover, Cho and Salmon (2007) proposed a typology to categorize unintended effects in communication literature, including obfuscation, boomerang effects, and social reproduction. This typology could offer a promising lens to understand the unintended effects of explanations. For example, understanding individuals' psychological mechanisms of obfuscation may help XAI scholars better review, update, and adjust their explanations. Also, by examining social reproduction effects, XAI scholars could be more aware of the unintended consequences such as expanding rather than reducing the knowledge gap in users' AI literacy.

Last, recent research on XAI has predominantly measured trust or interpretability as the major outcome of explanations. From communication perspectives, an explanation's effect can include a range of users' psychological constructs (e.g., behavioral and attitudinal change, psychophysiological reactions). Given the increasing need to integrate explanations into human–AI interaction, it is imperative that scholars consider and explore the effects beyond trust and fathom how users' cognitive, affective, and behavioral responses vary based on different forms of explanations. Figure 3 presents how communication could expand XAI research in the four abovementioned directions.

A two-level HMC framework

Thus far, we have demonstrated how future HMC works could take advantage of the XAI angles such as the human-in-the-loop approach and the focus on message production explainability. We have also provided examples on how communication concentration on message sources, features, receivers, and effects could enlarge the scope of future XAI. In this section, we propose a two-level HMC framework to present our vision based on the integration of these two areas (Figure 4). Thanks to the contribution of prior HMC scholarship, the first-level HMC has been conceptualized and founded, exhibiting the focuses on how people make sense and engage with technologies that enact the role of communicators (Guzman et al., 2023). Although the boundaries of HMC are still expanding and raising ontological, theoretical, and methodological questions, first-level HMC centers on technology as a communicative subject and investigates the implications of technologies at the individual, organizational, and cultural levels.

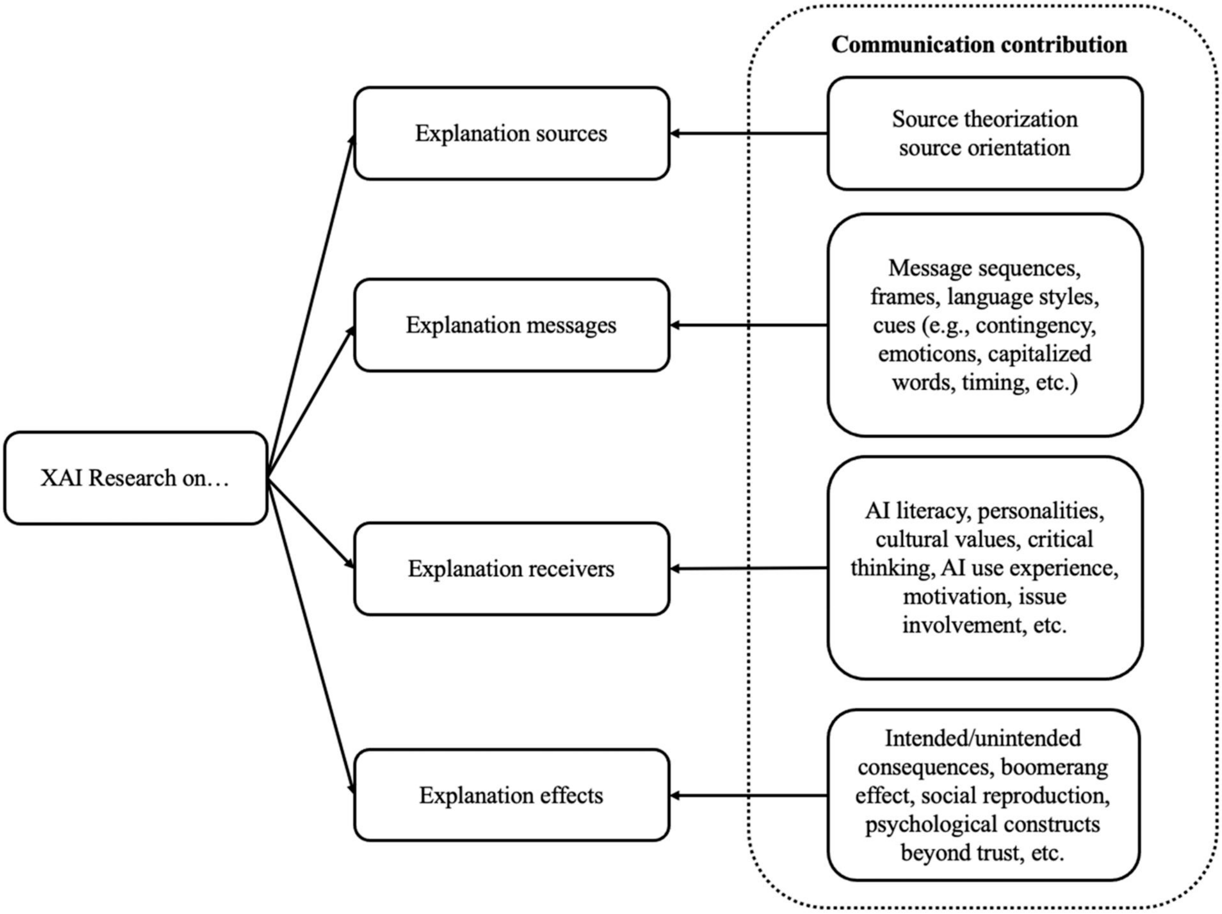
Communication contribution
XAI Research on…
Explanation sources
Explanation messages
Explanation receivers
Explanation effects
Source theorization source orientation
Message sequences, frames, language styles, cues (e.g., contingency, emoticons, capitalized words, timing, etc.)
AI literacy, personalities, cultural values, critical thinking, AI use experience, motivation, issue involvement, etc.
Intended/unintended consequences, boomerang effect, social reproduction, psychological constructs beyond trust, etc.

Figure 3. How communication research amplifies the scope of XAI.

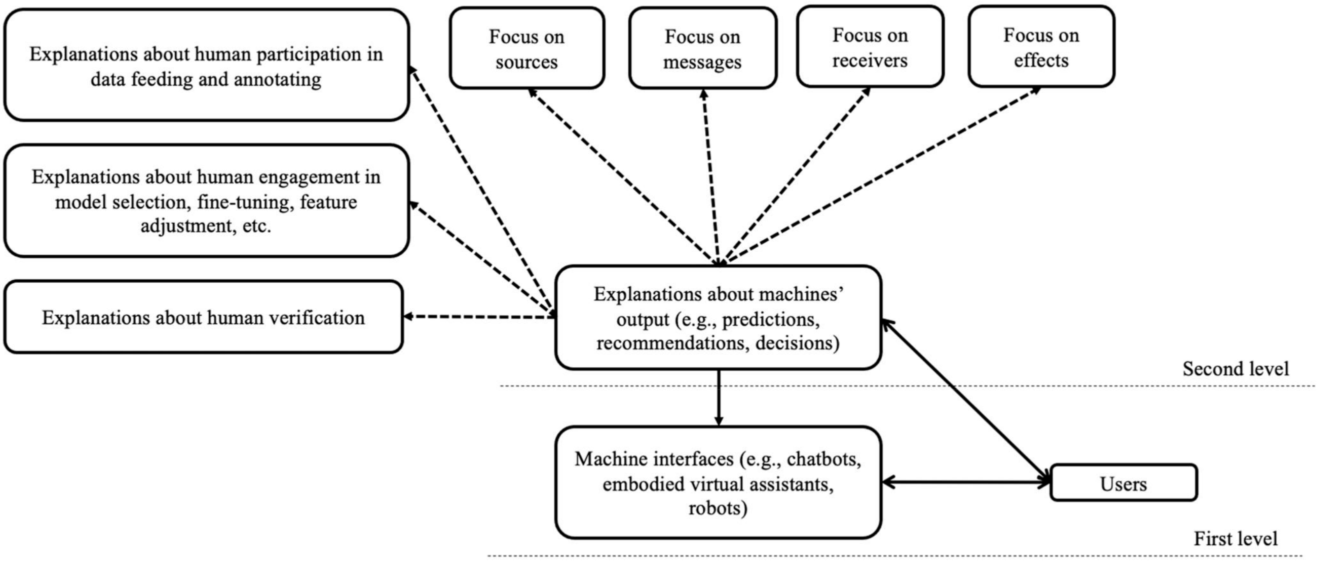
Explanations about human participation in data feeding and annotating
Explanations about human engagement in model selection, fine-tuning, feature adjustment, etc.
Explanations about human verification
Focus on sources
Focus on messages
Focus on receivers
Focus on effects
Explanations about machines' output (e.g., predictions, recommendations, decisions)
Second level
Machine interfaces (e.g., chatbots, embodied virtual assistants, robots)
Users
First level

Figure 4. Two-level human–machine communication framework.

Building on first-level HMC, we suggest that future research could explore a second-level HMC direction that theorizes how individuals receive, perceive, and evaluate the explanations about AI's working mechanisms. Such explanations could highlight the roles of human knowledge in AI's decision-making process (i.e., human-in-the-loop) or accentuate how AI's decisions are produced (i.e., message production explainability). Meanwhile, explanations at this second-level HMC could be designed and parsed with focuses on multi-layered sources and message features, as well as on explanation receivers and consequences.

Theoretical implications

The two-level HMC framework has the following theoretical implications. First, past communication research on AI has primarily compared HMC and human communication (Liu et al., 2023). That is, human–human communication has often been treated as a benchmark and HMC research has been conducted to infer AI's superiority or inferiority to humans along different dimensions, such as performing tasks, taking decisions, and making predictions (Gil de Zúñga et al., 2024; Spence, 2019). Yet, such comparison may not be sufficient in understanding the nuances of HMC. Questions like what features of AI make its performances different from humans still need much exploration and analyses (Gambino & Liu, 2022). The two-level HMC framework illuminates the importance of explaining AI's internal systems to users and portrays that in human–AI interaction, human elements can be involved in nearly all the production processes, including but not limited to data labeling, model training, model selection, fine tuning, and outcome verification.

Second, this two-level HMC framework illustrates that future works could look beyond users' interactions with machine interfaces and explore how users' evaluation of machine interfaces interact with their understanding of the black box of machines in determining users' attitudes toward and acceptance of AI recommendations. Apart from understanding machine cues, interface affordances, or humanlike attributes presented by machines (Sundar, 2020; Xu & Liao, 2020), this framework suggests that it is equally important to precisely convey how AI generates its decisions, especially when human knowledge is involved in the backstage AI systems.

Third, this framework indicates that in XAI, explanations are fundamentally messages. Explanations can be considered "a shared meaning-making process" between communicators (Ehsan et al., 2021, p. 2). Thus, to understand the effects of the explanations about algorithms is to understand the message sources, features, receivers, and effects. This framework demonstrates that communication scholars can make tremendous contributions to XAI, as communication scholars' expertise can deepen understandings of source credibility, message design, social cues, individual differences, and unintended consequences of message processing, all of which can serve as key concepts in understanding how users receive, evaluate, and accept explanations about AI's working mechanisms.

Fourth, the two-level HMC framework opens space for incorporating different theories and perspectives to guide future research on individuals' processing of explanations about AI. While first-level HMC treats machines as communicative subjects, the second-level HMC looks beyond how humans *communicate with* machines and calls for further research on how humans *understand* machines when communicating with them. In other words, if the first-level HMC seeks to stress *how AI can be communicative*, the second-level HMC probes *how AI can be communicated*. At this second-level HMC, XAI and HMC scholarships, together with other communication literature, including but are not limited to CMC, AI-MC, information processing, and persuasion can all be engaged to guide future work on humans' perceptions and evaluation of how AI works. It expands the scope of current HMC scholarship and serves as a theoretical framework to guide future human–AI interaction research when AI's internal workings are made transparent and when users are exposed to diversified forms of explanations about AI.

Ethical and practical implications

When investigating explanations about AI systems, scholars need to be aware of the ethical perils of using explanations to manipulate users' responses. For example, researchers should be cautious about who frames explanations, who verifies outcomes, and who develops algorithms. People who manage these tasks may have their own biases and may experience power influence and/or social pressure. Although XAI is anticipated to improve our AI literacy, it may also become a channel of reinforcing biases. Therefore, users should be informed of the role of human participation in these tasks and be aware of the potential manipulation of the explanations.

In addition to ethical practices, potential regulations could be enforced to supervise the transparency and interpretability of AI. The White House blueprint for an AI bill of rights and the EU AI Act could serve as the steppingstone for more responsible use of AI. As the White House blueprint (Office of Science and Technology Policy, 2022) suggests:

> Designers, developers, and deployers of automated systems should provide generally accessible plain language documentation including clear descriptions of the overall system functioning and the role automation plays, notice that such systems are in use, the individual or organization responsible for the system, and explanations of outcomes that are clear, timely, and accessible.

The EU AI Act also states that transparency "allows appropriate traceability and explainability" and includes "duly informing deployers of the capabilities and limitations of that AI system and affected persons about their rights" (Future of Life Institute, 2024), which is in congruence with the increasing demands for more interpretable, comprehensive, and transparent explanations.

At the same time, practical concerns may arise even when algorithm developers or technology companies acknowledge the necessity to unpack the black box of algorithms and disclose the human participation in algorithms. Tensions between disclosure and confidentiality may arise, and conflicts between open source and proprietary rights may emerge. Thus, even though XAI research may benefit greatly from communication perspectives, applying theory to practice will likely face barriers.

Despite these concerns, bridging HMC and XAI research can lay the foundation for designing and applying human-centered explanations. Knowledge of sources, message features, receivers, and effects can practically help XAI scholars refine explanations; knowledge of the human-in-the-loop approach and message production explainability can help communication scholars further investigate how users evaluate AI based on their working mechanisms.

Limitations and conclusions

We acknowledge that many more communication perspectives could be introduced to further amplify the scope of XAI. In that sense, we envision that this article only serves as the inception of the conversation between XAI and communication. Meanwhile, we are fully aware that both XAI and HMC are interdisciplinary fields and that no single framework is rooted entirely in either field. Indeed, the interdisciplinary nature of the two fields makes their frameworks all the more useful, informative, and predictive across fields.

As we approach an exciting but uncertain future of using and innovating AI technology, we face a growing demand for understanding how AI works, who develops and controls AI, and why AI makes certain recommendations. This article explores the areas in which the communication scholarship, especially HMC, and the XAI scholarship can be bridged. By introducing, analyzing, and integrating frameworks in both fields, we suggest that communication scholars can deepen their understanding of AI technology using XAI perspectives and XAI research can benefit from communication scholars' perspectives on concepts and theories revolving around messages. Exploring the relationships between communication and XAI is vital to the ongoing discussion and theoretical development for future AI research.

Data availability

No data were associated with this manuscript.

Conflicts of interest: None declared.

Acknowledgements

We are grateful for the editor, Dr. Jörg Matthes, and all the anonymous reviewers for their thoughtful suggestions and comments. We would also like to thank Dr. Bingjie Liu and Dr. Xiaohui Wang for their suggestions and encouragement during the ideation and revisions of this manuscript.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, *2*, 100006. https://doi.org/10.1016/j.mlwa.2020.100006

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Bradley, T. (2017, July 31). Facebook AI creates its own language in creepy preview of our potential future. *Forbes*. https://www.forbes.com/sites/tonybradley/2017/07/31/facebook-ai-creates-its-own-language-in-creepy-preview-of-our-potential-future/?sh=3795cb2b292c

Chen, C., & Sundar, S. S. (2023). Is this AI trained on credible data? The effects of labeling quality and performance bias on user trust. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–11. https://doi.org/10.1145/3544548.3580805

Cho, H., & Salmon, C. T. (2007). Unintended effects of health communication campaigns. *Journal of Communication*, *57*(2), 293–317. https://doi.org/10.1111/j.1460-2466.2007.00344.x

Dehnert, M., & Mongeau, P. A. (2022). Persuasion in the age of artificial intelligence (AI): Theories and complications of AI-based persuasion. *Human Communication Research*, *48*(3), 386–403. https://doi.org/10.1093/hcr/hqac006

Deng, C., Ji, X., Rainey, C., Zhang, J., & Lu, W. (2020). Integrating machine learning with human knowledge. *iScience*, *23*(11), 101656. https://doi.org/10.1016/j.isci.2020.101656

Edwards, A., & Edwards, C. (2017). The machines are coming: Future directions in instructional communication research. *Communication Education*, *66*(4), 487–488. https://doi.org/10.1080/03634523.2017.1349915

Edwards, C., Edwards, A., Stoll, B., Lin, X., & Massey, N. (2019). Evaluations of an artificial intelligence instructor's voice: Social identity theory in human-robot interactions. *Computers in Human Behavior*, *90*, 357–362. https://doi.org/10.1016/j.chb.2018.08.027

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding explainability: Towards social transparency in AI systems. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, *82*, 1–19. https://doi.org/10.1145/3411764.3445188

Fortunati, L., & Edwards, A. (2020). Opening space for theoretical, methodological, and empirical issues in human-machine communication. *Human-Machine Communication*, *1*, 7–18. https://doi.org/10.30658/hmc.1.1

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*, 107426. https://doi.org/10.1016/j.chb.2022.107426

Future of Life Institute (2024, January). *Recital 14a*. EU Artificial Intelligence Act. https://artificialintelligenceact.eu/recital/14a/

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*, 71–86. https://doi.org/10.30658/hmc.1.5

Gambino, A., & Liu, B. (2022). Considering the context to build theory in HCI, HRI, and HMC: Explicating differences in processes of communication and socialization with social technologies. *Human-Machine Communication*, *4*, 111–130. https://doi.org/10.30658/hmc.4.6

Gil de Zúñiga, H., Goyanes, M., & Durotoye, T. (2024). A scholarly definition of artificial intelligence (AI): Advancing AI as a conceptual framework in communication research. *Political Communication*, *41*(2), 317–334. https://doi.org/10.1080/10584609.2023.2290497

Gunkel, D. J. (2012). Communication and artificial intelligence: Opportunities and challenges for the 21st Century. *Communication +1*, *1*(1). https://doi.org/10.7275/R5QJ7F7R

Gunkel, D. J. (2023). *Person, thing, robot: A moral and legal ontology for the 21st century and beyond*. MIT Press.

Guzman, A. L. (2018). Introduction: 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. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, *90*, 343–350. https://doi.org/10.1016/j.chb.2018.08.009

Guzman, A. L., Jones, S., & McEwen, R. (2023). *The Sage handbook of human-machine communication*. Sage.

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society*, *22*(1), 70–86. https://doi.org/10.1177/1461444819858691

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. https://doi.org/10.1093/jcmc/zmz022

Holzinger, A. (2016). Interactive machine learning for health informatics: When do we need the human-in-the-loop? *Brain Informatics*, *3* (2), 119–131. https://doi.org/10.1007/S40708-016-0042-6

Ischen, C., Araujo, T., van Noort, G., Voorveld, H., & Smit, E. (2020). "I am here to assist you today": The role of entity, interactivity and experiential perceptions in chatbot persuasion. *Journal of Broadcasting & Electronic Media*, *64*(4), 615–639. https://doi.org/10.1080/08838151.2020.1834297

Katz, E., Blumler, J. G., & Gurevitch, M. (1974). *The uses and gratifications approach to mass communication*. Sage.

Kim, S. S., Watkins, E. A., Russakovsky, O., Fong, R., & Monroy-Hernandez, A. (2023). Help me help the AI: Understanding how explainability can support human-AI interaction. *Proceedings of the*

2023 CHI Conference on Human Factors in Computing Systems, *250*, 1–17. https://doi.org/10.1145/3544548.3581001

Kloker, A., Fleiß, J., Koeth, C., Kloiber, T., Ratheiser, P., & Thalmann, S. (2022). Caution or trust in AI? How to design XAI in sensitive use cases? *Proceedings of Americas Conference on Information Systems (AMCIS)*, *16*. https://aisel.aisnet.org/amcis2022/sig_dsa/sig_dsa/16

Kühne, R., & Peter, J. (2023). Anthropomorphism in human–robot interactions: A multidimensional conceptualization. *Communication Theory*, *33*(1), 42–52. https://doi.org/10.1093/ct/qtac020

Lai, V., Zhang, Y., Chen, C., Liao, Q. V., & Tan, C. (2023). Selective explanations: Leveraging human input to align explainable AI. *Proceedings of the ACM on Human-Computer Interaction*, *7* (CSCW2), 1–35. https://doi.org/10.1145/3610206

Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In W. E. Bijker, & J. Law (Eds), *Shaping technology/building society: Studies in sociotechnical change* (pp. 225–258). MIT Press.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1). https://doi.org/10.1177/2053951718756684

Lee, S. A., & Liang, Y. J. (2018). Theorizing verbally persuasive robots. In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology, and ourselves*. (pp. 119–143). Peter Lang.

Lee, S. A., & Liang, Y. J. (2019). Robotic foot-in-the-door: Using sequential-request persuasive strategies in human-robot interaction. *Computers in Human Behavior*, *90*, 351–356. https://doi.org/10.1016/j.chb.2018.08.026

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

Liang, Y. J., Lee, S. A., & Jang, J. W. (2013). Mindlessness and gaining compliance in computer-human interaction. *Computers in Human Behavior*, *29*(4), 1572–1579. https://doi.org/10.1016/j.chb.2013.01.009

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing design practices for explainable AI user experiences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, *463*, 1–15. https://doi.org/10.1145/3313831.3376590

Liao, Q. V., Singh, M., Zhang, Y., & Bellamy, R. (2021). Introduction to explainable AI. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, *127*, 1–3. https://doi.org/10.1145/3411763.3445016

Liao, T., & Humphreys, L. (2015). Layar-ed places: Using mobile augmented reality to tactically reengage, reproduce, and reappropriate public space. *New Media & Society*, *17*(9), 1418–1435. https://doi.org/10.1177/1461444814527734

Liao, T., & Tyson, O. (2021). "Crystal is creepy, but cool": Mapping folk theories and responses to automated personality recognition algorithms. *Social Media + Society*, *7*(2). https://doi.org/10.1177/20563051211010170

Liu, B. (2021). In AI we trust? Effects of agency locus and transparency on uncertainty reduction in human–AI interaction. *Journal of Computer-Mediated Communication*, 26(6), 384–402. https://doi.org/10.1093/jcmc/zmab013

Liu, B., Wei, L., Wu, M., & Luo, T. (2023). Speech production under uncertainty: how do job applicants experience and communicate with an AI interviewer? *Journal of Computer-Mediated Communication*, *28* (4), zmad028. https://doi.org/10.1093/jcmc/zmad028

Lombard, M., & Xu, K. (2021). Social responses to media technologies in the 21st century: The media are social actors paradigm. *Human-Machine Communication*, *2*, 29–55. https://doi.org/10.30658/hmc.2.2

Malle, B. F. (2006). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. MIT Press.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Molina, M. D., & Sundar, S. S. (2022). Does distrust in humans predict greater trust in AI? Role of individual differences in user responses to content moderation. *New Media & Society*, *26*(6), 3638–3656. https://doi.org/10.1177/14614448221103534

Nass, C., & Mason, L. (1990). On the study of technology and task: A variable-based approach. In J. Fulk & C. Steinfeld (Eds.), *Organizations and communication technology* (pp. 46–67). Sage.

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., Fogg, B. J., Reeves, B., & Dryer, C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, *43*(2), 223–239. https://doi.org/10.1006/ijhc.1995.1042

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72–78. https://doi.org/10.1145/191666.191703

Office of Science and Technology Policy (2022, October). *Blueprint for an AI Bill of Rights*. The White House. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

O'Keefe, D. J. (2015). *Persuasion: Theory and research* (3rd ed.). Sage Publications.

O'Keefe, D. J., & Jensen, J. D. (2006). The advantages of compliance or the disadvantages of noncompliance? A meta-analytic review of the relative persuasive effectiveness of gain-framed and loss-framed messages. *Annals of the International Communication Association*, *30*(1), 1–43. https://doi.org/10.1080/23808985.2006.11679054

O'Sullivan, P. B., & Carr, C. T. (2018). Masspersonal communication: A model bridging the mass-interpersonal divide. *New Media & Society*, *20*(3), 1161–1180. https://doi.org/10.1177/14614448166861

Palmgreen, P., & Rayburn II, J. D. (1985). A comparison of gratification models of media satisfaction. *Communication Monographs*, *52* (4), 334–346. https://doi.org/10.1080/03637758509376116

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, *2*, 1–22. https://doi.org/10.1145/3586183.3606763

Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, *46*(1), 69–81. https://doi.org/10.1037/0022-3514.46.1.69

Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Wortman Vaughan, J. W., & Wallach, H. (2021). Manipulating and measuring model interpretability. *Proceedings of the Human Factors in Computing Systems*, *237*, 1–52. https://doi.org/10.1145/3411764.3445315

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Shannon, C. E., & Weaver, W. (1964). *The mathematical theory of communication*. The University of Illinois Press.

Shi, J., & Dai, Y. (2022). Promoting favorable attitudes toward seeking counselling among people with depressive symptomatology: A masspersonal communication approach. *Health Communication*, *37*(2), 242–254. https://doi.org/10.1080/10410236.2020.1834209

Shi, J., & Dai, Y. (. (2023). Audience–campaign planner interaction in social media communication campaigns: How it influences intended campaign responses in the observing audience. *Human Communication Research*, *49*(3), 296–309. https://doi.org/10.1093/hcr/hqad003

Solomon, J., & Wash, R. (2014). Human-what interaction? Understanding user source orientation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 422–426. https://doi.org/10.1177/15419312145810

Spence, P. R., Westerman, D., Edwards, C., & Edwards, A. (2014). Welcoming our robot overlords: Initial expectations about interaction with a robot. *Communication Research Reports*, *31*(3), 272–280. https://doi.org/10.1080/08824096.2014.924337

Spence, P. R. (2019). Searching for questions, original thoughts, or advancing theory: Human-machine communication. *Computers in Human Behavior*, *90*, 285–287. https://doi.org/10.1016/j.chb.2018.09.014

Spence, P. R., Edwards, A., Edwards, C., & Jin, X. (2019a). 'The bot predicted rain, grab an umbrella': Few perceived differences in communication quality of a weather Twitterbot versus professional and amateur meteorologists. *Behaviour & Information Technology*, *38* (1), 101–109. https://doi.org/10.1080/0144929X.2018.1514425

Spence, P. R., Edwards, C., Edwards, A., & Lin, X. (2019b). Testing the machine heuristic: Robots and suspicion in news broadcasts. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 568–569. https://doi.org/10.1109/HRI.2019.8673108

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., & Teller, A. (2016, September) Artificial intelligence and life in 2030: The one-hundred year study on artificial intelligence. *Report of the 2015-2016 Study Panel*. Stanford University. https://ai100.stanford.edu/2016-report

Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press.

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital Media, Youth, and Credibility* (pp. 73–100). MIT Press.

Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication*, *25*(1), 74–88. https://doi.org/10.1093/jcmc/zmz026

Sundar, S. S., & Chen, J. (2023). From CASA to TIME: Machine as a source of media effects. In A. Guzman, R. McEwen, & S. Jones (Eds.), *The SAGE handbook of human-machine communication* (pp. 63–72). Sage. https://doi.org/10.4135/9781529782783

Sundar, S. S., & Nass, C. (2000). Source orientation in human-computer interaction: Programmer, networker, or independent social actor. *Communication Research*, *27*(6), 683–703. https://doi.org/10.1177/00936500002700

Tubaro, P., Casilli, A. A., & Coville, M. (2020). The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society*, *7*(1), 205395172091977. https://doi.org/10.1177/2053951720919776

Turkle, S. (2012). *Alone together: Why we expect more from technology and less from each other*. Basic Books.

van der Goot, M., & Etzrod, 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

Vincent, J. (2016,March 24). *Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day*. The Verge. https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Waddell, T. F. (2018). A robot wrote this? *How Perceived Machine Authorship Affects News Credibility*. *Digital Journalism*, *6*(2), 236–255. https://doi.org/10.1080/21670811.2017.1384319

Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in. In M. L., Knapp, & J. A. Daly (Eds.), *Handbook of interpersonal communication* (pp. 529–563) Sage.

Walther, J. B., Van Der Heide, B., Ramirez, A., Burgoon, J. K., & Pena, J. (2015). Interpersonal and hyperpersonal dimensions of computer-mediated communication. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology*. (pp. 3–22). John Wiley & Sons.

Weiser, M. (1991). The computer for the 21st century. *Scientific American*, *265*(3), 94–104.

Westerman, D., Cross, A. C., & Lindmark, P. G. (2019). I believe in a thing called bot: Perceptions of the humanness of "chatbots. *Communication Studies*, *70*(3), 295–312. https://doi.org/10.1080/10510974.2018.1557233

Wolf, C. T., & Ringland, K. E. (2020). Designing accessible, explainable AI (XAI) experiences. *ACM SIGACCESS Accessibility and Computing*, *125*, 6. https://doi.org/10.1145/3386296.3386302

Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P. Y. (2023). Supporting qualitative analysis with large language models: Combining codebook with GPT-3 for deductive coding. *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 75–78. https://doi.org/10.1145/3581754.3584136

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., & Zhu, J. (2019). Explainable AI: A brief survey on history, research areas, approaches, and challenges. In *Natural language processing and Chinese computing, NLPCC2019* (pp. 563–574). Springer. https://doi.org/10.1007/978-3-030-32236-6_51

Xu, K. (2020). Language, modality, and mobile media use experiences: Social responses to smartphone cues in a task-oriented context. *Telematics and Informatics*, *48*, 101344. https://doi.org/10.1016/j.tele.2020.101344

Xu, K., Chan-Olmsted, S., & Liu, F. (2022). Smart speakers require smart management: Two Routes from user gratifications to privacy settings. *International Journal of Communication*, *16*(2022), 192–214. https://ijoc.org/index.php/ijoc/article/view/17823

Xu, K., Chen, X., Liu, F., & Huang, L. (2024). What did you hear and what did you see? Understanding the transparency of facial recognition and speech recognition systems during human-robot interaction. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448241
256899

Xu, K., & Liao, T. (2020). Explicating cues: A typology for understanding emerging media technologies. *Journal of Computer-Mediated Communication*, *25*(1), 32–43. https://doi.org/10.1093/jcmc/zmz023

Xu, K., Liu, F., Mou, Y., Wu, Y., Zeng, J., & Schafer, M. (2020). Using machine learning to learn machines: A cross-cultural study of users' responses to machine-generated art works. *Journal of Broadcasting & Electronic Media*, *64*(4), 566–591. https://doi.org/10.1080/08838151.2020.1835136

Communication Theory, 2024, 00, 1–14
https://doi.org/10.1093/ct/qtae016
Original Article