Self-effects in AI-mediated communication

Àlex Valverde Valencia
Universitat Pompeu Fabra
alex.valverde02@estudiant.upf.edu
https://orcid.org/0009-0005-6789-8024

Autoefectos en la comunicación mediada por IA

ABSTRACT

The irruption of generative Artificial Intelligence, along with changes in content production and consumption models, sets the stage for the in-depth exploration of a specific field of media studies: self-effects. Defined as the cognitive, emotional, attitudinal or behavioral effects that messages have on the senders themselves (Valkenburg, 2017), this article aims to highlight the importance of studying such effects in AI-mediated communication (AI-MC) and content co-creation processes between AIs and humans with potential implications for digital wellbeing. First, the concept of Artificial Intelligence is narrowed down and its role in communication processes is discussed. Subsequently, the particularities of self-effects in AI-mediated communication are highlighted and the most important notions provided by current research on self-effects are stated for their application to AI-MC. Finally, the article discusses an important challenge presented by this field of study as well as some opportunities related to digital wellbeing and health in order to assess some potential social benefits of studying this type of effects.

RESUMEN

La irrupción de la Inteligencia Artificial generativa (GenAI), junto con los cambios en los modelos de producción y consumo de contenido, sienta las bases para la exploración en profundidad de un campo específico de los estudios de medios: los autoefectos. Definidos como los efectos cognitivos, emocionales, actitudinales o conductuales que los mensajes tienen sobre los propios emisores (Valkenburg, 2017), este artículo pretende destacar la importancia de estudiar dichos efectos en la comunicación mediada por IA (AI-MC) y los procesos de co-creación de contenidos entre IAs y humanos con potenciales implicaciones en el bienestar digital. En primer lugar, se acota el concepto de Inteligencia Artificial y se discute el rol que tiene en los procesos de comunicación. Posteriormente, se destacan las particu-laridades de los autoefectos en la comunicación mediada por IA y se destacan las nociones más importantes que aportan las investigaciones actuales sobre autoefectos para aplicarlas a la AI-MC .Por último, el artículo analiza un reto importante que presenta este campo de estudio así como algunas oportunidades relacionadas con el bienestar y la salud digitales con el fin de evaluar algunos beneficios sociales potenciales del estudio de este tipo de efectos.

KEYWORDS

Artificial Intelligence (AI); Human–AI interaction (HAII); Self-effects; Digital wellbeing; Communication research.

PALABRAS CLAVE

Inteligencia artificial (IA); Interacción Persona-IA (HAII); Autoefectos; Bienestar digital; Investigación en comunicación.

Autoefectes en la comunicació mediada per IA

RESUM

La irrupció de la Intel·ligència Artificial generativa (GenAI), juntament amb els canvis en els models de producció i consum de contingut estableixen les bases per a l'exploració en profunditat d'un camp específic dels estudis de mitjans: els autoefectes. Definits com els efectes cognitius, emocionals, actitudinals o conductuals que els missatges tenen sobre els mateixos emissors (Valkenburg, 2017), aquest article pretén destacar la importància d'estudiar aquests efectes en la comunicació mediada per IA (AI-MC) i els processos de co-creació de continguts entre IAs i humans amb potencials implicacions en el benestar digital. En primer lloc, s'acota el concepte d'intel·ligència artificial i es discuteix el rol que té en els processos de comunicació. Posteriorment, es destaquen les particularitats dels autoefectes en la comunicació mediada per IA i es destaquen les nocions més importants que aporten les investigacions actuals sobre autoefectes per aplicar-les a la AI-MC. Finalment, l'article analitza un repte important que presenta aquest camp d'estudi així com algunes oportunitats relacionades amb el benestar i la salut digitals per avaluar alguns beneficis socials potencials de l'estudi d'aquest tipus d'efectes.

1. Introduction: AI and the new paradigms in communication research

A new wave of AI-powered generative models has disrupted many areas in cultural studies, challenging the fundamental concepts of authorship (Lauber-Rönsberg & Hetmank, 2019; Deltorn & Macrez, 2019) and the uniqueness of human creativity (Arielli & Manovich, 2022), and creating a significant moment of exceptionality in communication research. This disruption has been enhanced by the release of Stable Diffusion, an open-source text-to-image artificial intelligence (AI) developed by Stability AI, Runway, and CompVis LMU. The public release of this model and its source code has sparked collective innovation around AI generative models, offering new opportunities for interaction between humans and AI (HAII), and setting the way for new research possibilities.

The exponential growth of these technologies has led scholars to reconsider the role of AI in communication processes beyond its traditional function as a tool or mediator. Instead, AI is now being conceived as a communicative subject (Guzman & Lewis, 2019) or a communicative Other (Gunkel, 2012). This new perspective is reinforced by the interactions users have with AI, which can be perceived as analogous to human interaction (Waddell et al., 2016). As a result, the paradigm of AI-mediated communication (AI-MC) needs to be broadened to encompass the concept of "machine subjects *with* which people make meaning instead of *through* which people make meaning" (Guzman & Lewis, 2019, p. 73).

Currently, the co-creation of meaning is expanded through the collective evolution of AI technologies, enabling new dimensions of content creation. One example of this is Dreambooth, introduced by Ruiz et al. (2022), which presents an approach for personalizing text-to-image diffusion models with a focus on subject-driven generation. With Dreambooth, any user can fine-tune a pre-trained text-to-image AI model to retrain it with a particular subject that was not present in its initial training (e.g. a person, a movie tone, an artistic style, etc.), and generate outputs that are based on that subject. This opens the door to personalized models for each individual, which are no longer limited by the previous database.

The ability to generate content with minimal skill or knowledge barriers represents a significant shift in the current media logic, especially given the new and complex interactions users are developing with AI (e.g., Brandtzaeg et al., 2022). These interactions place the user in front of a mirror, making them reconsider aspects of themselves (Turkle, 1984; Papacharissi, 2019), making the AI co-creation a process with effects on the creator. These phenomena are referred to as self-effects, defined as "the effects of messages on the cognitions,

emotions, attitudes, and behaviors of the message creators/ senders themselves" (Valkenburg, 2017, p.477).

This article aims to understand the context in which self-effects become relevant and provide insights for researching this field. First, the focus will be narrowed to AI's direct interaction with users and the roles that AI assumes in this type of process. Next, the article will address the field of media effects and self-effects by considering the particularities of Human-AI Interaction (HAII), highlighting the uniqueness of this research area. Finally, the article will discuss opportunities for studying self-effects related to digital well-being and health, with the goal of assessing the potential social benefits of an empirical understanding of these effects.

2. Narrowing the concept: The role of Generative AI in communication

As noted in the introduction, the development of Artificial Intelligence has expanded the boundaries of communication research. In line with the article's purpose, this section will focus on a specific area of AI that will be significant in the coming years. The first distinction to be made is between general and narrow AI. This research field will use the concept of narrow AI, which refers to goal-driven systems capable of performing one or a few tasks (AI HLEG, 2019) faster and more accurately than humans (Ninness & Ninness, 2020). To further refine the concept of AI used, we will focus on generative AI (GenAI), which is commonly defined as an algorithmic system capable of producing various types of content (e.g., audio, text, video, 3D models) inspired by the previous training datasets and able to generate unexpected outputs. GenAI has a unique position within communication processes, requiring us to reconsider the conventional view of AI as a channel.

The aforementioned change of paradigm regarding AI's role in communication (i.e. AI as a communicator) is not exclusive from the conception of AI as a mediator. This leads to the distinction of two traditionally separate fields: Human-Computer-Interaction (HCI) and Computer-Mediated-Communication (CMC). Although some scholars have attempted to reconcile the differences between these two concepts (Sundar & Lee, 2022), for the field of GenAI, it may be interesting to combine both approaches or view it as a spectrum, ranging from AI as a mediator to AI as a communicator, with various types of generative AIs situated in between. Multiple variables, such as AI involvement, communication purpose, functioning, interface, and human interaction can impact the placement of a particular generative AI within the spectrum.

To understand the concept of AI as both mediator and communicator, it is necessary to resort to recent developments that enable a more intimate communication process with AI. In Endacott & Leonardi (2022), significant progress has been

made in understanding how interactions with AIs can affect humans in impression management, providing an initial framework for exploring self-effects. The investigation is focused on communication processes where users interact with each other and AIs sometimes communicate on their behalf. When AIs communicate on behalf of someone, the user may not have complete control over how they are presented (Leonardi & Treem, 2020), as every action in that context can have a significant impact on how others perceive them (Metiu, 2006). Consequently, the user implicitly or explicitly delegates varying degrees of control over the impression management process to AIs (Endacott & Leonardi, 2022). These investigations lead to the assumption that this delegation of control over self-presentation may also be present in interpersonal interactions between AIs and humans, placing AI as the other agent in the communication process, situated on the mediator-communicator spectrum described before.

The impact of AI on communication processes and their outcomes is expanding, affecting not only interactions that involve self-presentation but also many other scenarios (Sundar & Lee, 2022). Therefore, the investigation of self-effects should not be limited to self-presentation processes that are co-created with AI. The variety of GenAI systems and their applications offer ample opportunities to explore self-effects in different contexts and scenarios, which will be discussed in the following section.

3. The particularities of self-effects in AI-mediated communication

To understand the importance of studying self-effects in the context of AI as a mediator and communicator, we need to first examine the concept of media effects and its evolution in recent times. Media effects refer to the impact that media content and messages have on the audience, and can be cognitive, emotional, attitudinal or behavioral effects (Potter, 2012). However, traditional media effects research has been limited to the concept of "mass" media, where consumption, impact, and anonymity were uniform across the audience (Valkenburg et al., 2016). This approach no longer fits the current media logic, where individualization and personalization are prevalent. This has given rise to the concept of mass self-communication, which enables researchers to examine not only the reception effects but also the generation processes and effects of media generation on the generators themselves (Castells, 2007). The new ways of individualization and personalization of media increase the engagement and the interactivity the user has with media and therefore, it can enhance media effects (Sundar et al., 2016). This seed, which Castells calls mass self-communication, is the prelude to the study of self-effects.

Empirical research has highlighted the significance of self-effects since the mid-20th century, for example, the effect

of role-playing on self-behavior (e.g. Elms, 1966). Despite the similarities in the mechanisms of self-effects between offline and online scenarios, the affordances of digital media, particularly social media, make online self-effects more likely or expected (Valkenburg, 2017). The heightened frequency of self-effects in online environments, along with the person-alization of consumption and production, and the distinctive role of AI in communication processes, elevates the research domain of self-effects in AI-MC as a unique area that can yield significant conclusions.

Several scholars have conceptualized and studied self-effects from diverse angles in digital media, such as Jensen Schau and Gilly (2003), Yee et al. (2009), and Peña et al. (2023). Nevertheless, due to the similarities and links between offline and online self-effects, it is feasible to apply offline theories to digital surroundings, as proposed by Valkenburg (2017) with self-effects in social media, distinguishing self-effects from reception effects (i. e. the effects of others' messages on the cognitions, emotions, attitudes, and behavior of recipients). Valkenburg (2017) identified four principal theories underlying self-effects: self-persuasion, self-concept change, expressive writing, and political deliberation. Although some studies have addressed these effects, such as Gore and Cross (2011), there is a lack of empirical research regarding self-effects in HAII. It is crucial to examine the variations and differences in these effects when the context is mediated by an AI, especially if we consider that new communication tools, platforms and agents can have unique and unidentified effects (Yao & Ling, 2020).

Throughout this article, several unique characteristics of HAII and AI-MC have been highlighted, underscoring the need for specific research on self-effects within GenAI. One essential characteristic of these communication processes is that both reception and self-effects are produced simultaneously, given that AI functions as a mediator and communicator in the generative process. Therefore, analyzing HAII in content generation processes will not only help us understand what impact AI-mediated generation has on the generator but will also provide insights into how the user interacts with the AI as another agent in the process and what effects the interaction with AIs has on the user on a scale closer to interpersonal interaction with humans, similar to the approach done by Brandtzaeg et al. (2022) with chatbots but a much larger scale.

The proliferation of GenAI in multiple content-generation applications and the growing relationship between users and AI technologies have significant implications in several areas of our daily lives. The study of self-effects can unlock the exploration of mechanisms to improve the positive aspects of these effects (e.g., Raglio et al., 2021) and mitigate the potential adverse outcomes of HAII and its impact on society (e.g., Dignum, 2017). Finally, the study of self-effects meets an urgent need to develop a practical understanding of AI-MC impact on a wide variety of behaviors and design, implement

and fairly regulate these systems (e.g. Hancock et al., 2020). Some challenges and opportunities in this field of research will be discussed in the last section of this paper.

4. Discussion: Challenges and opportunities of studying self-effects in AI-MC

The singularities of AI-MC as a research field have been expressed in the article with a double purpose, first to demonstrate such uniqueness in that area of study and, secondly, to bring to light the importance of studying self-effects in order to create safer and more ethical AI systems in the future. This last section will briefly explore some challenges and opportunities when starting self-effects research in AI-MC.

One of the biggest challenges is linked to the unique role of AI in communication processes, as reception effects and self-effects are produced simultaneously. Therefore, researchers must precisely define the object of study to avoid the measurement of different effects at the same time. Previous theoretical research, such as Gore & Cross's (2011) study on self-concept change, offers valuable insights into narrowing the focus on self-effects. In this study, the researchers limited the variables to measure this effect and obtained the expected results regarding the effect. This research highlights the need for specificity and particular indicators of change for each self-effect. Therefore, to assess the impact of AI-MC on the self for measuring self-persuasion and self-concept change, researchers should set different variables and patterns to evaluate each effect.

It is essential to consider the interdisciplinary nature of AI-MC and its connections with other research disciplines. AI-MC cannot be studied in isolation, and it is necessary to take into account both classical theories in communication and important contributions from related areas of inquiry (Yao & Ling, 2020). An interdisciplinary approach can provide an opportunity to develop more robust lines of research supported by various fields of expertise. For example, the study of self-effects in AI-MC can benefit from psychological research (e.g. Sundar, 2020) and algorithmic therapies (e.g. Raglio et al., 2021), as well as from classical media effects theories such as parasocial interaction or uses-and-gratifi-cations.

Perhaps the most important opportunity that arises from the study of self-effects, besides contributing to the more ethical development of AI systems, is the potential to improve the digital well-being of diverse groups, particularly those with specific clinical conditions. An in-depth understanding of how GenAI affects users may lead to algorithmic therapies that could benefit many individuals, ranging from therapies that alleviate the effects of degenerative diseases to enhancing the self-expression of those on the autism spectrum through

AI-generated content. In conclusion, the study of self-effects can help us understand this technology better; in doing so, we might better understand ourselves.

References

AI HLEG (2019). *Ethics guidelines for trustworthy AI*. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

Arielli, E., & Manovich, L. (2022). AI-aesthetics and the Anthropocentric Myth of Creativity. *Nodes* (19-20), 8-15.

Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research, 48*(3), 404-429. https://doi.org/10.1093/hcr/hqac008

Castells, M. (2007). Communication, power and counter-power in the network society. *International Journal of Communication*, 1, 238–266. https://ijoc.org/index.php/ijoc/article/view/46

Deltorn, J. M., & Macrez, F. (2019). Authorship in the age of machine learning and artificial intelligence. *SSRN*. https://doi.org/10.2139/ssrn.3261329

Dignum, V. (2017). Responsible artificial intelligence: designing AI for human values. *ITU Journal: ICT Discoveries, 1*(1). https://handle.itu.int/11.1002/pub/8129fdfb-en

Endacott, C. G., & Leonardi, P. M. (2022). Artificial intelligence and impression management: Consequences of autonomous conversational agents communicating on one's behalf. *Human Communication Research, 48*(3), 462-490. https://doi.org/10.1093/hcr/hqac009

Gore, J. S., & Cross, S. E. (2011). Defining and measuring self-concept change. *Psychological Studies, 56*(1), 135-141. https://doi.org/10.1007/s12646-011-0067-0

Gunkel, D. J. (2012). Communication and artificial intelligence: Opportunities and challenges for the 21st century. *Communication+ 1, 1*(1), 1-25. https://doi.org/10.7275/R5QJ7F7R

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

Jensen-Schau, H., & Gilly, M. C. (2003). We are what we post? Self-presentation in personal web space. *Journal of consumer research, 30*(3), 385-404. https://doi.org/10.1086/378616

Lauber-Rönsberg, A., & Hetmank, S. (2019). The concept of authorship and inventorship under pressure: Does artificial intelligence shift paradigms? *Journal of Intellectual Property Law & Practice, 14*(7), 570–579. https://doi.org/10.1093/jiplp/jpz061

Leonardi, P. M., & Treem, J. W. (2020). Behavioral visibility: A new paradigm for organization studies in the age of digitization, digitalization, and datafication. *Organization Studies, 41*(12), 1601-1625. https://doi.org/10.1177/0170840620970728

Metiu A. (2006). Owning the code: Status closure in distributed groups. *Organization Science, 17*(4), 418–435. https://doi.org/10.1287/orsc.1060.0195

Ninness, C., & Ninness, S. K. (2020). Emergent virtual analytics: Artificial intelligence and human-computer interactions. *Behavior and Social Issues, 29*(1), 100-118. https://doi.org/10.1007/s42822-020-00031-1

Papacharissi, Z. (Ed.). (2019). *A networked self and birth, life, death*. Routledge.

Peña, J., Barake, M. A., & Falin, J. M. (2023). Virtual leaders: Can customizing authoritarian and democratic business leader avatars influence altruistic behavior and leadership empowerment perceptions? *Computers in Human Behavior*, 141, 107616. https://doi.org/10.1016/j.chb.2022.107616

Raglio, A., Baiardi, P., Vizzari, G., Imbriani, M., Castelli, M., Manzoni, S., Vico, F., & Manzoni, L. (2021). Algorithmic music for therapy: effectiveness and perspectives. *Applied Sciences, 11*(19), 8833. https://doi.org/10.3390/app11198833

Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2022). Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. *arXiv* preprint arXiv:2208.12242. https://doi.org/10.48550/arXiv.2208.12242

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., & Lee, E. J. (2022). Rethinking communication in the era of artificial intelligence. *Human Communication Research, 48*(3), 379-385. https://doi.org/10.1093/hcr/hqac014

Sundar, S. S., Jia, H., Waddell, T. F., & Huang, Y. (2015). Toward a theory of interactive media effects (TIME) four models for explaining how interface features affect user psychology. In: S. S. Sundar, (Ed.), *The handbook of the psychology of communication technology*, (pp. 47-86). Wiley. https://doi.org/10.1002/9781118426456.ch3

Valkenburg, P. M. (2017). Understanding self-effects in social media. *Human Communication Research, 43*(4), 477-490. https://doi.org/10.1111/hcre.12113

Valkenburg, P. M., Peter, J., & Walther, J. B. (2016). Media effects: Theory and research. *Annual review of psychology, 67*, 315-338. https://doi.org/10.1146/annurev-psych-122414-033608

Waddell T.F., Zhang B., & Sundar, S.S. (2016) Human–computer interaction. In: *The International Encyclopedia of Interpersonal Communication*. Wiley. https://doi.org/10.1002/9781118540190.wbeic182

Yao, M. Z., & Ling, R. (2020). "What is computer-mediated communication?"—An introduction to the special issue. *Journal of Computer-Mediated Communication, 25*(1), 4-8. https://doi.org/10.1093/jcmc/zmz027

Yee, N., Bailenson, J. N., & Ducheneaut, N. (2009). The Proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research, 36*(2), 285-312. https://doi.org/10.1177/0093650208330254

CV

Àlex Valverde Valencia. He is a young Catalan researcher who studied Advertising and Public Relations at Pompeu Fabra University (UPF). He is currently enrolled in the M.A. in Digital Culture and Emerging Media with the aim of starting a PhD in Communication later on. His main academic interests are the interactions between Humans and Artificial Intelligence and the effects of these intereactions, both in the individual and in society as a whole, with an eye on the potential benefits of Human-AI Interaction in Education. He collaborated on the research day of the +RAIN Film Fest and was the promoter, along with Dr Frederic Guerrero Solé, of Barcelona's first community of creation and reflection on AI.