

AI Has Joined the Group: Collective Agency Over Sociotechnical Systems — A Perspective from Brazil

Author: Lauren Pachaly

AI Ethics and Society, Module 1

December 8th, 2025

Word count: 2,076

1. Introduction

Western epistemologies of technology—grounded in ideals of mastery, prediction, and control—tend to cast AI as an instrument of domination and knowledge extraction (Pasquinelli & Joler, 2021). Within such frames, human contribution is often associated with enabling machines “to use all of our knowledge to construct a computer program that knows” (McCarthy & Hayes, 1981). This paper challenges these framings by drawing on relational and decolonial epistemologies, which reconceive human agency over knowledge instruments as situated, collective, and relational, emerging from the social and cultural practices through which technologies are transformed in everyday life.

Brazil offers a compelling context for examining how relational agency materializes. WhatsApp, primary communication platform in the region, has developed into a sociotechnical framework embedded in everyday routines—not a fixed instrument, but a living system co-produced through collective practices. WhatsApp Groups, in particular, stand as an emblematic example of this dynamic. In these spaces, knowledge is situated in trust, local experience, and community norms that guide how information is produced, evaluated, and circulated. Users demonstrate agency not only by using the tool but by reshaping its social and epistemic roles.

However, these community-governed knowledge spaces are now facing substantial and decisive forms of pressure. As governments and technology companies express growing interest in entering group-chat environments—whether through state-driven invisible participation or through proactive AI agents—the very conditions that enable situated knowledge are at risk.

This paper therefore analyzes how user agency—understood as relational practice—has enabled communities in Brazil to reshape technological instruments, while also examining the risks posed by emerging institutional and corporate interventions. Section 2 situates user agency within wider literature; Section 3 explores WhatsApp as a collective infrastructure and a site of situated knowledge in Brazil; Section 4 introduces contemporary threats posed by new forms of participation; Section 5 examines associated risks; and Section 6 concludes.

Ultimately, this paper calls for a closer examination of what is at stake when AI enters community-governed knowledge spaces. It argues that users—who have historically reshaped

technological instruments through relational agency—must now mobilize that same agency to set the norms that will govern AI's participation in these environments.

2. Situating User Agency and Collective Influence Over Instruments in the Wider Literature

Understanding user agency—and the capacity of users to shape and redirect sociotechnical systems—is a multidisciplinary effort. Western technological traditions, rooted in ideals of mastery and optimization, often conceptualize AI as an instrument of knowledge extraction. Pasquinelli & Joler (2021) describe AI as a device that “perceives features, patterns, and correlations through vast spaces of data beyond human reach,” while human influence is typically reduced to the cultural construction and labeling of datasets. Ribes et al. (2019) reinforce this understanding by showing how human contribution becomes concentrated primarily during knowledge acquisition, where domain logics are encoded into data infrastructures. Together, these traditions construct what Ricaurte (2019) critiques as Western rationality and dualism—a colonial epistemology that separates the ‘knower’ from the ‘known’, mind from body, subject from object.

In contrast, relational epistemologies—rooted in Indigenous, Afro-diasporic, feminist, and decolonial traditions—argue that agency is grounded in networks of care, responsibility, and structural transformation. Ricaurte (2025) reframes agency as “response-ability,” emphasizing that agency emerges through collective care and the reconfiguration of power. McQuillan (2022) adds a political dimension, arguing that optimization-based AI mirrors authoritarian rationality, requiring alternatives rooted in democratic and collective governance. Nemer (2022) further illustrates—through the concept of ‘Mundane Technologies’—how agency operates not through exceptional acts of resistance but through quotidian practices and relational forms of self-organization that allow the user to reinvent technology.

Together, these literatures highlight the significance of collective practices in shaping technological meaning—a perspective that resonates strongly in Brazil, where communities have long adapted technologies to meet local needs and shape usage cultures.

3. Collective Shaping of Sociotechnical Systems in Brazil: WhatsApp Case Study

The widespread, deeply embedded use of WhatsApp in Brazil offers a powerful case study of how relational and collective agency operates in sociotechnical systems. Far from simply adopting a digital tool, Brazilian communities have transformed WhatsApp into a sociotechnical infrastructure. This disrupts paradigms that frame technologies as fixed instruments of extraction. Instead, it illuminates a model of collective agency in which communities co-produce technological meaning and function through situated practices.

WhatsApp as collective infrastructure

By 2025, around 150 million Brazilians—over 93% of Internet users—relied on WhatsApp regularly (Statista, 2025). The platform has become a relational infrastructure woven into daily life, supporting family ties, community coordination, commerce, religion, and political organization. Its widespread adoption is deeply tied to Brazil's telecom landscape, where carriers frequently zero-rate the data consumed on the platform (Nemer, 2022). Despite regulatory efforts to protect net neutrality, this “free” access has made WhatsApp function as “the Internet” for much of Brazil's low-income population (Omari, 2020). As a result, Brazilians use WhatsApp for nearly every aspect of daily life—from hyperlocal coordination, crowdsourced problem-solving, emotional support, information sharing to small-business operations (OpinionBox, 2025; Statista, 2024). Its saturation in social routines has transformed it from a messaging app into a collective infrastructure shaped by users' norms, priorities, and political conditions.

WhatsApp Groups as Sites of Situated Knowledge

WhatsApp groups are central to this infrastructure. Approximately 94% of users belong to at least one group (OpinionBox, 2025). These invitation-only spaces function as bottom-up information infrastructures—intimate knowledge circuits governed by community norms rather than algorithms or corporate moderation systems. Because entry is controlled by someone in the community, groups often reflect tight social bonds or affinity—family, school, neighborhood, faith, parenting, hobbies.

According to Edelman (2024), Brazilians rank “someone like me” among their most trusted sources of information. The Reuters Institute (Newman et al., 2025) similarly finds that Brazilians prioritize information from trusted contacts over media institutions. This reliance has also enabled rumors and misinformation to spread rapidly in groups, especially during elections—an issue that both communities and WhatsApp have had to navigate (Avelar, 2019).

Nonetheless, WhatsApp Groups remain central sites of situated knowledge, in Haraway’s sense of context-bound, non-universal ways of knowing (Haraway, 1988). They function as social filters for information sharing, where trust, personal recommendation, and lived experience become sources of validation. In this environment, community signals and agency play a larger role in discovery than algorithmic content moderation.

4. When power joins the group

The relational dynamics that have allowed WhatsApp Groups to become sites of situated knowledge now face an emerging set of pressures. While the state seeks to be an invisible listener (“Ghost”) within people’s private conversations, Big Tech—after a decade of attempting to monetize and shape the high-trust group-chat environment—seeks to become an active, visible participant (“Agent”). While one seeks to listen without presence, the other seeks to speak without belonging. Together, these ambitions signal a profound shift: a moment when relational knowledge spaces become infrastructures that institutions of power wish to occupy, study, and shape.

State as the invisible participant: The UK “Ghost Protocol”

A prominent example of institutional ambition to enter group-chat environments is the United Kingdom’s proposed “Ghost Protocol.” First revealed in 2018 through the GCHQ’s “Exceptional Access” proposal, it sought to enable state authorities to be silently added to encrypted group conversations without the knowledge or consent of participants (Levy & Robinson, 2018). Crucially, the proposal required the messaging app to suppress the standard notification that a new participant has joined the group. To the human users, membership would appear unchanged, while the “ghost” receives a copy of every message in real-time.

The global technology community and civil society organizations vehemently rejected this framing. An open letter (Various authors, 2019) signed by 47 entities—including Apple, Google, WhatsApp—argued that the proposal would introduce systemic vulnerabilities.

The mere existence of these proposals creates a "Panoptic effect" within the digital lifeworld (Lyon, 2006). The awareness that a "ghost" could be present chills free association and shifts community spaces from "safe" to "managed".

ChatGPT has joined the group

If the state seeks to listen without being seen, Big Tech seeks the opposite. For the tech industry, the group chat has long represented an untapped domain of high-trust interaction—precisely the kind of environment companies have struggled to monetize and algorithmically intervene in. With the release of AI-powered group chat agents, this ambition has become explicit. OpenAI's 2025 launch of group-chat assistants introduced AI assistants capable of interpreting conversational context, responding dynamically, and shaping how conversations unfold (OpenAI, 2025).

But participation is not the same as belonging. These AI agents enter conversations carrying the epistemic assumptions, datasets, and commercial incentives of their creators. They insert themselves into ongoing social negotiations, subtly affecting how information is shared, validated, and interpreted. Their presence does not simply augment conversation but restructures it, drawing participants into forms of interaction mediated by technical logics that reflect non-local, non-relational, and often corporate interests. The next section presents an initial assessment of the risks of having AI agents in group conversations. These risks, while not exhaustive, point to the magnitude of what is at stake.

5. What Is at Stake When AI Joins the Group

AI agents in group conversations do not simply add a new voice; they alter the epistemic architecture of the space. Rather than offering a comprehensive list of potential harms, this section outlines some key risks that help clarify the stakes of this technological shift.

Privacy Regression: From Encrypted Intimacy to Systemic Exposure

Group chats have historically been protected by end-to-end encryption and governed through mutual trust among known participants. Introducing AI agents creates new data flows, new forms of metadata aggregation, and new vectors of surveillance. Adding an AI agent to a group is functionally equivalent to adding an always-listening participant whose presence expands the surface for data capture. Even if encryption remains intact, the agent's ability to access, store, and transmit contextual content introduces privacy regressions that are difficult for users to detect or mitigate (Chou et al., 2024).

The Subtle Rewriting of Decision-Making

AI agents do not merely provide information; they help set the terms on which decisions unfold. As Song et al. (2024) show, such agents can exert social influence, shift user opinions, and reshape deliberative dynamics. By pre-structuring choices, proposing actions, or reducing epistemic friction through generated answers that bypass human deliberation, AI participants can gradually erode collective sense-making, situated interpretation, and the balancing of competing perspectives.

This dynamic also raises concerns about the homogenization of thought and the erosion of local expertise. AI-generated suggestions may appear neutral or authoritative, but they often reflect external datasets, generalized norms, and corporate logics misaligned with the group's relational context. As these outputs gain influence, the center of epistemic gravity may shift from community knowledge to machine-generated outputs.

The Automation of Reasoning and Dialogue

When conversational agents intervene in group decision-making—suggesting compromises, summarizing positions, or identifying areas of disagreement—they risk redefining the very nature of public reasoning. Rather than supporting debate, AI may channel discourse toward algorithmically preferred outcomes or flatten disagreement into artificial consensus.

The major risk is not AI replacing human conversation, but subtly redefining the norms governing deliberation. The ability of conversational agents to reframe conflict or mirror user

dispositions introduces what (Peter et al., 2025) term “anthropomorphic seduction,” in which users defer to AI outputs as epistemically authoritative. If community judgment becomes increasingly filtered through AI’s interpretive frameworks, the result is a narrowing of epistemic diversity and a weakening of the relational forms of collective reasoning that group chats have historically supported.

6. Conclusion

This paper has argued that agency in the context of AI should be understood not as transactional knowledge extraction but as a relational capacity enacted through collective practices. Drawing on Brazil as an example, this paper has explored WhatsApp Groups as one emblematic infrastructure shaped not *only* by corporate strategy but *also* by community creativity, necessity, and relational practice.

However, emerging interventions by states and technology companies threaten to destabilize these epistemic environments. AI’s entry into group conversations may bring efficiency and support coordination, but it also introduces invisible and consequential costs such as privacy regression, epistemic interference, and automation of collective reasoning. If left unexamined, these shifts risk eroding a significant contemporary digital site of situated and relational knowledge.

The central argument of this essay is therefore a call for vigilance and agency. Users and communities are not passive recipients of technological change; they hold the capacity to shape, resist, reinterpret, and redefine the tools that enter their lives. Safeguarding relational knowledge requires acknowledging not only the potential benefits of AI but also its costs, and affirming the responsibility—and the power—of users to set boundaries, push back, and articulate their own norms of participation.

Bibliography

- Avelar, D. (2019, October 30). WhatsApp fake news during Brazil election ‘favoured Bolsonaro.’ *The Guardian*.
<https://www.theguardian.com/world/2019/oct/30/whatsapp-fake-news-brazil-election-favoured-jair-bolsonaro-analysis-suggests>
- Chou, K.-H., Lin, Y.-M., Wang, Y.-A., Li, J. W., Kim, T. H.-J., & Hsiao, H.-C. (2024). Bots can Snoop: Uncovering and Mitigating Privacy Risks of Bots in Group Chats (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2410.06587>
- Edelman. (n.d.). 2024 Edelman Trust Barometer Global Report. Retrieved December 7, 2025, from
<https://www.edelman.com.br/sites/g/files/aatuss291/files/2024-06/2024%20Edelman%20Trust%20Barometer%20-%20Relat%C3%B3rio%20Complementar%20-%20Setor%20de%20tecnologia.pdf>
- Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3), 575. <https://doi.org/10.2307/3178066>
- Levy, I., & Robinson, C. (2023). Principles for a More Informed Exceptional Access Debate. *Lawfare*.
<https://www.lawfaremedia.org/article/principles-more-informed-exceptional-access-debate>
- Lyon, D. (Ed.). (2006). *Theorizing Surveillance* (0 ed.). Willan.
<https://doi.org/10.4324/9781843926818>
- McCarthy, J., & Hayes, P. J. (1981). Some Philosophical Problems from the Standpoint of Artificial Intelligence. In *Readings in Artificial Intelligence* (pp. 431–450). Elsevier.
<https://doi.org/10.1016/B978-0-934613-03-3.50033-7>
- McQuillan, D. (2022). *Resisting AI: An Anti-fascist Approach to Artificial Intelligence* (1st ed.). Bristol University Press. <https://doi.org/10.2307/j.ctv2rcnp21>

- Nemer, D. (2022). *Technology of the Oppressed: Inequity and the Digital Mundane in Favelas of Brazil*. The MIT Press. <https://doi.org/10.7551/mitpress/14122.001.0001>
- Newman, N., Ross Arguedas, A., Robertson, C. T., Nielsen, R. K., & Fletcher, R. (2025). Reuters Institute digital news report 2025. Reuters Institute for the Study of Journalism. <https://doi.org/10.60625/RISJ-8QQF-JT36>
- Omari, J. (2020). Is Facebook the Internet? Ethnographic Perspectives on Open Internet Governance in Brazil. *Law & Social Inquiry*, 45(4), 1093–1112. <https://doi.org/10.1017/lsi.2020.5>
- OpenAI. (2025, November 25). Introducing group chats in ChatGPT. <https://openai.com/index/group-chats-in-chatgpt/>
- OpinionBox. (n.d.). WhatsApp no Brasil: Os principais dados sobre o uso do app. Retrieved December 7, 2025, from https://blog.opinionbox.com/pesquisa-whatsapp-no-brasil/?utm_source=chatgpt.com
- Pasquinelli, M., & Joler, V. (2021). The Noosphere manifested: AI as instrument of knowledge extractivism. *AI & SOCIETY*, 36(4), 1263–1280. <https://doi.org/10.1007/s00146-020-01097-6>
- Peter, S., Riemer, K., & West, J. D. (2025). The benefits and dangers of anthropomorphic conversational agents. *Proceedings of the National Academy of Sciences*, 122(22), e2415898122. <https://doi.org/10.1073/pnas.2415898122>
- Ribes, D., Hoffman, A. S., Slota, S. C., & Bowker, G. C. (2019). The logic of domains. *Social Studies of Science*, 49(3), 281–309. <https://doi.org/10.1177/0306312719849709>
- Ricourte, P. (2019). Data Epistemologies, The Coloniality of Power, and Resistance. *Television & New Media*, 20(4), 350–365. <https://doi.org/10.1177/1527476419831640>
- Ricourte Quijano, P. (2025). Algorithmic Assemblages of Power: AI Harm and the Question of Responsibility. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, Avance en línea, 1–8. <https://doi.org/10.5209/tekn.101084>

- Song, T., Tan, Y., Zhu, Z., Feng, Y., & Lee, Y.-C. (2024). Multi-Agents are Social Groups: Investigating Social Influence of Multiple Agents in Human-Agent Interactions (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2411.04578>
- Statista. (n.d.). Top internet activities Brazil 2024. Statista. Retrieved December 7, 2025, from <https://www.statista.com/statistics/1052520/brazil-internet-activities/>
- Statista. (n.d.). Topic: WhatsApp in Brazil. Statista. Retrieved December 7, 2025, from <https://www.statista.com/topics/7731/whatsapp-in-brazil/>
- Various authors. (2019, May 22). Coalition Letter to GCHQ on the “Ghost Proposal.” Open Technology Institute / New America (hosted). https://newamericadotorg.s3.amazonaws.com/documents/Coalition_Letter_to_GCHQ_on_Ghost_Proposal_-_May_22_2019.pdf