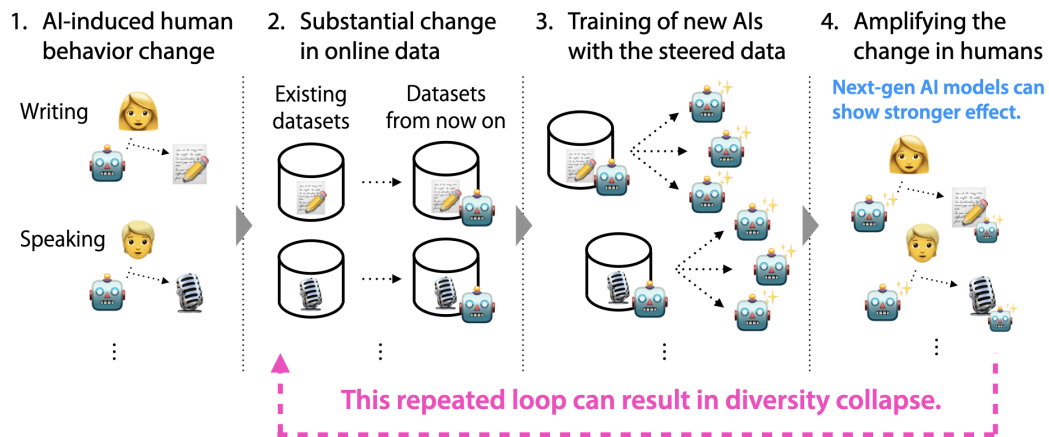


# Envisioning Human Communication after Human-AI Diffusion

Hiromu Yakura, Levin Brinkmann, Iyad Rahwan  
yakura@mpib-berlin.mpg.de  
Center for Humans and Machines  
Max-Planck Institute for Human Development  
Berlin, Germany



**Figure 1: A closed feedback loop between humans and AIs leads to human-AI diffusion: AI-induced behavioral shifts can reshape the online data environment, which then feeds into future models and amplifies subsequent human behavior change.**

## Abstract

Generative AI is increasingly mediating everyday communication. This can trigger *human-AI diffusion*: traits disproportionately produced by large language models (LLMs) diffuse into human cultural activities, and can enter a closed feedback loop when those human outputs become part of the data environment for future model training. Recent work has shown that this diffusion is a present and measurable phenomenon rather than a purely speculative risk. Based on this, this paper envisions two plausible futures: one where human and AI language progressively assimilate, leading to cultural collapse, and another where language becomes a contested social signal, leading to stigma and inefficient avoidance of beneficial tools. Both scenarios highlight the need for training-data and AI-use transparency, which is crucial not only for balancing short-term productivity and (in)equity, but also long-term preservation of cultural plurality.

## CCS Concepts

• **Social and professional topics** → *Computing / technology policy*; **Socio-technical systems**; • **Human-centered computing** → *Collaborative and social computing*.

## Keywords

Large language models, Human-AI feedback loop, social norms

## ACM Reference Format:

Hiromu Yakura, Levin Brinkmann, Iyad Rahwan. 2026. Envisioning Human Communication after Human-AI Diffusion. In *Proceedings of ACM CHI 2026 Workshop on Restoring Human Authenticity in AI-Mediated Communication*. ACM, New York, NY, USA, 4 pages.

## 1 Introduction

Human communication has repeatedly been reshaped by new media and communication technologies. Generative AI, especially large language models (LLMs), introduces a qualitatively new mechanism: the medium does not only transmit human culture, but can also distribute distinctive linguistic patterns at scale. Recent evidence suggests that humans are now adopting such linguistic patterns from LLMs [3, 15]. In our work [15], we find a measurable and abrupt increase in the use of words preferentially generated by ChatGPT in large-scale corpora of human spoken communication after ChatGPT’s release. This kind of lexicographic shift can be understood as an early indicator of a broader process we call *human-AI diffusion*: the diffusion of machine-associated traits into human behavior, even when we are not explicitly intending to imitate AI.

This paper uses that empirical starting point to *envision* plausible trajectories for human communication under human-AI diffusion. Here, we can consider several catastrophic scenarios, e.g., once machine-generated language influences humans, the resulting human outputs become part of the data environment from which future models learn, creating a closed feedback loop (Figure 1). The central question then becomes not whether AI will influence human language, but how this bidirectional feedback loop will evolve—and

what kinds of governance, norms, and technical interventions might steer it toward desirable futures.

## 2 Human-AI Diffusion in Linguistic Communication

While linguistic patterns exhibited by widely deployed LLMs have already spread into human written communication [7, 10], the question is whether they also reach *spontaneous* spoken language. If so, this means that LLM-associated traits are internalized and persisted in human culture, which existing large-scale evidence suggests already does. Specifically, we operationalize *ChatGPT-preferred words* via comparisons between human- and LLM-revised texts, and then track how those words change in real-world speech corpora over time [15]. Notably, we find that conversational podcasts show clear trend changes in multiple ChatGPT-preferred words after the release of ChatGPT in most topic categories (Figure 2). This pattern supports the claim that LLM-associated linguistic patterns can diffuse into everyday speaking, not only into formal writing.

This matters because once diffusion is underway, “AI-like” traits cease to be diagnostic of AI tool use. People can exhibit ChatGPT-associated linguistic patterns because they used an LLM directly, because their collaborators or institutions increasingly use LLMs, or simply because these patterns become ambient in the communicative environment and get internalized through exposure. In other words, we are entering an era in which it becomes increasingly difficult—and sometimes impossible—to infer whether an utterance reflects active AI assistance or human-internalized behavior change. The rest of this paper builds on this empirical premise and explores two plausible futures: one in which human and AI language progressively assimilate (with risks of diversity collapse), and another in which linguistic differentiation emerges as people strategically avoid AI-associated expression (with risks of stigma and inefficiency).

### 3 Scenario 1: Assimilation and Diversity Collapse

Scenario 1 describes a future in which human and AI language progressively assimilate. As this human-AI diffusion progresses, AI-associated linguistic patterns can become default across many contexts, especially in professional and institutional communication. Over time, this can narrow the distribution of expression in public discourse, not only across languages, but also across registers within a language (e.g., formal vs. informal, regional styles, community slang, and domain-specific voices). What begins as “helpful standardization” can therefore translate into a gradual erosion of linguistic diversity, with fewer locally distinctive markers and reduced stylistic variance.

The key risk is that assimilation is not a one-way effect; it can become an enhancing feedback loop (Figure 1). When AI-mediated communication shifts what people write and say, it also shifts what gets recorded, indexed, and reused online. That steered data environment then becomes part of the substrate on which future models are trained. This mechanism extends the concern described as *model collapse* [12], where repeatedly training generative models on their own outputs can cause distributional shrinkage and forgetting of rare modes. In this scenario, if AI-associated linguistic

patterns come to dominate high-visibility corpora, even via human production, future models may become increasingly confident in narrow patterns while becoming less capable of producing (and recognizing) diverse linguistic forms. In turn, rarer forms become even less visible to humans, making the feedback loop self-reinforcing.

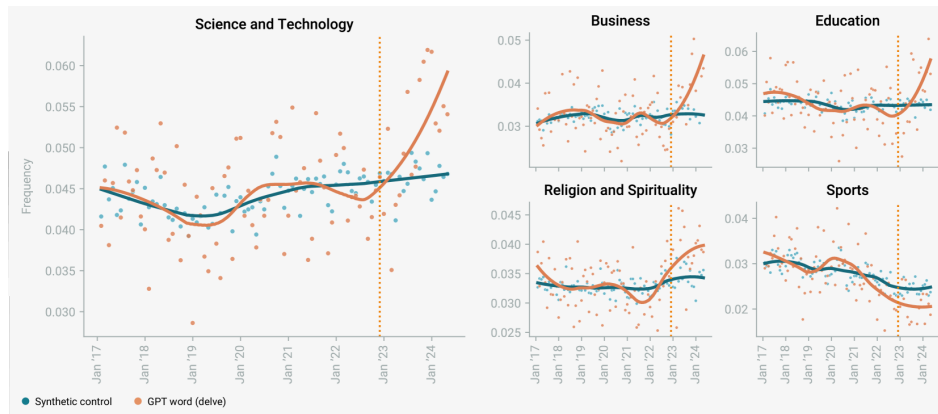
In the extreme, this trajectory can lead to cultural lock-in. For example, if new generations are increasingly educated through dialogue with future models (e.g., tutoring, writing support, etc.), they may acquire the narrowed register as the default and form the belief that it is the standard for “correct” communication. Once such beliefs are institutionalized through ubiquitous AI-mediated interaction, linguistic experimentation and minority patterns can be implicitly discouraged even without explicit policing. In addition, if LLMs are deployed as evaluators in admissions or job applications, applicants may strategically accommodate these systems [4], i.e., by adopting LLM-associated linguistic patterns more broadly, further accelerating diffusion. These futures are damaging for both humans and AI: for humans, it reduces cultural plurality and the space for innovation; for AI, it reduces the diversity of future training data and increases vulnerability to collapse.

### 4 Scenario 2: Differentiation, Avoidance, and Stigma

Scenario 2 describes a future in which Human-AI diffusion does not lead to smooth assimilation, but instead triggers social differentiation. Sociolinguistics has long documented that linguistic variation is not merely “noise,” but a resource for indexing identity and social position [9]. In an AI-mediated world, a similar dynamic could emerge around “AI-like” linguistic patterns: individuals and communities may intentionally avoid patterns perceived as machine-associated. In addition, they may deliberately accelerate linguistic innovation (e.g., slang, meme-terms, semantic shifts, and orthographic play) to keep a moving target of in-group expression that resists standardization [1, 6, 14]. While this can help us avoid cultural collapse, it also poses a risk of stigma and inequity.

Specifically, once diffusion is underway, such signaling becomes structurally unstable. “AI-like” traits are no longer reliable evidence of tool use; they can be internalized through exposure or institutional templates even when no AI was used. If institutions, platforms, or peer groups treat AI-like patterns as a basis for suspicion or devaluation, they risk false accusations and arbitrary boundary-making. These dynamics can have uneven impacts; for example, second-language speakers, who could be more susceptible to AI-associated linguistic patterns, might be more likely to be misclassified or undervalued, reinforcing existing inequities.

Furthermore, differentiation can impose real efficiency and welfare costs. In some settings, AI assistance can improve accessibility and productivity (e.g., language support, drafting, clarification) [8, 11]. But if social norms increasingly punish AI-associated expression, people may avoid useful tools even when their use would be appropriate and beneficial. The result is a persistent tension; as groups compete to maintain linguistic distinctiveness from machines, society may incur unnecessary interpersonal conflict and forego opportunities for equitable, productive communication.



**Figure 2: Trend changes of a ChatGPT-preferred word (i.e., *delve*) in conversational podcasts across topics, compared against synthetic controls, after the release of ChatGPT. Reproduced from Fig. 4 in [15].**

## 5 Discussion

The long-term consequences of human–AI diffusion remain deeply uncertain; we cannot confidently predict whether diffusion will stabilize, diversify, or homogenize human communication. Yet the scenarios outlined above illustrate that catastrophic trajectories are plausible: reinforcing feedback can narrow the space of expression, reshape incentives in education and evaluation, and couple cultural change to future model training. Responding to such risks requires transparency in two different dimensions: training transparency and use transparency.

*Training transparency* is the visibility into what data were collected, filtered, and used for training (and how AI-mediated resources are represented within those pipelines). Without this transparency, it becomes difficult to trace closed feedback loops and therefore difficult to mitigate them through policy or technical intervention. *Use transparency* is the visibility into how, where, and by whom AI systems are used in consequential communication and evaluation. Importantly, transparency policies can backfire; if rules are partial, unevenly enforced, or framed as punitive, they can incentivize clandestine *shadow AI* use rather than genuine accountability [2]. This highlights the challenge of effective policy making, which is now required to consider not only short-term productivity or (in)equity, but also long-term cultural impacts.

Also, these transparency needs naturally connect to preservation. If human communication and model behavior co-evolve, then safeguarding linguistic plurality requires maintaining reference points, such as longitudinal corpora with documented provenance, protected communicative spaces, or datasets that record when and how AI mediation is present. Here, we can learn from decades of work in language documentation and revitalization, which has developed methods and infrastructures for recording dialects and endangered or minority languages before they disappear from everyday use [5, 13]. Existing archiving efforts (e.g., dedicated endangered-language repositories) illustrate practical pathways for long-horizon stewardship, and suggest that preserving less-AI-mediated corpora can be treated as risk management.

## 6 Conclusion

Human–AI diffusion raises the possibility that linguistic traits circulate bidirectionally between humans and LLMs, forming a closed feedback loop with cultural consequences. While the long-run trajectory remains uncertain, we outlined two futures: 1) assimilation that risks diversity collapse and accelerates homogenization, and 2) differentiation that maintains surface distinctiveness but can generate stigma, inequity, and inefficient avoidance of beneficial tools. These scenarios highlight the need to tackle the policy and design challenges of AI-mediated communication, possibly by combining training and use transparency to support equitable access, accountability, and measurement of cultural change.

## Acknowledgments

This work was supported in part by the Japan Science and Technology Agency (JST) through the PRESTO program (JPMJPR246B).

## References

- [1] Michael Adams. 2009. *Slang: The People's Poetry*. Oxford University Press, Oxford, UK. doi:10.1093/oso/9780195314632.001.0001
- [2] Mengchen Dong, Hiromu Yakura, Omar Sherif, Jean-Francois Bonnefon, and Iyad Rahwan. 2025. Shadow AI Thrives Under Punitive Social Evaluation. *SSRN* 5410403 (2025), 1–20. doi:10.2139/ssrn.5410403
- [3] Mingmeng Geng, Caixi Chen, Yanru Wu, Yao Wan, Pan Zhou, and Dongping Chen. 2025. The Impact of Large Language Models in Academia: from Writing to Speaking. In *Findings of the Association for Computational Linguistics, ACL 2025*. Association for Computational Linguistics, Kerrville, USA, 19303–19319. doi:10.18653/v1/2025.findings-acl.987
- [4] Jonas Goergen, Emanuel de Bellis, and Anne-Kathrin Klesse. 2025. AI assessment changes human behavior. *Proceedings of the National Academy of Sciences* 122, 25 (2025), e2425439122. doi:10.1073/pnas.2425439122
- [5] Nikolaus P. Himmelmann. 1998. Documentary and Descriptive Linguistics. *Linguistics* 36, 1 (1998), 161–195. doi:10.1515/ling.1998.36.1.161
- [6] Daphna Keidar, andreas Opedal, Zhijing Jin, and Mrinmaya Sachan. 2022. Slangvolution: A Causal Analysis of Semantic Change and Frequency Dynamics in Slang. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Kerrville, USA, 1422–1442. doi:10.18653/V1/2022.ACL-LONG.101
- [7] Dmitry Kobak, Rita González-Márquez, Emőke-Ágnes Horvát, and Jan Lause. 2025. Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Science Advances* 11, 27 (2025), eadt3813. doi:10.1126/sciadv.adt3813
- [8] Keigo Kusumegi, Xinyu Yang, Paul Ginsparg, Mathijs de Vaan, Toby Stuart, and Yian Yin. 2025. Scientific production in the era of large language models. *Science* 390, 6779 (2025), 1240–1243. doi:10.1126/science.adw3000

- [9] William Labov. 2006. *The Social Stratification of English in New York City* (2 ed.). Cambridge University Press, Cambridge, UK.
- [10] Weixin Liang, Yaohui Zhang, Zhengxuan Wu, Haley Lepp, Wenlong Ji, Xuandong Zhao, Hancheng Cao, Sheng Liu, Siyu He, Zhi Huang, Diyi Yang, Christopher Potts, Christopher D Manning, and James Y. Zou. 2024. Mapping the Increasing Use of LLMs in Scientific Papers. In *Proceedings of the 1st Conference on Language Modeling*. OpenReview, Philadelphia, USA, 1–27.
- [11] Shakke Noy and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381, 6654 (2023), 187–192. doi:10.1126/science.adh2586
- [12] Iliia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross J. Anderson, and Yarín Gal. 2024. AI models collapse when trained on recursively generated data. *Nature* 631, 8022 (2024), 755–759. doi:10.1038/S41586-024-07566-Y
- [13] Hedvig Skirgøard, Hannah J. Haynie, Damián E. Blasi, Harald Hammarström, Jeremy Collins, Jay J. Latache, Jakob Lesage, Tobias Weber, Alena Witzlack-Makarevich, Sam Passmore, Angela Chira, Luke Maurits, Russell Dinnage, Michael Dunn, Ger Reesink, Ruth Singer, Claire Bowern, Patience Epps, Jane Hill, Outi Vesakoski, Martine Robbeets, Noor Karolin Abbas, Daniel Auer, Nancy A. Bakker, Giulia Barbos, Robert D. Borges, Swintha Danielsen, Luise Dorenbusch, Ella Dorn, John Elliott, Giada Falcone, Jana Fischer, Yustinus Ghanggo Ate, Hannah Gibson, Hans-Philipp Göbel, Jemima A. Goodall, Victoria Gruner, Andrew Harvey, Rebekah Hayes, Leonard Heer, Roberto E. Herrera Miranda, Natalia Hübner, Biu Huntington-Rainey, Jessica K. Ivani, Marilen Johns, Erika Just, Eri Kashima, Carolina Kipf, Janina V. Klingenberg, Nikita König, Aikaterina Koti, Richard G. A. Kowalik, Olga Krasnoukhova, Nora L. M. Lindvall, Mandy Lorenzen, Hannah Lutzenberger, Tânia R. A. Martins, Celia Mata German, Suzanne van der Meer, Jaime Montoya Samamé, Michael Müller, Saliha Muradoglu, Kelsey Neely, Johanna Nickel, Miina Norvik, Cheryl Akinyi Oluoch, Jesse Peacock, India O. C. Pearey, Naomi Peck, Stephanie Petit, Sören Pieper, Mariana Poblete, Daniel Prestipino, Linda Raabe, Amna Raja, Janis Reimringer, Sydney C. Rey, Julia Rizaew, Eloisa Ruppert, Kim K. Salmon, Jill Sammet, Rhiannon Schembri, Lars Schlabbach, Frederick W. P. Schmidt, Amalia Skilton, Wikaliler Daniel Smith, Hilário de Sousa, Kristin Sverredal, Daniel Valle, Javier Vera, Judith Voß, Tim Witte, Henry Wu, Stephanie Yam, Jingting Ye, Maisie Yong, Tessa Yuditha, Roberto Zariquiey, Robert Forkel, Nicholas Evans, Stephen C. Levinson, Martin Haspelmath, Simon J. Greenhill, Quentin D. Atkinson, and Russell D. Gray. 2023. Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances* 9, 16 (2023), eadg6175. doi:10.1126/sciadv.adg6175
- [14] Peng Wang, Xilin Tao, Siyi Yao, Jiageng Wu, Yuntao Zou, Zhuotao Tian, Libo Qin, and Dagang Li. 2026. Can Large Language Models Resolve Semantic Discrepancy in Self-Destructive Subcultures? Evidence from Jirai Kei. *arXiv* 2601.05004 (2026), 1–29. doi:10.48550/ARXIV.2601.05004
- [15] Hiromu Yakura, Ezequiel Lopez-Lopez, Levin Brinkmann, Ignacio Serna, Prateek Gupta, and Iyad Rahwan. 2024. Empirical evidence of Large Language Model’s influence on human spoken communication. *arXiv* 2409.01754 (2024), 1–29. doi:10.48550/ARXIV.2409.01754