

AI Slop III: Society and Model Collapse

Jun-E Tan



Introduction

AI slop – a flood of low quality AI-generated content – pollutes our communication environment with noise that drowns out important signals.

In this article, I explore some systemic effects of AI slop, first from the angle of society at large, with a particular focus on democracy and science as institutions that have been disrupted. Next, I look into the concept of model collapse, where AI models trained on AI-generated content gradually degrade to the point of failure.

This is the third and final article within the AI slop series. Interested readers can refer to Part 1 for examples of AI slop in various genres of media and a discussion on the incentives that drive its proliferation. Part 2 addresses consequences of AI slop on content consumers and content creators.

Views are short opinion pieces by the author(s) to encourage the exchange of ideas on current issues. They may not necessarily represent the official views of KRI. All errors remain the authors' own.

This view was prepared by Dr Jun-E Tan, a researcher from the Khazanah Research Institute (KRI). The author is grateful for research assistance from Dinie Joshry Shaeel.

Author's email address:
june.tan@krinstitute.org

Attribution – Please cite the work as follows: Jun-E Tan. 2025. AI Slop III: Society and Model Collapse. Kuala Lumpur: Khazanah Research Institute. License: Creative Commons Attribution CC BY 3.0.

Photo by Justin Ha on Unsplash

Information on Khazanah Research Institute publications and digital products can be found at www.KRIInstitute.org.

Society at large

AI slop leaching into the communication and media environment has generated immense impacts at the societal level. As argued in my previous work^{1,2}, societal harms associated with AI can be viewed with a *relational* perspective, where the unit of analysis of harm³ shifts from the level of individuals to *relationships or connections* between them.

It is therefore important to isolate the impacts of AI slop from effects on individuals (e.g. creators or consumers) to also focus on macro-level disruptions on social structures that are fundamental to the functioning of modern society. **AI slop changes the nature and quality of relationships in societies, and erodes the strength of institutions** as the information environment becomes crowded with AI-generated content. We offer two examples.

AI, democracy and public trust

Democracy as a system of government - through power conferred on elected representatives, or directly held by citizens themselves - relies on citizens having access to accurate and timely information to make informed decisions⁴. The information intake is not only important at an individual level, but also at a community level, where factual information ideally creates a shared social reality to support public discourse and policy deliberations.

With the proliferation of cheaply generated content, global elections in recent years have had to contend with AI-generated content. For example, in 2024, a year unusually concentrated with general elections globally, researchers tracked 50 elections and found that 80% of these elections had incidents connected to gen-AI, with 215 incidents logged. Out of these incidents, more than two thirds (69%) had “a harmful role” in the election⁵.

Politics and elections are therefore a fertile ground for AI-enabled disinformation campaigns. However, the more insidious effects of AI slop does not require the intent of deception and manipulation. With a communication and media environment overwhelmed with low quality, mass-generated content, the larger concern here is *not* that people will believe in false information, but that they will find it difficult to believe what they see and read at all.

The uncertainty about the veracity of information can result in people defaulting into existing biases and identity politics, tearing apart the social fabric that enables inter-group engagement that builds consensus and a shared vision of the future. The general lack of public trust also undermines the legitimacy of journalistic content and democratic processes, further exacerbating the problem in a vicious cycle.

The rapid improvement of gen-AI models means that the convincingness of synthetic media continues to increase, while the cost of creating such content plummets to almost zero. OpenAI's

¹ Tan (2023)

² Tan (2024)

³ Smuha (2021)

⁴ UNDP (2024)

⁵ Trauthig et al. (2025)

Sora 2, for example, is a case in point. Released in October 2025 with minimal guardrails⁶, the ability to generate realistic short videos with the likenesses of real (and deceased) people is put into the hands of millions, truly propelling us into a post-truth era.

AI's impact on science

Another example of AI slop's societal effects can be found in the field of science, a bedrock of human understanding of the world. In a cogently argued article, Princeton scholars Sayash Kapoor and Arvind Narayanan ask the question, "Could AI slow science?" They answer:

"Even if individual scientists benefit from adopting AI, it doesn't mean science as a whole will benefit. [...] So far, on balance, AI has been an unhealthy shock to science, stretching many of its processes to the breaking point⁷."

To set the scene, Kapoor and Narayanan cite an array of literature from the field of metascience (the science of science), showing that scientific progress has been slowing down despite increased production of scientific papers and studies. While the rate of paper publication has increased 500-fold between 1900 and 2015, the rate of genuine breakthroughs that have advanced leaps in understanding has either remained constant or even slowed down, depending on the measure used. This is what is known as the "production-progress paradox".

There may be different explanations for this paradox, but the researchers suggest that the most likely reason is that *the increase in production itself is slowing down scientific progress*. The volume of papers creates a lot of noise that drowns out the important signals of truly novel work, and top-cited papers which do not depart from the canon get most of the attention. Importantly, the imperative to publish (or perish) also reduces the time spent in developing breakthrough work, incentivising authors to work on less risky and more established areas.

The production-progress paradox is a systemic issue, existing before the generative AI boom. Post boom, in recent years, Large Language Models (LLMs) have been found to be widely used in academic papers published⁸, remarkably scaling the productivity of scientists⁹. The increase of subtle errors, characteristic of AI slop, compounds the problem. Traditional means of quality control, such as the peer review process, cannot cope with the volume and are also getting encroached by AI (with AI being used to review papers), creating questions of trust in the institution of science¹⁰.

Attempting to adapt to the situation, some scientific institutions including universities and academic publishers have built policies for AI use and guidelines for disclosure of use. While this is a good place to start, a study on thirty such policies in American universities ultimately

⁶ Valinsky (2025)

⁷ Kapoor (2025)

⁸ Liang et al. (2024)

⁹ Somers (2023)

¹⁰ Chalboub (2025)

concludes that currently available institutional guidance places the onus of accountability and compliance on individual researchers¹¹, something that may not be sustainable in the long term.

In any case, the space continues to evolve. Kende Kefale, a researcher from the University of Capetown suggests that in the longer term, universities may evolve from being “[places] where knowledge is primarily *generated* to [places] where it is rigorously *validated*” (emphasis in original) to remain relevant. For that, curricula have to be radically adapted; tenures and promotions have to assess and reward success differently; and a university’s brand will hinge on its reputation for rigour¹².

AI companies and their models

Beyond the impacts to societal systems, the overwhelm of synthetic content can even negatively impact the generative AI models themselves. Many of these models are trained on vast datasets, in many cases scraped from the internet.

As has been demonstrated^{13,14}, post 2022, AI slop has encroached into most of our online spaces, and is difficult to differentiate from human-made data. This means that current and future attempts to trawl the web for data to further train LLMs will inevitably include AI slop into training datasets.

“Model collapse” refers to **a gradual degradation of performance of gen-AI models when they are trained on AI-generated content**. In a widely cited paper in *Nature*, Shumailov and colleagues, who coined the term, find that “indiscriminate use of model-generated content in training causes irreversible defects in the resulting models, in which tails of the original content distribution disappear”¹⁵.

One way to think about this is making photocopies of photocopies, where the fidelity to the original gets progressively less after each round of copying. In early stages of model collapse, data that occur less within the dataset will be lost. At a late stage, the data distribution would converge so much that it would look nothing like the original data, baked in with compounded errors that were generated in previous iterations¹⁶.

It is in the companies’ interest to ensure that there is enough of diverse human-made content that has gone through adequate quality control, to train their next generations of models¹⁷. There is already talk of “peak data”, where AI companies run out of online content that can be scraped for model training¹⁸.

¹¹ Ganguly et al. (2025)

¹² Kefale (2025)

¹³ Tan (2025a)

¹⁴ Tan (2025b)

¹⁵ Shumailov et al. (2024)

¹⁶ Gomstyn and Jonker (2024)

¹⁷ *The Straits Times* (2025)

¹⁸ Cuthbertson (2025)

To somewhat mitigate this risk (and the risk of copyright lawsuits), leading AI companies now have licensing deals with major media companies, to use their content in exchange for visibility, attribution and compensation¹⁹. However, the scale of these mitigations may not be sufficient for the amount of training data needed. Smaller independent publishers who do not have access to these deals will disappear or struggle to remain afloat, as they lose traffic from search engine click-throughs with the rise of AI search²⁰.

Conclusion

While numbers on productivity gains from gen-AI have been projected and bandied about, we need to recognise that those benefits will not be evenly distributed and there will be short term shocks and long term consequences on all. In this article series on AI slop, I have sought to expand on these impacts from multiple angles, to illustrate the present and potential outcomes of minimising the barriers to content creation and flooding the communication environment with low quality AI media.

Discussions on negative consequences of AI-generated content should move beyond the narrow focuses of malicious use and material impacts on creatives. Much more research needs to be done on societal and institutional effects of AI slop, acknowledging that the consequences can come from thoughtless and unfettered use of gen-AI as well. This is especially urgent outside the context of the Global North, where technologies for AI detection work less well²¹, and cultural contexts and digital literacy levels are different.

Adapting to the post-truth world becomes increasingly crucial as a matter of societal resilience. Governments and institutions will have to put in place governance and regulatory mechanisms, such as setting in place AI use guidelines or regulations, labeling AI content, or allocating more resources to AI detection. Widespread awareness and literacy campaigns will also need to be run, to ensure that the public is equipped with critical thinking and fact checking skills.

The strength of AI models is directly linked to the strength of the information and media ecosystem. As AI companies create major disruptions in our global communication environment, they would do well to consider auxiliary impacts of their actions, if only to ensure that continual model improvement continues to be possible in the longer term. This of course has to be couched within a larger imperative, to build technological innovations that are responsibly developed and safely deployed, with outcomes that benefit society as a whole.

¹⁹ Duffy (2025)

²⁰ Benjamin (2025)

²¹ Elliott (2024)

References

- Benjamin, Jack. 2025. "One-Third of Indie Publishers Could Shut down by next Year as AI Search Hits Traffic." *The Media Leader* (blog). September 19, 2025. <https://uk.themedialeader.com/one-third-of-indie-publishers-could-shut-down-by-next-year-as-ai-search-hits-traffic/>.
- Chalboub, George. 2025. "We Must Set the Rules for AI Use in Scientific Writing and Peer Review." *Times Higher Education* (THE). October 2025. <https://www.timeshighereducation.com/opinion/we-must-set-rules-ai-use-scientific-writing-and-peer-review>.
- Cuthbertson, Anthony. 2025. "AI Has Run out of Training Data, Warns Data Chief." *The Independent*. October 3, 2025. <https://www.independent.co.uk/tech/ai-peak-data-goldman-sachs-b2838795.html>.
- Duffy, Katie. 2025. "A Complete List of Publishers and Their AI Licensing Deals." *Futureweek* (blog). September 4, 2025. <https://futureweek.com/a-complete-list-of-publishers-striking-ai-content-licensing-deals/>.
- Elliott, Vittoria. 2024. "AI-Fakes Detection Is Failing Voters in the Global South." *Wired*, September 2, 2024. <https://www.wired.com/story/generative-ai-detection-gap/>.
- Ganguly, Amrita, Aditya Johri, Areej Ali, and Nora McDonald. 2025. "Generative Artificial Intelligence for Academic Research: Evidence from Guidance Issued for Researchers by Higher Education Institutions in the United States." *AI and Ethics* 5 (4):3917–33. <https://doi.org/10.1007/s43681-025-00688-7>.
- Gomstyn, Alice, and Alexandra Jonker. 2024. "What Is Model Collapse?" IBM. October 14, 2024. <https://www.ibm.com/think/topics/model-collapse>.
- Kapoor, Sayash. 2025. "Could AI Slow Science?" April 15, 2025. <https://www.normaltech.ai/p/could-ai-slow-science>.
- Kefale, K. 2025. "Remaking Institutional Research in the Age of Generative AI." *Systems Theory for Institutional Research* (blog). July 26, 2025. <https://syres.uct.ac.za/2025/07/remake-ir-for-ai/>.
- Liang, Weixin, Yaohui Zhang, Zhengxuan Wu, Haley Lepp, Wenlong Ji, Xuandong Zhao, Hancheng Cao, et al. 2024. "Mapping the Increasing Use of LLMs in Scientific Papers." arXiv. <https://doi.org/10.48550/arXiv.2404.01268>.
- Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. 2024. "AI Models Collapse When Trained on Recursively Generated Data." *Nature* 631 (8022). Nature Publishing Group:755–59. <https://doi.org/10.1038/s41586-024-07566-y>.
- Smuha, Nathalie A. 2021. "Beyond the Individual: Governing AI's Societal Harm." *Internet Policy Review* 10 (3). <https://doi.org/10.14763/2021.3.1574>.
- Somers, Meredith. 2023. "How Generative AI Can Boost Highly Skilled Workers' Productivity." MIT Management Sloan School. October 19, 2023. <https://mitsloan.mit.edu/ideas-made-to-matter/how-generative-ai-can-boost-highly-skilled-workers-productivity>.

- Tan, Jun-E. 2023. "Visualising Societal Harms of Artificial Intelligence." *Khazanah Research Institute Publications*, November. <https://www.krinstitute.org/publications/visualising-societal-harms-of-artificial-intelligence>.
- . 2024. "Algorithmic Management and Societal Relations: The Plight of Platform Workers in Southeast Asia." *Khazanah Research Institute Publications*, August. <https://www.krinstitute.org/publications/algorithmic-management-and-societal-relations-the-plight-of-platform-workers-in-southeast-asia>.
- . 2025a. "AI Slop I: Pollution in Our Communication Environment." *Khazanah Research Institute Publications*, October. <https://www.krinstitute.org/publications/ai-slop-i-pollution-in-our-communication-environment#citation>.
- . 2025b. "AI Slop II: Content Consumers and Creators." *Khazanah Research Institute Publications*, November. <https://www.krinstitute.org/publications/ai-slop-ii-content-consumers-and-creators>.
- The Straits Times*. 2025. "Tech Companies Are Turning to 'Synthetic Data' to Train AI Models – but There Is a Hidden Cost," January 13, 2025. <https://www.straitstimes.com/opinion/tech-companies-are-turning-to-synthetic-data-to-train-ai-models-but-there-is-a-hidden-cost>.
- Trauthig, Inga, Philip N. Howard, Kate Dommett, and Dounia Mahloully. 2025. "The Role of Generative AI Use in 2024 Elections Worldwide." International Panel on the Information Environment (IPIE). <https://doi.org/10.61452/HZUE9853>.
- UNDP. 2024. "Information Integrity for Electoral Institutions and Processes: Reference Manual for UNDP Practitioners." United Nations Development Programme. https://www.undp.org/sites/g/files/zskgke326/files/2024-03/24119_undp_information_integrity_v08_rc_002.pdf.
- Valinsky, Jordan. 2025. "OpenAI's Sora Bans Martin Luther King Jr. Deepfakes after His Family Complained | CNN Business." CNN. October 17, 2025. <https://www.cnn.com/2025/10/17/tech/openai-mlk-sora-pause>.