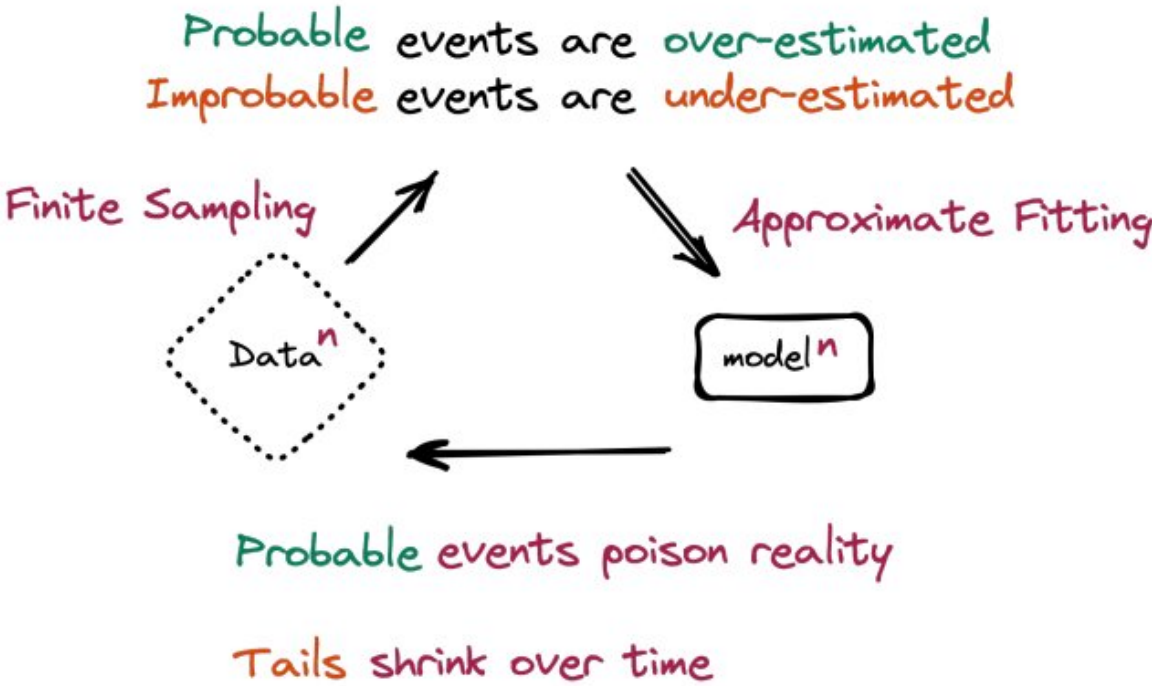


AI models feeding on AI data may face death spiral

June 19 2023, by Peter Grad



Model Collapse refers to a degenerative learning process where models start forgetting improbable events over time, as the model becomes poisoned with its own projection of reality. Credit: *arXiv* (2023). DOI: 10.48550/arxiv.2305.17493

Large language models are generating verbal pollution that threatens to

undermine the very data such models are trained on.

That's the conclusion reached by a team of British and Canadian researchers exploring the impact of successive generations of ChatGPT generated text that will be culled for future models.

In a paper published on the *arXiv* preprint server and titled, "The Curse of Recursion: Training on Generated Data Makes Models Forget," the team predicted that the recursive nature of AI training will eventually lead to "model collapse."

"We discover that learning from data produced by other models causes model collapse—a degenerative process whereby, over time, models forget the true underlying data distribution," the team said.

Team member Ross Anderson, of University of Cambridge and University of Edinburgh, likened the effect to the diminishing quality of musical output.

"If you train a music model on Mozart," he said in a personal blog, "you can expect output that's a bit like Mozart but without the sparkle ...and if [that version] trains the next generation, and so on, what will the fifth or sixth generation sound like?"

The authors note that model collapse is a threat similar to catastrophic forgetting and data poisoning.

In catastrophic forgetting, a model "forgets" previous data, sometimes abruptly, when learning new information. The impact is compounded over time.

In their new research, the team said, models don't forget previously learned data "but rather start misinterpreting what they believe to be real,

by reinforcing their own beliefs."

Data poisoning is the malicious insertion of false information. Of course, this practice predated the use of large language models. But with the use of large-scale web crawls, the insertion of even a small amount of malicious data, the team said, can lead to widespread contamination.

"What is different with the arrival of large language models is the scale at which such poisoning can happen once it is automated," the team said.

Researcher Ilya Shumailov, of the University of Oxford, warned that "major degradation happens within just a few iterations, even when some of the original data is preserved."

"Errors from optimization imperfections, limited models and finite data," he continued, "ultimately cause synthetic data to be of low[er] quality. Over time mistakes compound and ultimately force models that learn from generated data to misperceive reality even further."

The researchers said that the nature of recursive learning is to dispense with low-probability events, referred to by statisticians as "tails of the distribution"

In his blog, Anderson warned, "using model-generated content in training causes irreversible defects. The tails of the original content distribution disappear. Within a few generations, text becomes garbage."

"Low-probability events are ... vital to understand [complex systems](#)," the report noted.

The first large language models were trained on human-generated text. But with the rapid adoption of ChatGPT by industry and general users, enormous amounts of data are populating online sites.

The researchers urged that steps be taken to distinguish AI content from human-generated content and that efforts be made to preserve original content for future training purposes.

"Large language models are like fire," team member Anderson said, "a useful tool, but one that pollutes the environment. How will we cope with it?"

More information: Ilia Shumailov et al, The Curse of Recursion: Training on Generated Data Makes Models Forget, *arXiv* (2023). [DOI: 10.48550/arxiv.2305.17493](https://doi.org/10.48550/arxiv.2305.17493)

© 2023 Science X Network

Citation: AI models feeding on AI data may face death spiral (2023, June 19) retrieved 23 February 2024 from <https://techxplore.com/news/2023-06-ai-death-spiral.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.