Regenerative AI? Regurgitating AI? CHATGentropyT?
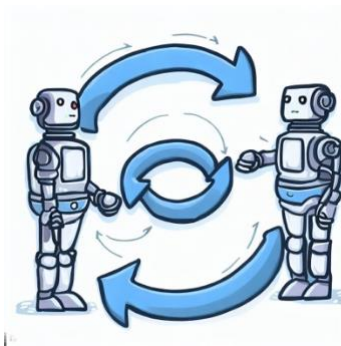
By Eric E. Cohen

Eric E. Cohen, CPA, is a technologist with a passion for collaboration toward the goal that "a piece of business information, once entered into any system, anywhere, never needs to be retyped as it moved through the business reporting supply chain." He's also a prolific author, engaged in virtually every effort to standardize accounting and audit data, a national expert to a wide variety of standards efforts, and co-founder of XBRL.

I really enjoy ChatGPT and its friends. As a sole practitioner, its nice to have an (artificial) colleague to run my thoughts past and help me organize them, even if they are just said back to me a slightly different way sometimes. Sadly, I learn more when I teach others than when I am taught, so the experience is fairly natural to me.

But what happens when the output of AI is based on being trained on greater and greater amounts of prior output of AI? When generative AI is responsible for more and more of the content being created, so future AI versions will be trained on the outcome of existing AI output? Is it a problem? Can AI get caught in a feedback loop, and hit "model collapse"?

This is a question raised in a recent article from "Venture Beat" found at https://venturebeat.com/ai/the-ai-feedback-loop-researchers-warn-of-model-collapse-as-ai-trains-on-ai-generated-content/. The article notes that, in six months, more people are using and relying on the output of generative AI (ChatGPT, Bard, et al.), leaving less human-generated content on which future AIs might feed. A group of researchers put out their thoughts on "The Curse of Recursion: Training on Generated Data Makes Models Forget." https://arxiv.org/abs/2305.17493v2, https://arxiv.org/pdf/2305.17493v2.pdf.

The journal article is filled with mathematical equations, so I can't easily follow their proofs, but the second law of thermodynamics (closet physics major) speaks to the direction of natural processes, that entropy will always increase over time (changes toward disorder are overwhelming more likely than those toward order).

As a (poor, but enthusiastic) musician, I use technology to take music I hear and capture it as notation. Through a process called quantization, the free and expressive imprecisions in performance are shoved into a grid, made to conform with more regular patterns. Expressiveness can be lost.

Quantizing isn't just loss in music. The same thing goes with an original goal of XBRL – instead of forcing reporting distinctives into 800 fixed reporting categories, let companies report what they want, and then associate it with known quantities, a process now known as "anchoring."
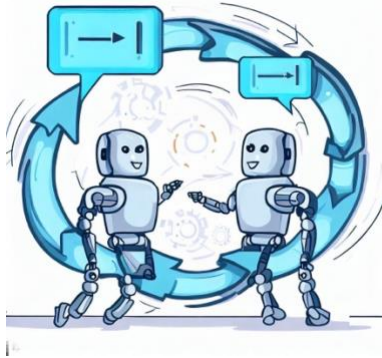
I can't say whether AI output is "quantized" per se. I can tell you that researchers who asked ChatGPT to tell them a joke 1000 times, it told the same 25 jokes some large portion of the time. "Over 90% of 1,008 generated jokes were the same 25 jokes." https://arstechnica.com/information-technology/2023/06/researchers-discover-that-chatgpt-prefers-repeating-25-jokes-over-and-over/.

But let's look at the original corpus of training material. Some says it is already filled with "noise, bias, and low-quality content." There is the old saw that "The entire realm of human knowledge passes through Internet chat rooms every four days." There's also the old saw that "The average person uses only 10% of their brain. And with the average Internet poster, that seems absolutely obvious."

Is it possible we aren't headed to model collapse … but if humans are adding energy to the system through selecting and (re)publishing the generative AI content, perhaps that's enough energy to stop the collapse.