

1 I would be so ashamed to use generative AI, heres why

Generative AI ("genAI") is not harm-free and I do not think the benefits outweigh the harm in its current iteration. I compiled my thoughts about genAI here to keep all the cited sources straight and points in one place. Is the title and thumbnail image harsh? Yes, and by design— as you can see, I think genAI is harmful and that we need to be clear about why using it is harmful. This is my first pass, Ill keep adding better sources as I find them.

1.1 What is genAI?

Generative "artificial intelligence" often shortened to "generative AI" or "genAI", are data models that produce novel text, images, videos, audio, and software code. You may recognize the name of genAI apps, which include Twitter/Xs Grok, Anthropics Claude, Microsofts Copilot, DeepSeek, Googles Gemini and Veo, Stability AIs Stable Diffusion, Midjourney, Lightricks LTX, and OpenAIs ChatGPT, DALL-E, and Sora. These models use machine learning classification and deep neural networks to learn underlying patterns and structures of the data on which they are trained. In simpler terms: these models take in extremely large amounts of data, analyze them to "learn" repeatable patterns, then produce something that matches those patterns when asked a question or fed a prompt by a user. These are pattern recognition and pattern regurgitating machines.

1.2 What isn't genAI

genAI is not true intelligence: Artificial intelligence is a catch-all and (Id argue) intentionally confusing term. These models are not intelligent in that they can think like humans do they regurgitate patterns based on their training data. Those patterns are only as strong as their training data. Large-language models, like those used by many popular apps, are often trained on a massive amount of data from the internet. For example, Grok (the Twitter/X chat app) is trained on web pages, text extracts, and real time tweets/posts on X. I dont personally trust the veracity of the information posted on Twitter/X, and I dont think that you should either.

genAI is not an expert: genAI is a pattern machine; it does not have a human experts deep knowledge of a particular topic or subfield. If you want surface level patterns, genAI may be fine, but Id still argue a human-curated site like Wikipedia would be more accurate more often.

genAI does not learn or understand: Similarly, some experts suggest avoiding language that labels model training as "learning," since they do not actually learn in the sense of knowing or understanding. Rather, these pattern machines produce output that resembles the patterns in their training data, which can include generating plausible-looking results that are completely fabricated, such as invented citations. This is commonly called model "hallucinations," although I would argue that this terminology still relies too heavily on company propaganda trying to convince us that these models have brain-like "intelligence" and "learning" and can therefore experience brain-related symptoms. You cannot train away these "hallucinations" of made-up data (despite users who claim that instructing the model to not invent sources could work); that's not how large language models work. They do not "understand" their sources, they do not "know" what is or is not made up, they cannot "learn" to avoid pitfalls, and they certainly do not have the capacity for true emotion or understanding, just the kind of programming that allows them to mimic the kind of output to suggest that they do (and the incentive to do it, as companies want us to connect with their products as "friends" or "partners").

Not all MLMs are genAI: machine learning models are widely used to analyze data, well beyond genAI apps. For instance, I use regression trees (a type of machine learning model) in my ecological conservation research. How is this different from genAI? Unlike these large language models that train across broad (typically stolen) sources, I train my research models on data that I have permission to use. These are data that I have verified as consistent, accurate, and precise before I ran them through the machine learning model. I have also used similar models to analyze wildlife images/videos for my research; again, I controlled the training data and parameters, so I knew exactly what the limitations of the output would be. Examples of machine learning models (AI) I use every day but did not build myself include the iNaturalist/Seek algorithm and the Merlin algorithm. iNaturalist/Seek uses an AI to identify living things from photos; the training data are images submitted to iNaturalist and verified by at least one other observer (real people verifying the accuracy of the training data!). Merlin uses an AI to identify birds from photos and audio recordings; the training data are photos and audio submitted to eBirds Macaulay library (verified by experts). When you see articles about the benefits of AI, ask yourself how often these are genAI/large language models or more limited machine learning models! Many people have told me that AI isn't all bad, because of scientific advancements made by AI; in almost all cases, those advancements were made with narrower machine learning models, not these general access genAI apps. The rest of this post is about those genAI/large language models, not machine learning models more generally. However, it is important to note that

non-genAI machine learning models can also be used in harmful ways (for surveillance, to replace workers, to produce biased results, etc.), and these too should be regulated and used with caution.