

I am an artist and author, and my answers reflect my experience as such. I have studied computer science, but do not focus on AI training except at the abstract level.

1. AI generations should not be a product that can be sold, and should not have a copyright. Common image and chat AIs are trained on any data they can get their hands on from all across the internet, regardless of the copyright status of its source. OpenAI specifically says they use public information from the wider internet (A). Though its specifics seem to be a company secret, the post simply states that it doesn't take information from beyond paywalls. However, even those things on the open internet still have some form of copyright. My art, even if posted publicly, is still my own.

As it currently is, this policy of scraping everything public for generators means that artists and authors are scrambling to lock down everything they've made just to keep it from being put into generators. If a work gets put into a generator, it can be spit right back out to consumers without a single penny or credit ever reaching the creator. It's terrifying.

2. The AI available today encroaches into my sectors specifically. They are made and trained on writing and art, and output the same. Any work I may be paid for as a concept artist, graphic designer, article author, or otherwise can be accomplished quicker and cheaper by asking an AI to generate something that is based partially on my own work anyway.

In personal experience, AI is also harming academic circles. My professor in a literature course was driven to paranoia when she couldn't tell whether our essays had been written by AI or not. This culminated in her attempting to fail the entire class on plagiarism accusations. It took a fight with the college's admin in order to convince her otherwise.

3. I have linked all my sources at the end of my response, corresponding to the letters in their inline citations. Most of my sources come from public news sites. Though not peer reviewed academic sources, these publishers are generally well known and trustworthy to a point. Further, AI's prevalence in the news demonstrates that it is a prevalent issue for the general public.

4. AI is such a new area that few places have even put laws out yet. The EU's AI act is the most substantial law I can find, and its ambitions seem worthy of repeating (C). In practice, the law seems vague and subjective, so a hard copyright law may be better suited and less ambiguous. As something that widely exists on the internet, international laws on the subject will absolutely come into play, similarly to the popups about cookies appearing worldwide following the 2018 California Consumer Privacy Act (D) and other regulations. An AI trained in the US may well take data from foreign individuals, and vice versa.

5. Legislation regarding AI is absolutely necessary. AI works are derivatives of those created by humans, and are currently reused and even monetized (B) without any credit to the producers supplying the training data. Training data is being harvested against the wishes of those making it (E). AI should not be usable as a product in place of human works, be those images, articles, computer programs, or otherwise.

6. As stated above, AI training can use anything available on the public internet. Currently, these datasets bank on the principle of Fair Use (F) in order to do so. To my understanding, data is generally gathered automatically via web crawlers, because the required sets are so massive. OpenAI's GPT models can also gather further data when given a prompt to do so (G).

In conclusion, the new wave of AI is riding off of millions of people's copyrighted works without giving any credit or compensation. It is already having devastating effects on the industries where writing and art are necessary. AI is currently being used with zero respect given to those it is learning from in the first place, and this needs to change.

- (A) <https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed>
- (B) <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>
- (C) <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- (D) <https://www.latimes.com/business/technology/story/2021-09-01/what-are-website-cookies-how-do-they-impact-internet-data>
- (E) <https://arstechnica.com/information-technology/2022/12/artstation-artists-stage-mass-protest-against-ai-generated-artwork/>
- (F) <https://hai.stanford.edu/news/reexamining-fair-use-age-ai>
- (G) <https://platform.openai.com/docs/guides/gpt-best-practices/tactic-use-embeddings-based-search-to-implement-efficient-knowledge-retrieval>
- (H) <https://arxiv.org/abs/2203.02155>