

Comments on the subject of AI and copyright, questions 1-8 and 18-21 by an artist and technical artist working in the industry of digital media and games.

General Questions

<https://www.federalregister.gov/d/2023-18624/p-66>

1. The benefits of this technology lies in rapid creation of iterations of any kind of media and digital goods. Depending on the end products needs this technology can save time and money. The risks of this technology are extensive, though. In every industrialized use-case it is inherently anti-working class and can easily lead to devaluing of work and mass job loss in any digital industrie as well as analytical, assistant and in part administrative job positions in any field. AI generators also streamline abuse and propaganda for the general public.

2. -

3. -

4. China is banning AI generated media without watermarks to clarify the source of generated content while also determining what kind of content generators are allowed to produce. Italy temporarily banned access to ChatGPT because of OpenAI's massive data collection and possible breach of security and privacy regulations. The EU started deploying laws regarding AI, especially generative AI, that demand transparency over AI training data, functioning and output as well as categorizing restrictions by how dangerous the AI models are to human health and rights. The potential dangers of this technology are apparent and require regulation quickly. While this needs time to develop, temporary bans are sufficient to contain the damage this technology is already causing.

5. AI technology reached a point where new legislation might be warranted.

Legislation needs to entail:

- Transparency and restrictions regarding data collection to protect personal rights and property and determine usage capability and functionality of AI models.
- Distribution of software to regulate how easy it is to spread an already trained model and violate laws with rogue applications.
- Stricter requirements for models to be contained, edited or destroyed to remain in control in cases of rights violations or unauthorized distribution.

Training

<https://www.federalregister.gov/d/2023-18624/p-73>

6. Every piece of media that is eligible for copyright-protections is also in danger of being scraped for AI training. All kinds of digital media is collected for the training of AI generators, depending on the functionality of the model. Articles, papers, books etc for language models. Photos, paintings, graphics etc for image generators.

6.1. The internet is the main source for data collection. Depending on the required data (images or text), every website accessible via internet is a target for scraping.

Third-party entities provide the bulk of training material. There is next to no website exclusively servicable to train AI models, data is scraped from third-parties who provide other services unrelated to AI training (galleries, libraries etc).

6.2. To my knowledge there is no licensing model effectively being used for AI training data. AI is trained without regarding licensing.

6.3. Non-copyrighted material is used in the same extent as copyrighted material as copyright is being disregarded by AI developers. Training data is not created or commissioned by AI developers, due to the magnitude of data necessary to make a model functional.

6.4. -

7.1 Training data is analyzed, checked for patterns and replicated. Data is filtered through a matrix of weights, which determine the output. Output is analyzed, checked against the desired outcome and used to correct weights on which data is replicated.

Output is always replicated data, the smaller the data set is, the more likely the replication will produce copies.

Rights of reproduction of works and derivatives are hurt in how data is used and in the very premise of how this software works based on backpropagation. Rights of display and distribution are hurt with public access to reproduction qualities of this software.

7.2 -

7.3 -

7.4 Depending on the model, output can be backtracked to particular pieces. However, as training data is converted into weights, there is no viable way to backtrack consistently without access to training data.

Copyrightability

<https://www.federalregister.gov/d/2023-18624/p-124>

18. AI systems are automatically reprocessing data towards statistical patterns, thus whatever is produced by AI can not be claimed authorship upon by the system or the human using the system. Authorship belongs to the authors of data being processed in the first place, the processing itself does not change that.

If a human trains an AI model exclusively on their own work, they may claim authorship by proxy.

19. -

20. AI-generated material should not be able to receive legal protection. Legal for protection for any kind of AI-generated material will result in copyright laundering.

Denying legal protection for AI-generated material is necessary to incentivise progress beneficial to industries and societies and deter from exploitation and predatory practices.

20.1 -

21. The copyright clause does not permit copyright protection for AI-generated material as AI-generated material can not be claimed authorship upon.

There is also no reason why protection would promote the progress of science and useful arts, in fact the use of current AI generators for artistic and informational material suggest protections would lead to inflation of issues to control and degeneration of any affected subject.