

To make this "intelligence" these companies say, two kind of people are being exploited. First are the said "content creators", which can be an artist, photographer, model and so on. Most of these people make a living out of this content they create. Companies simply take all of it to train AI without permission, and then AI starts creating the same kind of content. This way the companies can make money out of this content without needing to pay the creators whose work were based on. Then they're forced to compete with these big companies in the same market.

One obvious risk is these creators being displaced and the AI companies making a monopoly out of it, crushing workers and small businesses. Another thing that can happen is content creators being forced to become "prompters" or to make hybrid AI/human work in an attempt to survive and employers start using that as an excuse to make people work more for less money. Art and culture will be reduced to artificially generated content. These companies are also planning to give only fraction of cents to the people whose work was used to train AI, and that's definitely too far away from a fair compensation!

The employees who work on training AI by curating and labeling content, they're being exploited as well. They live in terrible work conditions, under pressure and being abused mentally by being exposed to the pornographic, violent and disturbing content they have to filter out of the "intelligence" of generative models, in order to receive a ridiculous salary in exchange of all of this trauma.

The amount of AI generated content online is already too much and is overshadowing legit creators, making it more difficult to them to get noticed. How is that fair when their content were used to train AI in the first place? This is just an example of how AI deprives the copyright owner of income and undermines their market. AI is also unable to create new stuff. It can't experience feelings or moments in life, not even have its own expressions or rational/critical thinking. It's limited to its training material. It's only probability and algorithmic calculation. It only generates things already seen. It can't add nothing new as expression, meaning, information, aesthetics, insights or understandings into nothing. All of that is just taken from the dataset it was trained on and recycled to make the output. Again, how's that fair use?

The billions of materials used to train models are gathered by bots, image repositories or from third party datasets, and are not limited to the fruit of labor of the workers, but also extends to what common people post online or even pictures of them that got uploaded without their consent, which is the case of Lapine whose sensitive medical pictures were included in the image datasets by LAION.

The public also deserves to know if what they're seeing, listening or reading was generated by or using AI. The law must require them to be labeled as so! Uploaders must do that and platforms must ensure their content is being classified correctly. A watermark or caption like "made with AI" can be provided by the platform itself. Tons of false pictures of false events and stuff are already spread online. The risk of misinformation was very much known during the pandemic, isn't that right? ChatGPT is also known to throw misinformation in almost 50% of its answers. Not to mention the biases these softwares have against minorities which is another big problem. In some cases facial recognition might incriminate innocent people just because they're black. There're also several cases where guys used deepfake tech to create porn using pictures woman post online.

Some ideas of legislation I think are a priority are: to enforce online platforms to offer an opt-in/opt-out option that people can chose to warn companies if they authorize their content to be used to train AI or not; enforce AI companies to respect these choices; let people chose if they want platforms to recommend content generated by/made using AI tools; and also AI companies have to be responsible by checking the origin of the material they're using; AI content must not be eligible for copyright or commercial use; A person must not be considered an author if they used AI, as this person only prompted/requested the AI to generate it for them; The responsibility of infringements must fall upon the owners of AI. Accusing the end user might apply at some cases, as in img2img prompts, but not if the infringement is inside the AI. There must not be exceptions and privileges for foundation or open-source models!

Developers, companies and organizations must be transparent and make training datasets open to the public. All of them must retain and disclose records regarding training materials, even mere third party creators of datasets! If you type a prompt using "in the style of Van Gogh" or the name of a character such as Mickey Mouse and then if the AI manage to reproduce the requested style/character, this make obvious the company used Van Gogh paintings or Mickey Mouse images to train the AI. Other methods might exist, but in general it's pretty much difficult to find proof that copyrighted material

was used for training, and forcing them to be transparent might help a lot to people enforce their rights.

The most important recommendation I have to the government is to stop giving that much voice to employers, CEOs, businessman and company representatives and start listening to actual machine learning specialists in order to get all the knowledge and studies they have about AI. Start giving a voice to the affected people as well. Make the debate more plural! And also please never do what Japan did.