



EUROPEAN GUILD
FOR ARTIFICIAL
INTELLIGENCE
REGULATION

We are a group of artists, creatives, publishers and associations from all over Europe united in bringing to the public attention how our data and intellectual properties are being exploited without our consent, on a scale never seen before. Such an unprecedented situation has led us to join our forces to reach out to the European Institutions and have our voices heard. If you believe that your data and creative work should not be exploited with impunity for profit by a handful of corporations, join us in supporting this battle.

Summer 2022 has seen the rise of a new, incredible technology: generative AIs. These forms of artificial intelligence can generate images, videos, texts, programs, audios, 3d models and other contents from textual prompts or other media given by the user. To do so, an AI needs to be trained on a dataset of media. The quality of a generative AI is defined by the quality of its dataset – for example, in regard to images, the more pictures and illustrations an AI learns on, the more styles the AI is able to replicate and the more things it can do. Therefore, the products sold by AI companies are the result of operations on datasets, which contain all sorts of data, including millions of copyrighted images, private pictures and other sensitive material. These files were collected by indiscriminately scraping the internet without the consent of the owners and people portrayed in them and are currently being used by AI companies for profit. This use of sensitive materials and biometric data (such as voice actors' voices) is foremost a violation of privacy and image rights, introducing new dangerous avenues for identity thefts through unprecedented means. At the same time, some of the companies offering content generation services via generative AI are using and manipulating works and names of artists to train their AIs, which allows them to offer on the market the chance to imitate the styles of these artists and their work with the promise of being able to generate original images for any use, making their product irresistible. This exploitation of our work and datas not only does not respect the rights that regulate our society, presenting a huge security risk: it is also severely damaging the art market, potentially scarring it forever. We see this as only the beginning of a crisis that will afflict all sorts of jobs and occupation, whether they are creative jobs or not. The art market is the first one to be affected only because of its structural vulnerabilities, which make it an easy prey. Every time a groundbreaking technology comes to life, our society has to oversee its deployment in order to avoid any harm or infringing of human rights. This hasn't happened with AI technology yet. It is time to change this.

These are the key points we want to enforce:

- 1) Any data related to people or works, in any form, be it digital data – such as text files, audios, videos or images – or captured from reality by camera, microphones or any other mean of registration, shall not be used to train AI model without the explicit and informed consent of its owner. We ask for an extension to the AIs of the principles protecting personal data previously introduced by the GDPR and the introduction of a new form of protection specifically for this kind of exploitation: the “training right”. This protection establishes three alternative scenarios for AI companies when it comes to using a content for training: the impossibility of using said content without explicit consent of the holder; the authorization of the holder to use the content without restrictions; the authorization to use the content regulated by a licensing commercial contract between the parts with clear terms and conditions.
- 2) Using the names of people, stage names or titles of works not covered by a license to exploit for AI training shall be prohibited for those software that allows the use of textual or vocal prompt to generate images, videos, texts or audio.
- 3) Using videos, images, audios and texts not covered by a license to exploit for AI training shall be prohibited for those software that allow the upload of media contents to generate an image, a video, a text or an audio, such as image-to-image software.
- 4) A “human and machine readable” indexing and certification system shall be established, reporting all AIs' activities and the full content of their datasets of images, texts, videos and sounds, be them fully or partially reproduced. Captions such as “entirely made by AI”, “made using AI-generated material” should become the standard.
- 5) The distinction between “copyrighted material” and “public domain” is no longer adequate to identify what can and cannot be used for the datasets. Learning datasets contain personal sensitive data, protected by the privacy laws, but not by copyright. We can find examples of material released when it would not have been possible to foresee its use in a dataset to train an AI model. Any data used in training a model shall be curated and authorized by its legitimate owner and willingly inserted in the dataset by its author with full knowledge of it. AI companies shall produce internally original materials for the training or license external material following terms and contracts previously established with the authors or rightful holders of said material.

Find out more on www.egair.eu