PRESS RELEASE

Study Reveals: AI Training is Copyright Infringement

Brussels/Berlin - September 5, 2024: A computer scientist and a legal scholar shed light on the black box of processing steps in AI training - for the first time on this scale. The presentation of the interdisciplinary study "Copyright & Training of Generative AI - Technological and Legal Foundations" took place today in the European Parliament.

In spring, the Copyright Initiative commissioned Prof. Dr. Tim W. Dornis (University of Hanover) in collaboration with Prof. Dr. Sebastian Stober (University of Magdeburg) with a tandem expert opinion on the technological and legal aspects of training generative AI models. Their interdisciplinary research provides urgently needed new insights into the technically necessary intermediate steps in the training of generative artificial intelligence. For the first time on this scale, a computer scientist and a legal scholar are jointly creating evidence regarding the processing steps in AI training. During the event, many open questions about protected materials were answered in a well-founded, reliable manner and in line with the current state of the art.

The work of Prof. Dornis and Prof. Stober focuses on the copyright assessment of the processing of protected material in AI training: "As a closer look at the technology of generative AI models reveals, the training of such models is not a case of text and data mining. It is a case of copyright infringement – no exception applies under German and European copyright law," says Prof. Dornis. Prof. Stober explains that "parts of the training data can be memorized in whole or in part by current generative models - LLMs and (latent) diffusion models - and can therefore be generated again with suitable prompts by end users and thus reproduced."

Axel Voss, MEP and host of today's event in the European Parliament, expressly thanks the scientists Dornis and Stober and is pleased that "the study not only proves that the training of Generative AI models is not covered by text and data mining, but that it also provides further important indications and suggestions for a better balance between the protection of human creativity and the promotion of AI innovation."

"This study is explosive because it proves that we are dealing with large-scale theft of intellectual property. The ball is now in the politicians' court to draw the necessary conclusions and finally put an end to this theft at the expense of journalists and other authors," commented Hanna Möllers, legal advisor to the DJV and representative of the European Federation of Journalists (EFJ).

Katharina Uppenbrink, Managing Director of the Initiative Urheberrecht, emphasizes: "It is a groundbreaking result if we now have proof that the reproduction of works by an AI model constitutes a copyright-relevant reproduction and, in addition, that making them available on the European Union market may infringe the right of making available to the public."

The composer and spokesperson for the Copyright Initiative, Matthias Hornschuh, comments: "There would be a new, profitable licensing market on the horizon, but no remuneration is flowing, while generative AI is preparing to replace those whose content it lives from in its own market. This jeopardizes professional knowledge work and cannot be in the interests of society, culture or the economy. All the better that the authors of our tandem study provide the technological and copyright basis for finally turning the legal consideration of generative artificial intelligence from its head to its feet."

Dornis, Tim W. and Stober, Sebastian, Copyright and training of generative AI models - technological and legal foundations (in German)
(September 4, 2024).
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4946214

The abstract can be found here (in English and German language):
https://urheber.info/panel/pages/diskurs+abstract-interdisciplinary-study

The executive summary is available in English and German language at
https://urheber.info/panel/pages/diskurs+executive-summary-deutsch
https://urheber.info/panel/pages/diskurs+executive-summary-english

English-language publications are in preparation.
The presentation in Berlin will take place at the end of September.

Authors' Rights Initiative (Initiative Urheberrecht)

Through its 45 member organizations, the Copyright Initiative represents the interests of around 140,000 authors and performing artists in the fields of fiction and non-fiction, visual arts, journalism, design, music, audio and audiovisual media, photography, illustration, drama, game development, dance and many more: https://urheber.info/about-us

Contact:

Katharina Uppenbrink
Managing Director
Authors' Rights Initiative (Initiative Urheberrecht)
Phone: +49 160 9095 4016
Email: katharina.uppenbrink@urheber.info
Website: www.urheber.info