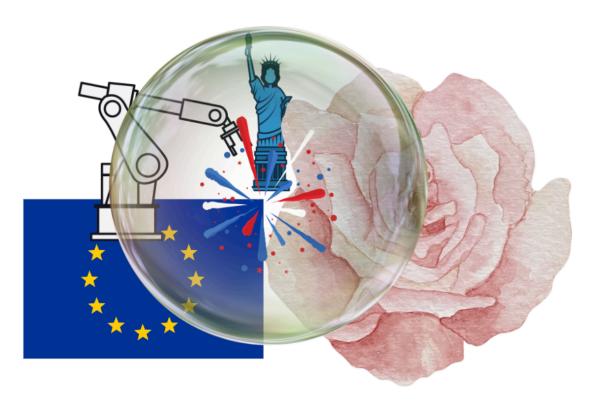


← Back

# Generative Al and transparency of databases and their content, from a copyright perspective

May 21, 2024



### By Ana Andrijevic\*

In May 2024, the Organisation for Economic Co-operation and Development (OECD) updated its Principles on Artificial Intelligence (AI), including the principle of transparency, which has contributed to shaping policy and regulatory debates on AI and generative AI (i.e. deep learning models that can create new content, such as text, computer code, and images, in response to a user's short, written description – a "prompt"). From a copyright perspective, the principle of transparency has become increasingly relevant in several respects: the transparency of the databases (or datasets) and their content used to train AI models, the transparency in AI models, and the transparency regarding the use of AI tools in the creative process.

This contribution focuses on the transparency of the databases and their content used to train AI models, which are increasingly being kept under lock and key and made inaccessible for perusal by AI companies to maintain their competitive edge. Since these databases contain a wide range of protected literary and artistic works (e.g. literary works, photographic works, paintings and drawings, musical works, and more), the interests of AI companies collide with those of authors and rights holders, whose works are used without authorization or compensation. In this contribution, we explore this discrepancy of interests from the principle of transparency angle.

### Lack of transparency of databases and their content used to train Al models

Over the last few years, AI companies have become more cautious about disclosing the databases used to train their AI models, as illustrated, for instance, by Meta (for the training of Llama) and OpenAI (for the training of GPTs, i.e. Generative Pre-trained Transformers). In

particular, OpenAl's strategy has shifted from openness to limiting the amount of information relating to their training datasets. Thus, in the span of a couple of years (2018 to 2020), the US company has gone from disclosing the use of BooksCorpus (a dataset of 7,000 self-published books retrieved from smashwords.com, which are largely protected under copyright) for the training of GPT-1 (released in June 2018), to indicating the use of several vaguely labeled datasets to train GPT-3 (released in July 2020), including two internet-based books corpora (Books1 and Books2). Although the content of Books1 and Books2 remains unknown, the defendants in *Tremblay et al. v. OpenAl et al.* (consolidated on March 12, 2024), one of several complaints filed in the USA against Al companies in 2023, have investigated the issue at hand.

With the launch of GPT-4 on March 14, 2023, OpenAI has become increasingly secretive, citing "the competitive landscape and the safety implications of large-scale models like GPT-4" to explain its choice. Nevertheless, as OpenAI CEO Sam Altman recently acknowledged, there is no doubt that their datasets contain large amounts of copyrighted work, since, in his words, "it would be impossible to train today's leading AI models without using copyrighted materials." In addition to Meta and OpenAI, other AI companies such as Google and Nvidia have also refrained from disclosing their training datasets and their content over time. The same applies to Stability AI which initially opted for disclosure of its datasets for training its AI model Stable Diffusion, a strategy that proved unsuccessful as it became one of the first AI companies to be taken to court in the USA and in the UK in 2023. However, the research paper on its AI model Stable Video Diffusion (released on November 21, 2023 datasets.

# Generative AI and transparency from a copyright perspective

From a regulatory perspective, the latest amended EU AI Act, approved by the European Parliament in March 2024, includes art. 53 par. 1 let. d, the aim of which is to promote transparency on the data (including copyrighted data) used by providers for training their General Purpose AI ("GPAI") models. It reads: "Providers of general-purpose AI models shall: (d) draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office."

This provision, which was introduced into the EU AI Act at a late stage in the drafting of the regulation, has been the subject of much criticism on the grounds that the requirement posed by art. 53 par. 1 let. d of the EU AI Act was ambiguous and overly demanding.

However, the recent inclusion of recital 107 provides a further clarification that allays some of the concerns raised by AI model providers, and states that the summary "should be generally comprehensive in its scope instead of technically detailed to facilitate parties with legitimate interests, including copyright holders, to exercise and enforce their rights under Union law, for example by listing the main data collections or sets that went into training the model, such as large private or public databases or data archives, and by providing a narrative explanation about other data sources used" (emphasis added). Nevertheless, the AI Office's role does not imply to verify or proceed to "a work-by-work assessment of the training data in terms of copyright compliance." On the subject of transparency, while a summary of the databases used by AI companies can be a first source of relevant information, the lack of disclosure of their content (especially if they are private) remains a challenge for rights holders to establish conclusive evidence of copying.

In the USA, the Copyright Office issued a Notice of inquiry and request for comments on August 30, 2023, seeking comments on copyright law and policy issues raised by AI, "including those involved in the use of copyrighted works to train AI models, the appropriate levels of transparency and disclosure with respect to the use of copyrighted works (...)." More specifically, the US Copyright Office inquired whether developers of AI models should be required "to collect, retain, and disclose records regarding the materials used to train their models." to enable copyright owners to determine whether their works have been used. The US Copyright Office received more than 10,000 written comments by the December 2023 deadline; these are currently under review. It is worth noting, however, that unsurprisingly, AI companies such as Google and Meta (and as previously mentioned, OpenAI) agree that datasets and their content should not be divulged.

Yet, on April 9, 2024, the Generative Al Copyright Disclosure Act was introduced by California Democratic Congressman Adam Schiff and, if approved, would require "a notice be submitted to the Register of Copyright with respect to copyrighted works used in building generative Al systems, and for other purposes." More specifically, section 2(a)(1) of the Generative Al Copyright Disclosure Act would require "[a] person who creates a training dataset, or alters a training dataset (...) in a significant manner, that is used in building a generative Al system" to submit to the Register of Copyrights a notice that contains: "(A) a sufficiently detailed summary of any copyrighted works used – (i) in the training dataset (...); or (ii) to alter the training dataset (...)" and "(B) the URL for such dataset (in the case of a training dataset that is publicly available on the internet at the time the notice is submitted)." Therefore, unlike the requirement of art. 53 par. 1 let. d of the EU Al Act, which is limited to a summary of the content used to train the GPAI, the proposed Generative Al Copyright Disclosure Act would mandate a notice regarding all copyrighted works used in building or altering the training dataset.

Despite the imprecise nature of this proposal (e.g. databases can be created by more than one person, as in the case of entities such as the nonprofit organizations LAION or Common Crawl or

undeniably give rights holders of protected literary and artistic works sturdier evidence of copyright infringement by the Al company. This point is not innocuous, as the plaintiffs in several lawsuits brought against Al companies did not have access to the content of the training datasets and could therefore only assume that their works had been used by Al companies based on the outputs generated by their Al tools. For instance, in *Tremblay et al. v. OpenAl et al.*, the defendants argue that ChatGPT generates very accurate summaries of their copyrighted works. However, without access to OpenAl's data, it cannot be ruled out that these summaries were generated from other sources (e.g. other summaries written by third parties).

In fact, plaintiffs have rarely provided conclusive evidence of copying in proceedings against AI companies. One such case is *Concord Music Group Inc. et al. v. Anthropic PBC*, in which the plaintiffs (the rights holders) were able to present clear examples of the reproduction of their lyrics by Claude, Anthropic's AI tool. Similarly, *The New York Times* was able to demonstrate in *The New York Times Company v. Microsoft Corporation et al.* that the defendants' AI tools can generate output that "recites Times content verbatim, closely summarizes it, and mimics its expressive style," as illustrated by several examples produced by the plaintiff. Thus, despite the lack of transparency regarding the databases and their content used by the defendants, the plaintiffs were able to provide concrete evidence of copying by analyzing the outputs proposed by these tools. However, these complaints are unlikely to contribute to greater transparency, which begs the question of alternative solutions.

# Possible remedies for copyright holders against Al companies

In cases where copyrighted works have already been harvested, the first challenge is to identify which works have been used by AI companies. The proposed solutions diverge between the EU and the USA: On the one hand, the EU AI Act requires GPAI providers to "draw up and make publicly available a sufficiently detailed summary about the content used for training" of the GPAI (art. 53 par. 1 let. d of the EU AI Act), and on the other hand, the US Generative AI Copyright Disclosure Act would require AI companies to disclose both databases and their content (section 2(a)(1)). Thus, while the latter proposal is more advantageous for authors and rights holders (since it enables them to identify the content of databases), the burden and responsibility of requesting the removal of their works or, if necessary, filing a complaint, still rests with them.

For the time being, *curative* opt-outs (as opposed to *preventive* opt-outs permitted by art. 4 par. 3 of the Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (DSM Directive)<sup>[54]</sup> referred to in art. 53 par. 1 let. c of the EU AI Act<sup>[55]</sup>) are left to the rights holders themselves or to the goodwill of AI companies, whose opt-out mechanisms have proved unsatisfactory. What is more, these processes sustain an important disadvantage: They only apply to future uses, not current ones. Indeed, the possibility of deleting the content of trained models is still under development, although there are interesting avenues for large language models (LLMs). However, curative opt-outs are disadvantageous and not a sustainable solution, as rights holders have to ensure, for each AI company, that their works have not been used.

An upstream approach is therefore needed to promote the transparency of databases and their content. The creation of partnerships between AI companies and providers of literary and artistic works, such as between Getty Images and Nvidia, <sup>[59]</sup> Universal Music and BandLab Technologies, <sup>[60]</sup> Google with Reddit, <sup>[61]</sup> OpenAI with *Le Monde* and *Prisa Media*, <sup>[62]</sup> is one development that has been observed. Yet, this type of collaboration mainly involves AI companies with large-scale content suppliers but does not extend to smaller players. They are, however, taken into account for certification, which is another solution to promote transparency of databases and their content, as is done by Fairly Trained, <sup>[63]</sup> whose mission is to certify AI companies that get a license for their training data. <sup>[64]</sup>

# Conclusion

Fom a legal point of view, the transparency of databases and their content involves a balance between, on the one hand, the interest of AI companies in preserving a competitive advantage, favored by the EU AI Act, and, on the other hand, the interests of rights holders, who could benefit from the obligation to disclose databases and their content under the US Generative AI Copyright Disclosure Act. However, solutions aimed at improving database transparency, in both the EU and the USA, remain unsatisfactory, as the onus is still on rights holders to opt out (where possible and with the aforementioned constraints) or lodge a complaint. Yet, solutions are available, which include partnerships between AI companies and providers of copyrighted works, and certification.

It should nevertheless be noted that if, in the USA, the *fair use* doctrine is admitted concerning the unauthorized copying of protected works to train AI companies' models, the US Generative AI Copyright Disclosure Act will lose its relevance, as will the issue of transparency of databases and their content. However, this matter will remain relevant within the EU insofar as rights holders expressly reserve the right to make reproductions and extractions of their works for text and data mining (see art. 4 par. 3 of the DSM Directive, referred to in art. 53 par. 1 let. c of the EU AI Act), which thus favors the interests of rights holders over those of AI companies.

### About the Author:

Ana Andrijevic is a PhD candidate at the University of Geneva. She is also a visiting researcher at Harvard Law School where she is an affiliated researcher at the Berkman Klein Center for Internet & Society (Harvard University).

### Sources:

- 1. OECD, OECD Legal Instruments, OECD (02.05.2024), available at: <a href="https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449">https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449</a>
  See also OECD, OECD updates AI Principles to stay abreast of rapid technological developments, OECD (03.05.2024), available at: <a href="https://www.oecd.org/newsroom/oecd-updates-ai-principles-to-stay-abreast-of-rapid-technological-developments.htm">https://www.oecd.org/newsroom/oecd-updates-ai-principles-to-stay-abreast-of-rapid-technological-developments.htm</a>
- 2. OECD, OECD Legal Instruments, par. 1.3 on Transparency and explainability, which states in particular that: "Al Actors should commit to transparency and responsible disclosure regarding Al systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art: iii. when feasible and useful, to provide plain and easy-to-understand information on the sources of data/input (...)." ↑
- 3. See for instance OECD.AI, OECD Principles, Transparency and explainability (Principle 1.3), available at: <a href="https://oecd.ai/en/dashboards/ai-principles/P7">https://oecd.ai/en/dashboards/ai-principles/P7</a>↑
- 4. World Intellectual Property Organization (WIPO), *Generative AI, Navigating Intellectual Property*, Geneva (2024), p. 2, available at: <a href="https://www.wipo.int/export/sites/www/about-ip/en/frontier-technologies/pdf/generative-ai-factsheet.pdf">https://www.wipo.int/export/sites/www/about-ip/en/frontier-technologies/pdf/generative-ai-factsheet.pdf</a>↑
- 5. See for instance Tremblay et al. v. OpenAl et al., Case 3:23-cv-03223-AMO, 13.03.2024 and The New York Times Company, v. Microsoft Corporation et al., Case 1:23-cv-11195, 27.12.2023. ↑
- 6. Getty Images (US) v. Stability AI Inc., Case 1:23-cv-00135-UNA, 03.02.2023 and Getty Images (UK) et al. v. Stability AI Ltd., [2023] EWHC 3090 (Ch), Case No: IL-2023-000007, 1.12.2023. ↑
- 7. Andersen et al. v. Stability Al Ltd. et al., Case 3:23-cv-00201, 13.01.2023 and Jingna Zhang et al. v. Google LLC et al., Case 3:24-cv-02531, 26.04.2023. ↑
- 8. Ashley Carman and Lucas Shaw, Sony Music Warns Companies to Stop Training AI on Its Artists' Content, Bloomberg (16.05.2024), available at: <a href="https://www.bloomberg.com/news/articles/2024-05-16/sony-music-warns-companies-to-stop-training-ai-on-its-artists-content">https://www.bloomberg.com/news/articles/2024-05-16/sony-music-warns-companies-to-stop-training-ai-on-its-artists-content</a>↑
- 9. Meta, Introducing Meta Llama3: The most capable openly available LLM to date, Meta (18.04.2024), available at:

  <a href="https://ai.meta.com/blog/meta-llama-3/">https://ai.meta.com/blog/meta-llama-3/</a> It indicates that: "Llama 3 is pretrained on over 15T tokens that were all collected from publicly available sources." ↑
- 10. Chloe Xiang, *OpenAl Is Now Everything It Promised Not to Be: Corporate, Closed-Source, and For-Profit*, Vice (28.02.2024), available at: <a href="https://www.vice.com/en/article/5d3naz/openai-is-now-everything-it-promised-not-to-be-corporate-closed-source-and-for-profit">https://www.vice.com/en/article/5d3naz/openai-is-now-everything-it-promised-not-to-be-corporate-closed-source-and-for-profit</a> \( \)
- 11. Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, *Improving Language Understanding by Generative Pre-Training*, OpenAl (2018), p. 4, available at: <a href="https://cdn.openai.com/research-covers/language-unsupervised/language\_understanding\_paper.pdf">https://cdn.openai.com/research-covers/language-unsupervised/language\_understanding\_paper.pdf</a> ↑
- 12. Hugging Face, *Datasets: bookcorpus*, available at: <a href="https://huggingface.co/datasets/bookcorpus">https://huggingface.co/datasets/bookcorpus</a> ↑
- 13. Tremblay et al. v. OpenAl et all., par. 39. ↑
- 14. Tom B. Brown et al., Language Models are Few-Shot Learners, p. 8, available at: https://arxiv.org/pdf/2005.14165 ↑
- 15. See footnote n°5. ↑
- 16. Edward Lee, *Status of all 24 copyright lawsuits v. Al companies*, 17.05.2024, available at: <a href="https://chatgptiseatingtheworld.com/2024/05/17/status-of-all-24-copyright-lawsuits-v-ai-companies-may-17-2024/↑">https://chatgptiseatingtheworld.com/2024/05/17/status-of-all-24-copyright-lawsuits-v-ai-companies-may-17-2024/↑</a>
- 17. Tremblay et al. v. OpenAl et al., par. 40 to 43. ↑
- 18. OpenAl, GPT-4 Technical Report, OpenAl (04.03.2024), p. 2, available at: https://arxiv.org/pdf/2303.08774 ↑
- 19. House of Lords Communications and Digital Select Committee, *OpenAl written evidence (LLM0113)*, London (05.12.2023), p. 4, available at: <a href="https://committees.parliament.uk/writtenevidence/126981/pdf/">https://committees.parliament.uk/writtenevidence/126981/pdf/</a> ↑
- 20. Jingna Zhang et al. v. Google LLC et al., par. 31 and 32. ↑
- 21. In Abdi Nazemian et al. v. NVIDIA Corporation, Case 3:24-cv-01454, 08.03.2024, par. 22 and 23 and Andre Dubus III et al. v. NVIDIA Corporation, Case 4:24-cv-02655, 02.05.2024, par. 21 and 22, defendants refer to the training of the NeMo Megatron, released in September 2022 and trained on "The Pile" dataset. However, if we take a more recent example, NVIDIA indicates for instance that its AI model PeopleNet was trained on a "proprietary dataset with more than 7.6 million images," without any further information. For more, see NVIDIA, PeopleNet Model Cart, NVIDIA (11.04.2024), available at: <a href="https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/peoplenet">https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/peoplenet</a>
- 22. Andersen et al. v. Stability Al Ltd. et al. and Getty Images (US) v. Stability Al Inc.  $\underline{\uparrow}$
- 23. Getty Images (UK) et al. v. Stability AI Ltd.  $\underline{\uparrow}$
- 24. Stability AI, Introducing Stable Video Diffusion, Stability AI (21.11.2023), available at: <a href="https://stability.ai/news/stable-video-diffusion-open-ai-video-model">https://stability.ai/news/stable-video-diffusion-open-ai-video-model</a> ↑
- 25. Andreas Blattmann et al., Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets, Stability Al (21.11.2023), p. 2, available at:

https://static1.squarespace.com/static/6213c340453c3f502425776e/t/655ce779b9d47d342a93c890/1700587395994/stable\_video\_d iffusion.pdf ↑

- 26. European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Act (COM(2021)0206 − C9-0146/2021 − 2021/0106(COD), available at: <a href="https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138">https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138</a> EN.html#title2 ↑
- 27. European Parliament, Artificial Intelligence Act: MEPs adopt landmark law, Brussels (13.03.2024), available at:

  <a href="https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law">https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law</a> As indicated: "The regulation is still subject to a final lawyer-linguist check and is expected to be finally adopted before the end of the legislature (through the so-called corrigendum procedure). The law also needs to be formally endorsed by the Council." <a href="https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law">https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law</a> As indicated: "The regulation is still subject to a final lawyer-linguist check and is expected to be finally adopted before the end of the legislature (through the so-called corrigendum procedure). The law also needs to be formally endorsed by the Council."</a>
- 28. See definition of providers in art. 3 par. 3 of the EU AI Act. ↑
- 29. See definition of GPAI in art. 3 par. 63 of the EU AI Act. 1
- 30. With regard to the AI Office, see art. 3 par. 47 of the EU AI Act: "AI Office' means the Commission's function of contributing to the implementation, monitoring and supervision of AI systems and AI governance carried out by the European Artificial Intelligence Office established by Commission Decision of 24.1.2024; references in this Regulation to the AI Office shall be construed as references to the Commission." ↑
- 31. Paul Keller, *A first look at the copyright relevant parts in the final AI Act compromise*, Kluwer Copyright Blog (11.12.2023), available at: <a href="https://copyrightblog.kluweriplaw.com/2023/12/11/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise/">https://copyrightblog.kluweriplaw.com/2023/12/11/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise/</a> ↑
- 32. Andres Guadamuz, The EU Al Act and Copyright, TechnoLlama (14.03.2024), available at: https://www.technollama.co.uk/the-eu-ai-act-and-copyright 1
- 33. Keller. ↑
- 34. *ld*. 1
- 35. See recital 108 of the EU AI Act. ↑
- 36. Library of Congress, Copyright Office, *Artificial Intelligence and Copyright*, No. 2023-6, in: Federal Register, Vol. 88, No. 167, Washington, DC (30.08.2023), available at: <a href="https://www.govinfo.gov/content/pkg/FR-2023-08-30/pdf/2023-18624.pdf">https://www.govinfo.gov/content/pkg/FR-2023-08-30/pdf/2023-18624.pdf</a> ↑
- 37. Id., p. 59942. 1
- 38. Id., p. 59947. 1
- 39. Id. 1
- 40. US Copyright Office, *Artificial Intelligence and Copyright*, Washington, DC, available at: <a href="https://www.regulations.gov/docket/COLC-2023-0006/comments">https://www.regulations.gov/docket/COLC-2023-0006/comments</a> <a href="https://www.regulations.gov/docket/COLC-2023-0006/comments">https://www.regulations.gov/docke
- 41. US Copyright Office, Washington, DC (23.02.2024), p. 5, available at: <a href="https://copyright.gov/laws/hearings/USCO-Letter-on-Al-and-Copyright-Initiative-Update-Feb-23-2024.pdf?loclr=blogcop↑</a>
- 42. US Copyright Office, *Comment from Google*, Washington, DC (01.11.2023), pp. 11 and 12, available at: <a href="https://www.regulations.gov/comment/COLC-2023-0006-9003">https://www.regulations.gov/comment/COLC-2023-0006-9003</a> ↑
- 43. US Copyright Office, *Comment from Meta Platforms, Inc.*, Washington, DC (01.11.2023), pp. 19 and 20, available at: <a href="https://www.regulations.gov/comment/COLC-2023-0006-9027">https://www.regulations.gov/comment/COLC-2023-0006-9027</a> ↑
- 44. Available at: https://schiff.house.gov/imo/media/doc/the\_generative\_ai\_copyright\_disclosure\_act.pdf ↑
- 45. Rep. Schiff introduces groundbreaking bill to create AI transparency between creators and companies, Washington, DC (09.04.2024), available at: <a href="https://schiff.house.gov/news/press-releases/rep-schiff-introduces-groundbreaking-bill-to-create-ai-transparency-between-creators-and-companies">https://schiff.house.gov/news/press-releases/rep-schiff-introduces-groundbreaking-bill-to-create-ai-transparency-between-creators-and-companies</a> ↑
- 46. LAION, *About*, available at: <a href="https://laion.ai/about/↑">https://laion.ai/about/↑</a>
- 47. Common Crawl, Frequently asked questions, available at: https://commoncrawl.org/faq↑
- 48. Tremblay et al. v. Stability AI et al., par. 5 and 51. ↑
- 49. Concord Music Group, Inc. et al., v. Anthropic PBC, Case 3:23-cv-01092, 18.10.2023. ↑
- 50. Id., par. 66 to 69. ↑
- 51. The New York Times Company, v. Microsoft Corporation et al.  $\underline{\uparrow}$
- 52. *Id.*, par. 4. <u>↑</u>
- 53. See for instance *Id.*, par. 99, 100, and 104 to 107. <u>↑</u>
- 54. Art. 4 par. 3 of the DSM Directive: "The exception or limitation provided for in paragraph 1 [Exception or limitation for text and data mining] shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online." 

  1
- 55. Art. 53 par. 1 let. C of the EU Al Act: "Providers of general-purpose Al models shall: (c) put in place a policy to comply with Union copyright law, and in particular to identify and comply with, including through state of the art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790." 

  ↑
- 56. See "Have I Been Trained?, which is mentioned in Andersen et al. v. Stability AI et al., p. 6, footnotes n°1 and 2. ↑
- 57. Kali Hays, OpenAl offers a way for creators to opt out of Al training data. It's so onerous that one artist called it 'enraging', Business Insider (29.09.2023), available at: <a href="https://www.businessinsider.com/openai-dalle-opt-out-process-artists-enraging-2023-9">https://www.businessinsider.com/openai-dalle-opt-out-process-artists-enraging-2023-9</a> <a href="https://www.businessins.com/openai-dalle-opt-out-process-artists-enraging-2023-9">https://www.businessins.com/openai-dalle-opt-out-process-artists-enraging-opt-out-process-artists-enraging-opt-out-proce

- 58. Ronen Eldan and Mark Russinovich, Who's Harry Potter? Approximate Unlearning in LLMs, 04:10.2023, available at: https://arxiv.org/pdf/2310.02238 ↑
- 59. Rick Merritt, Moving Pictures: NVIDIA, Getty Images Collaborate on Generative AI, NVIDIA (21.03.2023), available at: https://blogs.nvidia.com/blog/generative-ai-getty-images/ 1
- 60. Universal Music Group, Universal Music Group and BandLab Technologies announce first-of-its-kind strategic Al collaboration, Universal Music (18.10.2023), available at: https://www.universalmusic.com/universal-music-group-and-bandlab-technologies-announce-first-of-itskind-strategic-ai-collaboration/ ↑
- 61. Anna Tong, Echo Wang, Martin Coulter, Exclusive: Reddit in AI content licensing deal with Google, Reuters (21.02.2024), available at: https://www.reuters.com/technology/reddit-ai-content-licensing-deal-with-google-sources-say-2024-02-22/ 1
- 62. OpenAI, Global news partnerships: Le Monde and Prisa Media, OpenAI (13.03.2024), available at: https://openai.com/index/global-newspartnerships-le-monde-and-prisa-media/ 1
- 63. Fairly Trained, About, available at: https://www.fairlytrained.org/about ↑
- 64. Fairly Trained, Licensed Model Certification, available at: https://www.fairlytrained.org/certifications ↑













Disclaimer: This article is for educational purposes only and is not meant to provide legal advice. Readers should not construe or rely on any comment or statement in this article as legal advice. For legal advice, readers should seek a consultation with an attorney.

Previous

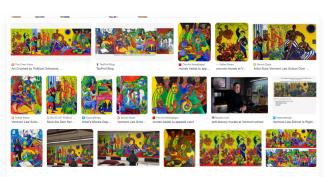
Next →

# **Related Posts**



# Dispute over Thomas Cole's Painting Comes to Court

September 24, 2013



Case Review: Kerson v. Vermont Law School, Inc.

June 21, 2024



**Fundi** The S

Februa

# **INSTAGRAM**













# Generative AI and transparency of databases and their content, from a copyright perspective - Center for Art Law





















# DISCLAIMER

Center for Art Law is a New York State non-profit fully qualified under provision 501(c)(3) of the Internal Revenue Code.

The Center does not provide legal representation. Information available on this website is purely for educational purposes only and should not be construed as legal advice.

Become a member

Subscribe

# TERMS OF USE AND PRIVACY POLICY

Your use of the Site (as defined below) constitutes your consent to this Agreement. Please read our Terms of Use and Privacy Policy carefully.

© 2024 CENTER FOR ART LAW