IIC (2023) 54:491–494
https://doi.org/10.1007/s40319-023-01321-y

EDITORIAL

ChatGPT and Generative AI Tools: Theft of Intellectual Labor?

Alain Strowel

Accepted: 22 March 2023 / Published online: 4 April 2023
© Max Planck Institute for Innovation and Competition, Munich 2023

Since its release in November 2022, ChatGPT, the Large Language Model trained by OpenAI to interact in a conversational way, has attracted more than 100 million users, as well as the attention of the press. Numerous opinions of "thought leaders" have been published, some heralding "an intellectual revolution" (e.g. Henry Kissinger, Eric Schmidt and Daniel Huttenlocher[1]), others criticizing "the false promise of ChatGPT" (e.g. Noam Chomsky[2]). Many old markets and tech businesses feel exposed.[3] The education sector is also trying to address the new challenges, which include ethics in research and student plagiarism. At the same time, generative AI has prompted the interest of tech companies and investors: Microsoft will inject more than USD 10 billion in OpenAI and Google has invested USD 300 million in Anthropic, etc.[4]

While ChatGPT and other generators of content – for example, those spitting images (DALL-E, Stable Diffusion, DeviantArt, Midjourney) – are so far free for their users (though in exchange for some personal data), paid versions are being proposed (as is already the case for ChatGPT), and several business models are in the making. The prospect of new funding and revenues is likely to lead to additional scrutiny of those tools. Do these AI models use the literary and artistic data needed for their training in a legal and fair way?

[1] Wall Street Journal, 24 February 2023.

[2] New York Times, 8 March 2023. *See also* Patrick Goethals (2023) "De grootste intellectuele hold-up uit de geschiedenis", De Standard, 24 January 2023. The title of this opinion has inspired mine.

[3] Tripp Mickle, Cade Mets and Nico Grant (2023) "The Chatbots Are Here, and the Internet Industry Is in a Tizzy", New York Times, 8 March 2023.

[4] *See* the other investments listed by CBInsights, "The state of generative AI in 7 charts", 25 January 2023 (at https://www.cbinsights.com/research/generative-ai-funding-top-startups-investors/). Both Microsoft and Google are trying to insert generative tools in their search tool.

A. Strowel (✉)
Professor, UCLouvain & Université Saint-Louis, attorney, Brussels, Belgium
e-mail: alain.strowel@uclouvain.be

1 Backlash Among Artists

Several artists have openly expressed their strong opposition to the use of these tools, which they see as undermining their jobs and crafts.[5] Our 2022 study for the European Commission,[6] which included interviews with stakeholders, also indicated that artists are at minimum ambivalent about the introduction of these technologies, seeing some advantages (e.g. for video games or photos), but also the risks. One narrative widely shared since the end of 2022 is that AI tools are piggybacking on previous creations. Some artists and creatives have, for example, founded EGAIR, the "European Guild for AI Regulation", which collects money to "Help protect our art and data from AI companies".[7] The feeling of a hold-up on creative labor sometimes translates in the legal claim that AI machines are stealing real people's intellectual property (IP).

2 Pending Litigation (US and UK)

At least five cases, most of them class actions, have been introduced against generative AI in the US and the UK. Most of those actions target the tools for creating graphic materials: for example, Stability AI, the company behind Stable Diffusion, is being sued in a US class action initiated by several graphic artists, and Getty Images is the plaintiff in two other cases (in the US and UK) against this company. Another class action was brought against the software platform GitHub (and its parent Microsoft) for its offer of Copilot, a tool designed by OpenAI to provide computer code under an advanced autocomplete mode. In the US, the main arguments brought forward concern copyright (and Digital Millennium Copyright Act) violations. Unfair competition is another ground commonly put forward by the plaintiffs. In its US case, Getty Images also relies on trademark infringement, as some photographs generated by Stable Diffusion retain Getty's watermark. On top of intellectual property, data protection rules might be invoked in some of those anti-generative AI actions,[8] which are likely to spread across Europe.

[5] *E.g.* when the image "Théâtre d'opéra spatial", generated by Midjourney, won a fine art competition, Jason M. Allen's victory created a backlash among artists. Similarly, the use by the San Francisco Ballet of the same AI tool to generate an image for the promotion of its "Nutcracker" performance was met with outrage by graphic artists. *See also* the story about the Polish artist Greg Rutkowski (M. Heikkilä (2022) This artist is dominating AI-generated art. And he is not happy about it, MIT Technology Review, 16 September 2022).

[6] *See* Luc Desaunettes, Sari Depreeuw and Alain Strowel (2022) for the AI part of the "Study on copyright and new technologies: copyright data management and artificial intelligence", Publications Office of the European Union, https://data.europa.eu/doi/10.2759/570559.

[7] *See* the European Guild for Artificial Intelligence Regulation manifesto: https://www.egair.eu/#manifesto.

[8] In a class action against Lensa AI, an app for generating customized cartoon avatars developed by Prisma Labs, the plaintiffs rely on a violation of some privacy law that protects biometric information, as the app has been collecting facial geometry data for training its model.

3 The Copyright Framework

In the US, the copyright discussion will revolve around the fair use exception, thus whether the use of copyrighted materials for training AI models is fair, i.e. transformative. One precedent is likely to play for the AI companies: in the *Google Books* case, the court found that "Google's unauthorized digitizing of copyright-protected works, creation of a search functionality, and display of snippets from those works were non-infringing fair uses. The purpose of the copying is highly transformative, the public display of text is limited, and the revelations do not provide a significant market substitute for the protected aspects of the originals. Google's commercial nature and profit motivation do not justify denial of fair use".[9] But the extent of the use of protected materials to train generative AI tools is different, and the output is more likely to be a good substitute for the training materials, at least for graphic works.

In the EU, the legal discussion is framed by two text and data mining (TDM) exceptions (Arts. 3 and 4 of Directive (EU) 2019/790), which are relevant for any AI based on Machine Learning. Asked about the first exception, TDM for scientific purposes by research organizations, ChatGPT responded to me that it is "subject to certain conditions, such as the obligation to provide a fair remuneration to the rightholders" (!), which is plainly wrong but shows how the tool was trained to show empathy (?). To undertake TDM outside of the research context, one has to use "lawfully accessible works" unless such use has been "expressly reserved by their rightholders in an appropriate manner, such as machine-readable means" (Art. 4). Some see this provision as offering "a huge opportunity for creators to build new collective structures" for exercising their right to oppose AI training and to "rely on open protocols to communicate their norms and terms"[10] to those applying Machine Learning. The question remains whether the community of artists could wake up and develop an "appropriate" opt-out tool, and whether AI companies will adapt their content scrapping practices to accommodate such tool.[11]

4 What's Next?

"The battle over the role of data in training AI could, according to the Financial Times, become as important to the tech industry as the patent wars at the dawn of the smartphone era".[12] This would mean that many court cases will be initiated – let's remember the numerous IP sagas involving Apple, Samsung, Nokia, Google,

[9] *Authors Guild v. Google Inc.* 804 F3d 202, 229 (2d Cir 2015).

[10] Paul Keller (2023) "Protecting creatives or impeding progress? Machine learning and the EU copyright framework", 17 February 2023, at https://openfuture.eu/blog/.

[11] There is a precedent which does not augur well: in the cases involving press publishers opposing the snippets on Google News, the opt-out was the solution proposed by Google, but the press could not thrive online if dereferenced on the dominant search tool, and no agreement on a standardized technical tool that would allow to express the opt-out while remaining searchable could be reached.

[12] Richard Waters (2023) "Generative AI: how will the new era of machine learning affect you?" Financial Times, 25 January 2023.

HTC, etc. But there are good reasons to distinguish the IP battle around generative AI: the smartphone wars focused on patents and opposed large companies with extensive IP portfolios and deep pockets. The numerous artists complaining about the infringement of their work are in a very different position (no need to explain here), and their claims rely as much on broad and uncertain principles (unfair practice or unjust enrichment) as on IP rights. Litigation will, in any case, raise insurmountable evidence problems. While an artistic style should remain outside the scope of copyright, the question of what an individual style means will resurface and be subject to discussion. The discrepancy between the individual gains and claims regarding the reuse of content, on the one hand, and the benefits of those generative tools for many people, including artists, on the other hand, will not encourage courts to grant injunctions or damages. Some will frame the dispute as being between those entitled (by traditional copyright) and those empowered (by the new tools). The extensive use of open licenses (such as Creative Commons licenses) for content shared online might further complicate the issues. This is clear for the on-going Copilot dispute as most of the code on GitHub is probably open source and therefore reusable, meaning that the plaintiffs have to rely on their right of attribution guaranteed by permissive licenses.

The outcome of the tensions between AI companies and creators might change if artists are able to unite and design "a new era of consenting interactions around artist data", thanks to tools able to verify the training (e.g. haveibeentrained.com[13]) and to express an opt-out. But will artists organize and be able to agree with tech companies on this new opt-out mechanism?[14] Or should other solutions be considered, such as a compulsory license for training AI? The debate is just starting.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

[13] *See* https://spawning.ai.

[14] Pranav Dixit (2023) "Meet the Three Artists Behind a Landmark Lawsuit Against AI Art Generators", BuzzFeedNews, 20 January 2023. At the end of 2022, Stability AI made clear that artists not wanting their creation used for training purposes could opt out, but artists responded that the default should rather be opt in.