THE CREATIVE WAR AGAINST AI ART: COPYRIGHT INFRINGEMENT, EXPLOITATION OF ARTISTS AND PROTECTING ARTISTIC INTEGRITY

TSUI, YU HEI IRIS
Common Law LLB and English Literature, College of Arts and Humanities

ABSTRACT

With the expanding use of generative AI tools like Midjourney, Stable Diffusion and DALL-E in the creative industry also comes looming legal and ethical concerns. Generative AI is trained on large datasets created by human artists, an unethical process that involves scraping copyrighted material from the Internet without consent or compensation. This article argues how, contrary to the claims that it 'democratises' art, generative AI has the opposite effect of automating and privatising the creative process, taking protections away from individual artists, while also failing to replicate the artistic cognition that makes creativity morally valuable. However, how copyright law has already been evolving to catch up with AI's development by defining how AI art can be considered theft. From the US Copyright Office's decisions on AI creations Zarya of the Dawn and A Recent Entrance to Paradise, which highlighted the importance of human authorship; to the European Court's 2017 ruling that Pirate Bay was infringing copyright by encouraging users to make copies of protected works; and finally to the lawsuits against Stability AI and OpenAI, what the law has shown is that it does not afford the same rights to machines as it does to humans.

INTRODUCTION

Since its foundation in 2015, OpenAI has produced the now ubiquitously used large language model ChatGPT, soon followed by fellow household names DALL-E, Sora, and Musenet. In 2022, DeviantArt followed suit with its creation of image-generation tool, DreamUp.ai, and in the same year, similar text-to-image generators Stable Diffusion and Midjourney added themselves to the plethora of Generative AI (GenAI) art platforms.

Consequently, they have also prompted a fierce and impassioned backlash from artists who, having spent years honing their craft, now find themselves increasingly replaced by vastly efficient, almost instantaneous mechanical counterparts that generate art with the ease of text prompts typed into a search bar. But beyond the typical objections of AI taking over jobs and incomes, the glamour of GenAI's efficiency conceals a corrupt and unethical underbelly, perpetrated by gross negligence on the part of companies, which takes advantage of the loopholes in copyright law and profits off copyrighted works of human artists.

This exploitation is revealed by a very simple question: where do GenAI tools obtain their images and artwork from? Three elements make up this process: 'web scraping', 'machine learning', and 'diffusion models'. 'Web scraping' is a process of collecting freely available information across the Internet, usually used by businesses to analyse end-to-end customer service processes and identify priority objectives (Khder, 2021). 'Machine learning' consists of giving computers a huge dataset and training it to generate images, text or music in response to a prompt; generator is forced to adjust its estimates every time it gets an image wrong until it eventually tine-tunes is images to suit preferences (Berryman, 2024, p.2). Finally, 'diffusion models' operate through 'a probabilistic process in which the AI system iteratively reconstructs the original data from a series of noise-infused versions' (Du et al, 2024, p.178). None of these are illegal or problematic on their own; but when combined, they could hinge on breaching copyright law (Appel et al, 2023). Here is a demonstration of how this occurs: GenAI programmes such as Midjourney and Stable Diffusion are trained on a library of 5 billion URL links collated by the non-profit organisation LAION. None of these programmes see images the way humans do: they simply diffuse a stream of computer generated noise (variations of text, images, or music) into what we might recognisably call art.

It is this process that has provoked artists for many reasons which can be condensed into three categories. First, most obviously, is the knock-on effect that AI has had on their income and jobs, as well as its devaluation of the years of effort artists have poured into their work in order to hone their skills (Goetze, 2024). Second, it turns the entire concept of artistic cognition on its head by removing the exchange between artist and audience that makes art what it is. Third, the fact that GenAI companies are exploiting loopholes in the copyright law framework. David Holz, founder of Midjourney, stated: 'To my knowledge, every single large AI model is basically trained on stuff that's on the Internet. And that's okay, right now. There are

no laws specifically about that. [...] And it took like 20 or 30 years for it to really become something that the legal system is starting to figure out' (Claburn, 2022). Until the law catches up, GenAI will continue to scrape the copyrighted works of human artists without legal repercussions.

THE CLAIM THAT AI 'DEMOCRATISES' ART

By its supporters, GenAI is acclaimed for its ability to 'democratise' art by lowering the barriers of entry and giving laypeople the tools to create art without having to undergo the lengthy, arduous process of learning creative skills. It is precisely this claim that has prompted a backlash from artists: the deep fear and outrage that they will be replaced by mechanised, automated counterparts (Crabapple, 2023). The rise of GenAI does not herald the end of human art, because artists create art out of personal love for the process and not for financial incentives, but it means that companies will turn to GenAI as a more cost-efficient option of producing art, pushing professional, skilled artists out of jobs and exposing them to exploitation (Goetze, 2024). Contrary then to what GenAI supporters claim, GenAI enables companies to reduce costs and commoditise art at an unprecedented level, not 'democratising' but rather privatising art.

The claim that AI 'democratises' art by allowing people to forgo the process of learning creative skills completely misses the point of artistic creation. Creation demands a process of inspiration, development, creation, and reflection, all of which demand a level of artistic cognition. This is what makes human art morally valuable and its protection legally essential, and this is what AI cannot replicate. AI does not have the consciousness to understand the difference between a lightbulb or a sun; it only diffuses computer generated noise into an arbitrary configuration of pixels, numbers, or words. Harvard professor and philosopher Sean Dorrance Kelly (2019, p.72) argues that to equate the human mind with AI is to suggest that the human brain operates according to a set of computational algorithms. We may be able to recognise the greatness in an AI artwork, but 'if we know that the output is merely the result of some arbitrary act or algorithmic formalism, we cannot accept it as the expression of a vision for human good.' AI has no control over its final product, because it is incapable of understanding the components of the artistic process. This element of control is a central tenet as to how copyright law protects human artists but not AI, as will be discussed later.

Crucial to the creative process is artistic reflection, which is brought about by the audience's criticism. The artist creates and presents his art, and the critic responds by engaging with and evaluating it, either by examination of its content or its form. It is for this reason that literary and artistic theories have been developed as movements throughout history, from 18th century Romanticism to 20th century New Criticism. As art critic David Levi Strauss (2012) argues, art 'needs something outside of itself as a place of reflection, discernment, and connection with the larger world.' It is this constant two-way conversation between the art we see and the art we make that the law recognises as worth protecting. Therefore, it is the lack of this exchange that distinguishes AI's creative process from our own; it does not engage with the art as it learns, but merely capitalises on it, copying styles of human artists.

Finally, because AI is trained on the work of human artists, it fails to innovate. In order to push boundaries, artists must first have cognition of conventions and traditions. With this understanding, artists can experiment, challenge, and break these rules to create something new. This was how Einstein discovered relativity theory and Picasso created cubism. This is something that Margaret Boden (1990, p.6) names 'transformational creativity', something that AI cannot replicate because it fails to engage with and internalise those old rules, and how to skilfully break them. Musical theorist Arnold Schoenberg, for example, invented a mathematical technique for music composition by understanding the public conception of music and breaking the rule: as Kelly (2019, p.71) argues, '[i]t is only creativity of this tradition-defying sort that requires some kind of social sensitivity', a responsiveness to social necessity that a machine cannot replicate. AI simply lacks the cognition to recognise traditional rules and the responsiveness to speak to what is needed now.

For these reasons, GenAI does not 'democratise' art by lowering the parameters to entry; instead, it digitises the work of human artists and puts this into the public domain without providing the original creators with compensation. The law has been slow to protect creators, but has begun to recognise the differences between AI- and human-generated art. It does so through the legal concept of 'human authorship'.

WHO KILLED THE ARTIST? THE VALUE OF ARTISTIC COGNITION

Built into the legal concept of 'human authorship' are two understandings: first, art is expression, and second, a human created it (US Copyright Office, 2021). AI clearly fails the first condition: as explained, it cannot express, and therefore should not be legally protected in the same way. The second condition concerns the requirement that a work must have human

involvement at its core in order for it to be copyrightable. This concept has prevailed across various legal systems, but the US will be taken as the primary example here.

The argument mounted by AI users is that they can be considered the artist. This was what artist Stephen Thaler and web artist Kris Kashtanova claimed concerning the copyright of their works 'A Recent Entrance to Paradise' (Thaler v Secretary of Commerce (2022)) and Zarya of the Dawn. Both of their cases were dismissed by the US Copyright Office (2021 and 2023), which stated that the images, produced by the technology of Creativity Machine and Midjourney respectively, were 'not the product of human authorship'. The US Copyright Act (1958) definitively excludes non-humans from the definition of 'author', underpinned by Judge Howell's understanding that, despite the adaptability of original works of authorship into various mediums of expressions, 'human creativity is the sine qua non at the core of copyrightability' (Thaler v Secretary of Commerce (2022, p.8)).

This is a crucial element in its reasoning: US law recognises that there must be a strong relationship of control between the artist and their art. That AI artists have no idea how the detail of their images is generated suggests a huge disconnect between artist and art, which in turn demonstrates a lack of control AI has over its product. As the US Copyright Office (Federal Register 2023) reasoned, 'When an AI technology receives solely a prompt from a human and produces complex written, visual, or musical works in response, the traditional elements of authorship are determined and executed by the technology – not the human user.' It is AI that determines the expressive elements of its output, not the human inputting word prompts; because the law does not give AI the same rights it affords to humans, AI art cannot be copyrighted.

There is a reason why we do not afford the same rights to machines as we do to humans: AI fails to replicate what makes art morally valuable and its protection legally essential. It does not build on the exchange between artist and audience, the interaction between the art we make and the art we experience; it simply diffuses computer generated noise into a configuration of pixels, notes or words that resembles the patterns of the human artists it stole from. Integral to the courts' reasoning is that art exists to further the public good by incentivising individuals to create and invent. This is demonstrated by the contrasting decisions in the US case Authors Guild v Google (2015) and the European Court of Justice's decision Stichting Brein v Ziggo BV, XS4ALL Internet BV (2017). Both concerned a digital platform (Google, and The Pirate Bay) making original human-created artwork available to laypeople. Google won its case, but interestingly, Pirate Bay did not. Three important factors distinguish these two cases. First, Google advanced public interests by facilitating access to a huge range of books and generating new audiences, increasing commercial benefits for authors and publishers; second, Google's copying fell within the 'fair use' doctrine due to its main purpose of advancing education and research; third, Google's digitisation of the work did not substitute the market for the originals, but enhanced the commercial profits for those authors. But Pirates Bay goes beyond these: it classifies works based on their genre and popularity, filters content, and deletes faulty files, actively encouraging users to make copies of those works without compensating the original artists.

The law recognises the importance of restricting protection of copyrighted works only to the artistic exchange between creator and audience. The law wants to make art free because it understands that, while machines do not need an incentive to create, humans do; and that incentivising individuals to innovate furthers the public good.

THE EVOLUTION OF COPYRIGHT LAW: DEFINING THE PARAMETERS

How much protection does the law offer human artists? At present, the answer is not very much. Copyright law is often reluctant to find GenAI platforms liable for vicarious infringement, requiring them to have done two things: one) knowledge of the infringement, and two) material contribution to it, for example that the platform itself was designed to steal.

Despite all of this, however, the foundations for the legal protection of artistic integrity are there. Essentially, within copyright law, AI art created by GenAI can already be considered theft. As AI art is built on the art of millions of human artists, and incorporates those artists' intellectual property into its own designs, GenAI platforms require a licence in order to avoid copyright liability which they do not have. Working on this premise, the law has been developing to define the parameters of how far GenAI can go without infringing copyright.

Turning first to legislation, statutes have required GenAI companies to comply with requirements of transparency. The EU has taken the first step towards setting the guardrails through the AI Act which came into force in August 2024. Chapter V, Article 53, specifically addresses general-purpose AI (GPAI) models, subjecting their providers to legal obligations condensed into two requirements. First, their copyright policy must comply with EU law; and if they feel they fall under a text or data mining exception, they must apply for an 'opt-out' reservation under Article 4 of the EU Digital Market Directive 2019. Second, providers are obliged to 'draw up and make publicly available a sufficiently detailed summary about the content used for

training of the general-purpose AI model' (2024). The Act provides standards that GenAI companies must meet in order to continue operating under copyright law: providing technical documentation for evaluation, ensuring that information is made publicly available, and putting in place a policy compliant with Union copyright law. The US has similarly undertaken the legislative route by introducing the 'Bill for the Generative AI Copyright Disclosure Act' (U.S. Congress, 2024). If implemented, it would require AI companies to submit any copyrighted works in their training datasets to the Copyright Office before releasing new generative AI systems. This would not ban AI from training on copyrighted material but would at least be a huge step in data transparency, and will empower copyright owners to take action against unauthorised use of their works.

Even before these statutes came into place, artists and creative corporations have begun to fight back in legal forms themselves: most prominently, two 2023 lawsuits brought by The New York Times against OpenAI and Microsoft, and Getty Images against Stable Diffusion. In its complaint, The Times (2023, para 64) explained the exploitation that OpenAI's commercial success is built on: what drives the widespread use of ChatGPT is the LLMs' 'ability to produce natural language text in a variety of styles', a result achieved by making numerous reproductions of copyrighted works owned by The Times and its peers in the course of training the AI. The ability of journalistic institutions to monetise their content is dependent on its exclusive rights of reproduction, adaptation, publication, performance, and display under copyright law; and it is these rights that GenAI companies negligently and grossly violate. As a result, GPT models could substitute the market for journalism by obviating the need to purchase access through The Times itself (2023, para 49). In its complaint, The Times also exposes an important flaw in the argument that AI will one day take over the creation of art: AI is in fact dependent on human art; and by draining revenue out of journalistic and artist sectors, AI creates a risk that new stories will go untold.

Similarly, Getty Images has made several allegations against Stability AI for committing copyright, database, and trademark infringement. Two of its claims are, first, that Stability AI has used its images as data inputs for the purposes of training and developing Stable Diffusion; and second, that it is responsible for secondary copyright infringement on the basis that Stable Diffusion has unlawfully imported the pre-trained Stable Diffusion software into the UK (Getty Images v Stability AI (2023)). The High Court refused Stability AI's application to dismiss these claims, and the case is now awaiting trial. As many law firms have noted, its outcome will have far-reaching implications: if Getty Images wins its case, it could potentially prevent developers from 'jurisdiction shopping' to train AI models in more favourable legislative landscapes around the world before importing them into a different jurisdiction.

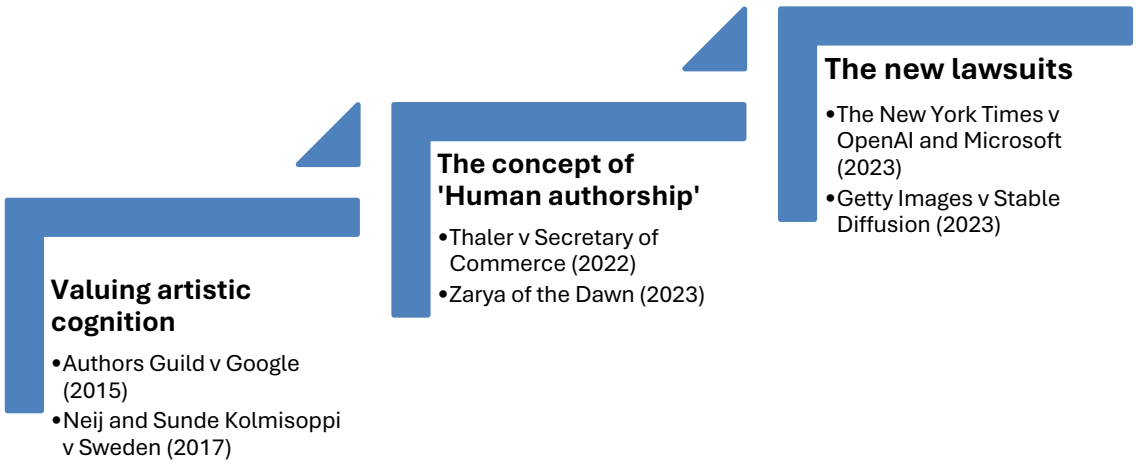

Figure 1: Timeline of relative case law

These lawsuits give courts a chance to shape the law in a way that protects copyright and artists' livelihoods. Taking the cue from increased court cases and artist strikes, governments have also begun making movements to negotiate the interests of creative industry stakeholders and AI companies, giving courts guidance in striking a balance between advancing the development of AI and technology and respecting the rights of copyright holders.

CONCLUSION: THE LOOMING AI DYSTOPIA

AI has assumed roles in society that have caused some disturbances, for example, taking over jobs in agricultural and industrial sectors, but nothing has been more troubling than its growing influence over the creative industry. Yet, the danger is not so much that art will die. The danger then is that GenAI corporations will, by infringing copyright and privatising the creation process, suck the lifeblood out of human artists.

As governments begin demanding transparency from GenAI platforms, and as creative institutions take legal action to protect their copyright, the law will continue to define the parameters to limit the use of AI in the creative sector (Mansell, 2021). This is not the first time digital technology has exacerbated existing social issues; but as with any societal crisis, regulations must keep pace with AI's development in order to ensure that rights are afforded to humans, and not to machines.

REFERENCES

Appel, G., Neelbauer, J. and Schweidel, D.A. 2023. '*Generative AI Has an Intellectual Property Problem'* [Online]. [22 February 2025]. Available at: https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem

Authors Guild, Inc. v Google, Inc. (No. 13-4829, 16 October 2015).

Berryman, J. 2024. Creativity and Style in GAN and AI Art: Some Art-historical Reflections. *Philosophy & Technology* **37**(61), pp.1-17.

Boden, M. 2003. *The Creative Mind: Myths and Mechanisms*. 2nd edn. London: Routledge.

Crabapple, M. 2023. Restrict AI Illustration from Publishing: An Open Letter. [Online]. [Accessed 25 February 2025]. Available at: https://artisticinquiry.org/AI-Open-Letter

Du, H., Zhang, R., Niyato, D., Kang, J., Xiong, Z., Kim, D.I., Shen, X., and Poor, H.V. 2024. Exploring Collaborative Distributed Diffusion-Based AI-Generated Content (AIGC) in Wireless Networks. IEEE Network **38**(3), pp.178-186.

European Union Artificial Intelligence Act 2024 (c.5). OJ L, 2024/1689, 12.7.2024 [Online] [Accessed 14 November 2024] Available at: https://artificialintelligenceact.eu/the-act/

Federal Register (Vol. 88, No. 51) 2023. [Online]. [Accessed on 14 Nov 2024]. Available at https://www.govinfo.gov/content/pkg/FR-2023-03-16/pdf/2023-05321.pdf

Getty Images (US) Inc v Stability AI Ltd [2023] EWHC 3090 (Ch).

Kelly, S.D. 2019. 'A philosopher argues that an AI can't be an artist: Creativity is, and always will be, a human endeavour.' *MIT Technology Review* **122**(2), pp.68-75.

Khder, M.A. 2021. Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *International Journal of Advances in Soft Computing and its Applications* **13**(3), pp.145-168.

Stichting Brein v Ziggo BV and XS4ALL Internet BV (C-610/15) Court of Justice of the European Union.

Mansell, R. 2021. Enclosing or democratising the AI artwork world. *Cambridge Journal of Law, Politics, and Art,* 1, pp.247-251.

Robert J. Kasunic, US Copyright Office, Zarya of the Dawn (Registration 3VAu001480196) (2023).

Claburn, T. 2022. David Holz, founder of AI art generator Midjourney, on the future of imaging: Optimizing for beauty while trying to suppress sensationalism [Online]. [Accessed on 21 January 2025]. Available at: https://www.theregister.com/2022/08/01/david_holz_midjourney/

Strauss, D.L. 2012. *'FROM METAPHYSICS TO INVECTIVE: Art Criticism Still Matters*'* [Online]. [Accessed on 11 November 2024]. Available at: https://brooklynrail.org/2012/05/art/from-metaphysics-to-invective-art-criticism-as-if-it-still-matters/

Thaler v Vidal, under Secretary of Commerce for Intellectual Property and Director of the United States Patent and Trademark Office (2022) United States District Court for the District of Columbia (2021-2347) Supreme Court of the United States. [Online]. [Accessed on 27 November 2024]. Available at: https://www.supremecourt.gov/search.aspx?filename=/docket/docketfiles/html/public/22-919.html

United States Copyright Office (2021) The Compendium of U.S. Copyright Office Practices of 2021, Chapter 306.

The New York Times Company v Microsoft Corporation and OpenAI and Others (Complaint) (2023) (United States District Court, Southern District of New York).

United States Copyright Office Review Board. 2023. Decision Affirming Refusal of Registration of a Recent Entrance to Paradise.

United States Copyright Office. 2023. Cancellation Decision re: Zarya of the Dawn (Registration #VAu001480196).

United States Congress (2024) Generative AI Copyright Disclosure Act of 2024, H.R.7913.

U.S. Congress. (1958) United States Code: Copyright Act, 17 U.S.C. §§ 201-216 (1958).