



NUS Law Working Paper No 2023/025

Good Models Borrow, Great Models Steal: Intellectual Property Rights and Generative AI

Simon Chesterman

chesterman@nus.edu.sg

[October 2023]

© Copyright is held by the author or authors of each working paper. No part of this paper may be republished, reprinted, or reproduced in any format without the permission of the paper's author or authors.

Note: The views expressed in each paper are those of the author or authors of the paper. They do not necessarily represent or reflect the views of the National University of Singapore.

Good Models Borrow, Great Models Steal: Intellectual Property Rights and Generative AI

Simon Chesterman*

Preliminary draft for LKYSPP Conference October 2023, and later submission to *Policy & Society Journal* Special Issue: Governance of Generative Artificial Intelligence

This article addresses three critical policy questions that will determine the impact of generative AI on the knowledge economy and the creative sector. The first concerns how we think about the training of such models — in particular, whether the creators or owners of the data that are “scraped” (lawfully or unlawfully, with or without permission) should be compensated for that use. The second question revolves around the ownership of the output generated by AI, which is continually improving in quality and scale. These questions are inherently linked to the realm of intellectual property, a legal framework designed to incentivize and reward human creativity and innovation. For instance, the United Kingdom has historically maintained a distinct category with limited rights for new “computer-generated” works, while Singapore recently introduced an exemption allowing for computational data analysis of existing works. The third section of this article explores the broader implications of these policy choices, weighing the advantages of reducing the cost of content creation and the value of expertise against the potential risks to various careers and sectors of the economy, which may be rendered unsustainable. Some lessons might be found in the music industry, which also went through a period of unrestrained piracy in the early digital era, epitomized by the rise and fall of the file-sharing service Napster. Similar litigation and legislation may help navigate the present uncertainty, along with an emerging market for “legitimate” models that respect the copyright of humans and are clear about the provenance of their own creations.

* National University of Singapore and AI Singapore. Email: chesterman@nus.edu.sg. I’m grateful to N Chesterman, Mikhail Filippov, Gao Yuting, Hahn Jungpil, and David Tan for comments on earlier versions of this text. This research is supported by the Ministry of Education, Singapore, under its MOE AcRF TIER 3 Grant (MOE-MOET32022-0001).

Introduction	2
1 I Think, Therefore I’m Paid	3
2 Author, Author!	8
3 Brave New World?	13
4 Conclusion	14
References	17

Introduction

When people think of the risks associated with artificial intelligence (AI), Hollywood looms large. Movies have long conjured the worst-case scenarios: from Hal refusing to open the pod bay doors in *2001*, to a murderous Arnold Schwarzenegger travelling back through time. If there is a robot apocalypse, however, it is unlikely to resemble a *Terminator* movie. A more probable scenario is what was recently seen off-screen in — ironically enough — the Writers Guild of America (WGA) strike of 2023.

Hollywood’s scriptwriters were protesting, in part, about the threat of many jobs being replaced by new generative AI tools that can perform similar functions at little or no cost. The concern is not that humanity will wake up to discover that it has been replaced by AI; rather, it is that AI will progressively reduce the economic viability of certain careers by salami-slicing fulltime jobs into tasks that can be commoditised and outsourced. This can be thought of as the dark side of the gig economy (Prassl 2018). Where Uber, Grab, and the like offered flexible arrangements that were attractive for young workers who later discovered that no foundation had been laid for a career, ChatGPT threatens to take existing careers and break them into gig work for hire.

Such precarity is not limited to scriptwriters. After an initial panic by academics worldwide that this new technology might enable students to cheat on their papers, it became clear that generative AI had larger implications for the knowledge economy, comparable perhaps to the impact of the industrial revolution on manufacturing. “Knowledge workers” was the term introduced in 1959 by management consultant Peter Drucker for non-routine problem solvers (Drucker 1959). People who “think for a living” earn through their ability to analyse and write – something that ChatGPT can replicate in almost no time and at almost no cost. Journalists, already taking a beating as readers turn from traditional to social media, now face the prospect of technology taking over the writing task as well. Yet that same threat confronts anyone who analyses or writes for a living, such as lawyers and even – gasp –

academics. Applications are not limited to prose, as ChatGPT has demonstrated proficiency in coding as well as poetry (Dwivedi, et al. 2023). Similar developments have shaken the art world, with generative AI images flooding social media and, increasingly, traditional media. Video and multimodal content is close behind.

This article will consider three policy questions facing governments around the world in relation to how generative AI will impact the knowledge economy and the creative sector. The first concerns how we think about the training of such models, in particular whether the creators or owners of the data that are “scraped” — lawfully or unlawfully, with or without permission — should be compensated for that use. The second question is who (if anyone) should own the output of generative AI, which is being produced at ever greater quality on ever greater scale. Both issues are linked to intellectual property, a body of laws that was adopted to incentivize and reward human creativity and innovation. Section three of the article considers the larger implications of the answers to those questions, weighing the benefits of lowering the cost of creation and the value of expertise against the possibility that diverse careers and sectors of the economy may be rendered unsustainable (Mims 2023).

1 I Think, Therefore I’m Paid

AI has always depended on access to data (Roberts, et al. 2021). Large language models (LLMs) in particular are trained on huge datasets, comprising publicly available material as well as copyrighted and pirated material available online (O’Leary 2013; Zikopoulos, et al. 2012). That scale transformed public debate about the impact of AI with the release of ChatGPT by OpenAI in November 2022, quickly followed by competitors such as Google’s Bard, Anthropic’s Claude, and Meta’s Llama. Excitement and trepidation about the uses for systems able to respond to natural language queries with human-like responses — in text as well as images — suggested that the long-heralded economic promise of AI might be at hand. Goldman Sachs breathlessly reported that generative AI could increase global GDP by seven percent (2023).

How (if at all) should the rights of creators, whose text and images train such models, be recognized and compensated? The use of pirated or illegally obtained material appears at first blush to be a simple case of theft of intellectual property, but has been notoriously difficult to prove. Around the world, concepts like fair use are being stretched by the wholesale consumption of books, photographs, and other materials. In some jurisdictions,

new rights to data-mine have sought to balance the interests of developers against those of creators.

If a work is not in the public domain, even temporary unauthorized use can be an infringement. This is the subject of ongoing litigation brought by Getty Images against Stability AI, for example, alleging that the Stable Diffusion model was trained on millions of copyrighted images and metadata. Getty claims that this deprived it of the revenue from licensing those images. Evidence of the alleged infringement includes content generated by Stable Diffusion with distortions of the watermark Getty uses to protect its product.

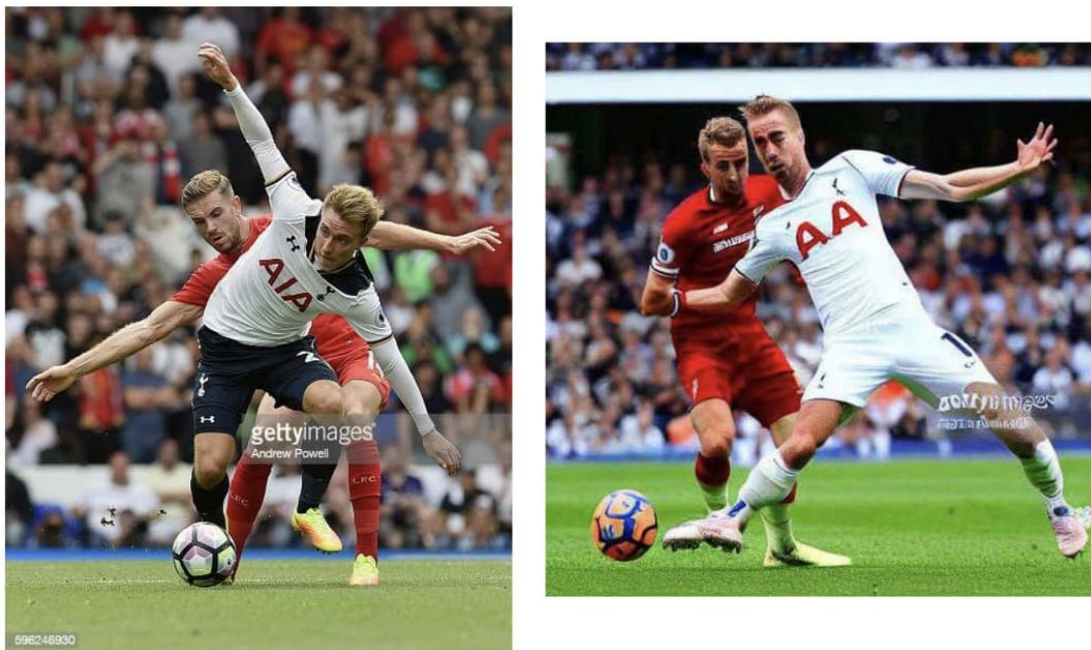


Figure 1. From the complaint in Getty Images litigation, Getty image (left) and Stability AI image (right)

Given the secretive nature of much model training, proving infringement is rarely as easy as this. Even in the Getty case, it appears possible that infringement may need to be demonstrated on a case-by-case basis, establishing substantial similarity for each image one by one — rather than the systemic infringement alleged by Getty (Tan 2024a).

Even if infringement can be established, fair use is a defence that balances the rights of creators and the interests of the wider public in distributing and using their works. It generally considers the purpose of the use, the nature of the work, the amount used, and the effect on the market for the original work. When an individual records a televised broadcast to watch at a later time, for example, that can be considered fair use. Projecting such a recording for an audience and charging tickets, by contrast, would not be (Beebe 2008).

An example of the current state of debate is the 2023 US Supreme Court case involving two versions of an image of the late musical artist Prince. Lynn Goldsmith took the original photograph in 1981.



Figure 2: A black and white portrait photograph of Prince taken in 1981 by Lynn Goldsmith

Three years later, Vanity Fair licensed the photograph to be used as source material by Andy Warhol for an illustration of an article entitled “Purple Fame”. Warhol’s silk screen image was one of more than a dozen he created, despite the license — for which Goldsmith was paid \$400 — being limited to one image to be used one time only in the magazine.



Figure 3: A purple silkscreen portrait of Prince created in 1984 by Andy Warhol to illustrate an article in Vanity Fair

After Prince’s death in 2016, the other images and two pencil drawings, collectively referred to as the “Prince Series” were published, including on the cover of another magazine owned by Vanity Fair’s parent company, Condé Nast.



Figure 4: An orange silkscreen portrait of Prince on the cover of a special edition magazine published in 2016 by Condé Nast

Litigation followed, with the Supreme Court ultimately concluding that, while Warhol's art might be fair use if hung in a museum, using the image for a magazine cover was precisely the kind of purpose for which Goldsmith licensed her own photos. She was therefore entitled to compensation (*Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith* 2023).

Returning to generative AI, a key question is whether using data to train models, which are then used to produce works that directly compete with the authors of those data, constitutes fair use. This appears to be distinct from other forms of data mining. When Google began scanning vast quantities of books in 2002, there were challenges that this infringed copyright. Google was, for the most part, successful in arguing that it made the information available but was not itself providing a substantial substitute or competing with the market for the original works (*Authors Guild v. Google* 2015; Maguire 2020).

The ability of generative AI to produce text and images that may, in fact, compete directly with past and present works produced by the authors and artists whose works trained those models is central to several of the lawsuits currently underway, including prominent authors such as John Grisham, Jonathan Franzen, and Elin Hilderbrand who are suing OpenAI, the creator of ChatGPT (Alter and Harris 2023; Reisner 2023b).

Mark Lemley, among others, has argued that model training should be regarded as fair use on the basis that machine learning is a transformative use of the underlying data. He and coauthor Bryan Casey also argue that this will encourage the creation of new databases with greater transparency, as well as recognizing that licensing materials for such large training sets is impractical given their scale (Lemley and Casey 2021). Lemley, who is part of Stability

AI's defence team, has gone on to argue that the infringement question may be inapplicable to generative AI "for the simple reason that generative AI is not about copying existing works but about creating new ones" (Guadamuz 2023; Lemley 2023).

In the absence of statutory reform, lawsuits are likely to proliferate.¹

Singapore is an example of a jurisdiction that has tried to thread this needle through legislation. Amendments to its Copyright law in 2021 include a permitted use to make a copy of a work for the purpose of "computational data analysis", which includes extracting and analysing information and using it to "improve the functioning of a computer program in relation to that type of information or data" (Copyright Act 2021, ss. 243-244).

The provision still requires lawful access to the underlying data, but appears more open to datamining and model training than traditional conceptions of fair use (Lim 2023) or the "non-commercial" text and data analysis exception adopted in the United Kingdom in 2014 (Copyright, Designs and Patents Act 1988, s. 29A). An information sheet produced by the Intellectual Property Office of Singapore (IPOS) explicitly states that the provision is intended to allow "training machine learning" (IPOS Factsheet 2022). Yet, analysing text or images for the purpose of making recommendations or optimising workflows is quite distinct from using those text and images to generate more text and images. The difference is not just the usage, where copying is central to the process, but also the economic impact of that usage (Tan 2023; Torrance and Tomlinson forthcoming).

This is no longer a hypothetical problem. In addition to the possibility of diluting human authors' works, it is possible that they will simply be swamped by the volume of generative AI produced. An early example was the science fiction magazine *Clarkesworld* had to shut down unsolicited submissions because it was being flooded with AI content (Silberling 2023). Amazon, which is now one of the world's largest publishers of books, was becoming so overwhelmed by submissions that it now imposes a limit that its self-published authors may "only" publish three books per day (Creamer 2023).

Returning to the question of lawful access, much of the data used by LLMs for training is pirated in the first place. More than 70,000 pirated books were found when Peter Schoppert analysed the "Books3" dataset (Reisner 2023a; Schoppert 2023). No one is seriously suggesting that generative AI should not be trained. But it is reasonable to expect that

¹ #update <https://aicopyright.substack.com/p/i-will-get-an-order-out-when-i-get>
<https://originality.ai/blog/openai-chatgpt-lawsuit-list>

models are not trained on stolen data, and that those who profit from this technology pay something to the creators whose works serve as its fuel (Tan 2024b).

2 Author, Author!

A second set of questions concerns who should own the outputs of generative AI. In an unscientific experiment, the author decided to ask ChatGPT itself and got two very different answers:² “I do not have the ability to own intellectual property or any other legal rights,” ChatGPT replied at first. “Any text or other content that I generate is the property of OpenAI, as the creator and owner of the tool that I am.”

The author pointed out that OpenAI itself now explicitly states that it will not claim copyright over any content generated by ChatGPT (Ellison 2022; Guadamuz 2022; Schade 2023).³ This led to a revised answer: “The text generated is not the intellectual property of the model itself. Instead, the intellectual property rights belong to the person or entity who has commissioned the model to generate the text.”

Clear and concise, but also wrong.

In most jurisdictions, automatically generated text does not receive copyright protection at all. The U.S. Copyright Office has stated that legislative protection of “original works of authorship” is limited to works “created by a human being” (17 USC § 102(a)). It will not register works “produced by a machine or mere mechanical process that operates randomly or automatically without *any* creative input or intervention from a human author.” (Compendium of U.S. Copyright Office Practices, 3rd edition 2019) (emphasis added). The word “any” is key and begs the question of what level of human involvement is required to assert authorship (Gervais 2020; Phelan and Carey 2023).

Early photographs, for example, were not protected because the mere capturing of light through the lens of a *camera obscura* was not regarded as true “authorship” (de Cock Buning 2018, p. 524). It took an iconic picture of Oscar Wilde going all the way to the US

² The “conversation” was conducted in January 2023 with ChatGPT’s publicly available model at the time; responses have been edited for brevity.

³ OpenAI had, in fact, initially claimed ownership of all output from DALL·E. This was amended in late 2022 to state that users now own the generated content, a position extended to the product of ChatGPT.

Supreme Court before copyright was recognised in mechanically-produced creations (*Burrow-Giles Lithographic Co v. Sarony* 1884).



Figure 5: Oscar Wilde, in the Smithsonian Magazine, May 2004

Arguments continued in other jurisdictions, however, with Germany withholding full copyright of photographs until 1965 (Nordemann 1999).

The issue today is distinct: not whether a photographer can own images passively captured by a machine, but who might own new works actively created by one. Computer programs like word processors do not own the text typed on them, any more than a pen owns the words that it writes. But AI systems now generate news reports, compose songs, paint pictures. These activities generate value — can and should they be protected by the law?

At present, the answer in most places is no. Unless there is an identifiable human author, copyright will not apply. The policy behind this is often said to be incentivizing and rewarding innovation. This has long been dismissed as unnecessary or inappropriate for computers. “All it takes,” Pamela Samuelson wrote in 1986, “is electricity (or some other motive force) to get the machines into production” (Samuelson 1986, p. 1199).

Indeed, protecting such works might *disincentivise* innovation — by humans, at least. AI has already unleashed an economic tornado in the art world, massively lowering the cost of producing original images (Menéndez 2023). If we wish to have a thriving arts sector that gainfully employs humans, it is arguable that their creations should be protected while machine creations should not be. Automatically generated content may not be eligible for copyright protection, but edited and curated content that draws on such material could still be owned by the person doing the editing and curating.

The author fed that into ChatGPT, which agreed that this was correct — sensibly adding that legal advice should be sought if there were any further questions.

An alternative approach, adopted in Britain is to have more limited protections for “computer-generated” work, the “author” of which is deemed to be the person who undertook “the arrangements necessary for the creation of the work”. “Computer-generated” is defined as meaning that the work was “generated by computer in circumstances such that there is no human author of the work” (Copyright, Designs and Patents Act 1988). Similar legislation has been adopted in New Zealand (Copyright Act 1994), India (Copyright Amendment Act 1994), Hong Kong (Copyright Ordinance 1997), and Ireland (Copyright and Related Rights Act 2000). Though disputes about who took the “arrangements necessary” may arise, ownership by a recognized legal person or by no one at all remain the only possible outcomes (Brown, et al. 2019, pp. 100-01; *Nova Productions v. Mazooma Games* 2007). The duration is generally for a shorter period, and the deemed “author” is unable to assert moral rights — including the right to be identified as the author of the work (Copyright, Designs and Patents Act 1988).

A World Intellectual Property Organization (WIPO) issues paper recognized the dilemma, noting that excluding these works would favour “the dignity of human creativity over machine creativity” at the expense of making the largest number of creative works available to consumers. A middle path, it observed, was to offer “a reduced term of protection and other limitations” (Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence 2020). Several commentators have suggested similar approaches (Abbott 2020, pp. 71-91; du Sautoy 2019, p. 102).

As human authorship becomes more ambiguous, that middle ground may help preserve and reward flesh and blood authorship, while also encouraging experiments in collaboration with our silicon and metal partners.

Europe is actively considering such a measure (Séjourné 2020). The Singapore Academy of Law’s Law Reform Committee proposed something similar in 2020 (Rethinking Database Rights and Data Ownership in an AI World 2020), but only traditional human authorship remains recognised under the new Copyright Act adopted the following year. AI-assisted works may still warrant protection if there is a causal connection to a human exercising input or control, though determining the threshold for that connection is left to the courts (Tan and Tan 2022).

An indication of the difficulty can be seen in the case of Jason M. Allen, who was denied protection for the work “Théâtre D’opéra Spatial”, which won first prize at the Colorado State Fair in 2022 — before he revealed that it was created using Midjourney.



Figure 6: Théâtre D’opéra Spatial — Midjourney, using a prompt from Jason M. Allen

In his subsequent request for copyright protection, he claimed that he revised his prompts “at least 624 times” to achieve the final work, which he also edited with Photoshop. The U.S. Copyright Office Review Board found that, because Allen was “unwilling to disclaim the AI-generated material”, it was unable to recognize the work (US Copyright Office 2023).

This was consistent with other high-profile examples of AI-generated works that have been denied protection, such as images generated by Midjourney for Kris Kashtanova’s graphic novel *Zarya of the Dawn* (while allowing protection for human-arranged portions of the work) (Edwards 2023) and Stephen Thaler’s AI-generated “A Recent Entrance to Paradise” (Brodkin 2023).



Figure 7: A Recent Entrance to Paradise

Thaler has been a frequent litigant in efforts to persuade courts and intellectual property offices that AI systems themselves can create and own patents and copyrightable works. Despite brief successes in Australia and South Africa (the former reversed on appeal), it remains the case — for the time being — that AI systems themselves can neither create nor own copyrightable or patentable works in their own right (Chesterman 2020; Padmanabhan and Wadsworth forthcoming).

3 Brave New World?

Generative AI has the potential to transform the arts as well as the knowledge economy. The precise impact is presently unknowable, with a recent study suggesting a “jagged frontier” of innovation across different fields based on a survey of complex, realistic, and knowledge-intensive tasks (Dell'Acqua, et al. 2023). Protecting IP rights too strictly could hinder the development of new tools and works enhanced by AI; failing to protect those rights could render millions of jobs unsustainable and undermine the viability of the arts sector in particular.

In the near-term, the most important regulatory steps are two forms of transparency in how such models are developed and deployed. Development should at least disclose the origins of the data used to train them, with appropriate compensation paid; deployment should make clear the relative contribution of AI to new “works”, with a new category of computer-generated work offering a reasonable middle ground between purely human- and purely AI-generated content.

With regard to development and economic sustainability, the music industry offers interesting parallels (Huber 2023). It also went through a period of unrestrained piracy in the early digital era, which radically transformed the economics of copying and gave rise to file-sharing services such as Napster (Tan 2017). Lawsuits and legislative changes led to most media platforms adopting copyright policies and takedown protocols (Digital Millennium Copyright Act (DMCA) 1998; Seng 2014), while those like Napster were shut down completely (Menn 2003). Producers and distributors developed technical means to limit copying, but a certain amount of piracy is often priced in as the cost of doing business (Aguilar, et al. 2018; Herings, et al. 2018).

It is possible that a similar evolution will take place in AI, at least with regard to LLMs. For all the concerns that IP protection will constrain the development of new models, the market for “legitimate” models appears to be growing. Adobe, for example, has built its Firefly tools using training sets consisting only of public domain and licensed works. Shutterstock has also committed to building AI tools with a Contributor Fund to compensate artists (Hayes 2023). Other models might also be used, such as the manner in which YouTube allows certain usages of music and other copyrighted material by sharing advertising revenue with owners of the original work through its Content ID system (Edwards 2018).

On the deployment of AI models, this connects to the larger question of whether consumers should *know* whether a given work is the product of a machine or a human. That might

seem like a simple question, but AI-assisted decision-making increasingly blurs that line. For many years, certain customer relations chatbots have started on automatic for basic queries, moving through suggested responses that are vetted by a human, escalating up to direct contact with a person for unusual or more complex interactions (Kucherbaev, et al. 2018).

For the raw text and images produced by AI, at least, it should be possible to disclose their provenance. To guard against misrepresentation, various efforts are underway to detect AI-generated text through anti-plagiarism software, though these have had mixed success, at best (Barrett, et al. 2023; Morris 2023). A more difficult but effective approach would be to “watermark” text and images in a manner that is invisible to users but detectable using a key (Li, et al. 2023; Sun, et al. 2023). Given the likely spread of the underlying software, this would be practical only if it is required by law. Even then, however, the spread of deepfake porn points to the difficulty of policing any such rules.

Much of the energy in this context comes from governments around the world concerned about generative AI being used to produce ever more realistic content at ever greater scale. “Fake news” existed long before Donald Trump — it appeared in the *New York Times* at least by 1894 (“The “A.P.” News” 1894) and in a headline by 1901 (Bartlett 1901) — but AI-generated videos of Ukrainian President Volodymyr Zelenskyy “surrendering” in 2022 made clear how it might be operationalised as a weapon of war (Geng 2023, pp. 159-60). Yuval Noah Harari has gone further to argue that such usages of generative AI threaten democracy itself (Harari 2023). Hyperbole aside, greater understanding of what content is produced by AI and how it is generated would aid regulators in the world in making better decisions, rather than relying on the market and the good graces of technology companies.

4 Conclusion

T.S. Eliot once observed that “good authors borrow, great authors steal”. Occasionally, this is taken literally, a case in point being the German writer Helene Hegemann, whose 2010 best-selling novel *Axolotl Roadkill* lifted entire pages from another novel. When confronted with the apparent theft, the seventeen-year-old responded that “There’s no such thing as originality, just authenticity” (Ellis 2010).

Eliot was not, of course, condoning plagiarism. His larger point was to challenge naïve idealization of the creative process: in arts, as much as in science, each new thinker and

writer builds on the work of those who have come before. Painters inspire and echo one another; writers offer variations on plots and structures that can be mapped and catalogued (Booker 2006; Koestler 1964).

This is perhaps clearest in music, where the limits of the heptatonic scale and chord progressions mean that melodies will inevitably echo one another, as Ed Sheeran successfully argued in a case concerning similarities between his hit song “Thinking Out Loud” and Marvin Gaye’s “Let’s Get It On” (Seabrook 2023).

With regard to the first question considered in this article, it may seem pointless to argue that AI models should pay for the use of data when the entire Internet has already been absorbed (Guadamuz 2023). In addition to the market for “legitimate” models, however, there is evidence that further refinement of those models and the training of new ones depends not just on the volume of data but its quality. In particular, early suggestions that LLMs might continue improving based on synthetic data that they themselves create have foundered on projections that such AI-generated data will “poison” future models (Martínez, et al. 2023; Rao 2023). Presuming that there is an ongoing market for data and the political will to regulate it, the idea that generative AI will have its own “Napster moment” is at least plausible.

As regards AI-produced content, existing laws appear capable of holding the line on protection of the rights of human creators. Much of the regulatory attention is focused on the threats posed by the quality and scale of synthetic content and its ability to overwhelm the market by sheer volume or exert influence on populations through deception. Jurisdictions like Singapore that adopted laws intended to address misinformation and disinformation online were criticized as draconian (Jayakumar, et al. 2021), but similar tools are increasingly being considered by western liberal democracies also (Bollinger and Stone 2022; Giusti and Piras 2021).

Underpinning all of this is the question of how societies choose to regulate this sector. In theory, governments regulate activities to address market failures, or in support of social or other policies. In practice, relationships with industry and political interests may cause politicians to act — or refrain from acting — in less principled ways (Baldwin, et al. 2011, pp. 15-24). Though the troubled relationship between Big Tech and government is well documented (Alfonsi 2019; Romm 2020), this article assumes good faith on the part of regulators.

From a market perspective, failing to protect human-authored works used in the training of generative AI — or offering too much protection to the computer-generated outputs — would reduce the incentive for additional human creations. It is conceivable that this would not be a net loss if AI content more than makes up for the deficit. That would certainly be the case with regard to quantity — and may yet be so with regard to quality. (In discussions of AI content, it is common to hear sage observations that AI will never produce Michelangelo’s “David” or Jane Austen’s *Pride and Prejudice*. That may be true. Yet I will never produce such works either — nor will *you*.) Nonetheless, if the concerns about model poisoning are correct, even the AI models themselves will continue to require human creativity to achieve further improvements.

In any case, regulation is not simply about market optimization. If societies value the arts, then investments should be made in them. Again, there is precedent for this — even if it is not particularly inspiring. Photography all but killed portraiture, though painting remains a niche activity (Graw and Lajer-Burcharth 2016). Motion pictures and television did not lead to the end of live theatre, but far fewer see it today than a century ago. Such art forms, along with dance, opera, orchestral music and the like continue with government subsidies — and, even then, are often regarded as bourgeois conceits (Bennett 2016). It is possible that human-generated text and images will become similarly rarefied, preserved as a callback to a different era, like Shakespeare’s Globe Theatre.

There are far larger implications, of course, connected to how we relate to knowledge. For the past two decades, “to Google” came to reflect how many questions were formulated. The answers came with a ranked list of responses, which had the salutary consequence of making clear that there were multiple possible answers — along with subtle indications that some of them might be supported by advertisers who paid for the whole enterprise. If, as appears likely, generative AI leads to our interactions with ChatGPT, Bard, Claude, and so on becoming the first point of inquiry, it is probable that answers will be clear, succinct, and opaque. At that point, understanding the inputs that go into generative AI and who is responsible for its outputs will have political as well as economic consequences.

References

- Abbott, Ryan (2020). *The Reasonable Robot: Artificial Intelligence and the Law*. Cambridge: Cambridge University Press.
- Aguiar, Luis, Jörg Claussen, and Christian Peukert (2018). "Catch Me If You Can: Effectiveness and Consequences of Online Copyright Enforcement," *Information Systems Research* 29, 656-678.
- Alfonsi, Carlotta (2019). "Taming Tech Giants Requires Fixing the Revolving Door," *Kennedy School Review* 19, 166-170.
- Alter, Alexandra, and Elizabeth A. Harris (2023). "Franzen, Grisham and Other Prominent Authors Sue OpenAI." *New York Times*, 20 September 2023.
- Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 US ____ (Supreme Court of the United States).
- Authors Guild v. Google*, 721 F.3d 132 (2nd Cir).
- Baldwin, Robert, Martin Cave, and Martin Lodge (2011). *Understanding Regulation: Theory, Strategy, and Practice*. 2nd ed. Oxford: Oxford University Press.
- Barrett, Clark, Brad Boyd, Ellie Burzstein, Nicholas Carlini, Brad Chen, Jihye Choi, Amrita Roy Chowdhury, Mihai Christodorescu, Anupam Datta, Soheil Feizi, Kathleen Fisher, Tatsunori Hashimoto, Dan Hendrycks, Somesh Jha, Daniel Kang, Florian Kerschbaum, Eric Mitchell, John Mitchell, Zulfikar Ramzan, Khawaja Shams, Dawn Song, Ankur Taly, and Diyi Yang (2023). "Identifying and Mitigating the Security Risks of Generative AI," arXiv 2308.14840 [cs.AI].
- Bartlett, Capt. (1901). "'Fake' News for Spain." *New York Times*, 6 October 1901.
- Beebe, Barton (2008). "An Empirical Study of U.S. Copyright Fair Use Opinions, 1978-2005," *University of Pennsylvania Law Review* 156 560-.
- Bennett, James T. (2016). *Subsidizing Culture: Taxpayer Enrichment of the Creative Class*. New York: Routledge.
- Bollinger, Lee C., and Geoffrey R. Stone (eds) (2022). *Social Media, Freedom of Speech, and the Future of our Democracy*. Oxford: Oxford University Press, 2022.
- Booker, Christopher (2006). *The Seven Basic Plots*. London: Continuum.
- Brodkin, Jon. "US Judge: Art Created Solely by Artificial Intelligence Cannot Be Copyrighted." *Ars Technica*, 22 August 2023.
- Brown, Abbe, Smिता Kheria, Jane Cornwell, and Marta Iljadica (2019). *Contemporary Intellectual Property: Law and Policy*. 5th ed. Oxford: Oxford University Press.
- Burrow-Giles Lithographic Co v. Sarony*, 111 US 53 (1884).
- Chesterman, Simon (2020). "Artificial Intelligence and the Limits of Legal Personality," *International and Comparative Law Quarterly* 69, 819-844.
- Compendium of U.S. Copyright Office Practices, 3rd edition 2019. (U.S. Copyright Office, Washington, DC, 2019), available at www.copyright.gov/comp3/docs/3-15-19/compendium-draft.pdf.
- Copyright Act 1994 (NZ).
- Copyright Act 2021 (Singapore).
- Copyright Amendment Act 1994 (India).
- Copyright and Related Rights Act 2000 (Ireland).
- Copyright Ordinance 1997 (HK).
- Copyright, Designs and Patents Act 1988 (UK).
- Creamer, Ella (2023). "Amazon Restricts Authors from Self-Publishing More Than Three Books a Day After AI Concerns." *Guardian*, 20 September 2023.

- de Cock Buning, Madeleine (2018). "Artificial Intelligence and the Creative Industry: New Challenges for the EU Paradigm for Art and Technology by Autonomous Creation." In *Research Handbook on the Law of Artificial Intelligence*, eds. Woodrow Barfield and Ugo Pagallo. Cheltenham: Edward Elgar, 511-535.
- Dell'Acqua, Fabrizio, Edward McFowland III, Ethan Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani 2023. *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality* (Harvard Business School Working Paper, September 2023).
- Digital Millennium Copyright Act (DMCA) 1998 (US). 105-304.
- Drucker, P.F. (1959). *The Landmarks of Tomorrow*. New York: Harper and Row.
- du Sautoy, Marcus (2019). *The Creativity Code: Art and Innovation in the Age of AI*. Cambridge: Harvard University Press.
- Dwivedi, Yogesh K., Nir Kshetri, Laurie Hughes, Emma Louise Slade, Anand Jeyaraj, Arpan Kumar Kar, Abdullah M. Baabdullah, Alex Koohang, Vishnupriya Raghavan, Manju Ahuja, Hanaa Albanna, Mousa Ahmad Albashrawi, Adil S. Al-Busaidi, Janarthanan Balakrishnan, Yves Barlette, Sriparna Basu, Indranil Bose, Laurence Brooks, Dimitrios Buhalis, Lemuria Carter, Soumyadeb Chowdhury, Tom Crick, Scott W. Cunningham, Gareth H. Davies, Robert M. Davison, Rahul Dé, Denis Dennehy, Yanqing Duan, Rameshwar Dubey, Rohita Dwivedi, John S. Edwards, Carlos Flavián, Robin Gauld, Varun Grover, Mei-Chih Hu, Marijn Janssen, Paul Jones, Iris Junglas, Sangeeta Khorana, Sascha Kraus, Kai R. Larsen, Paul Latreille, Sven Laumer, F. Tegwen Malik, Abbas Mardani, Marcello Mariani, Sunil Mithas, Emmanuel Mogaji, Jeretta Horn Nord, Siobhan O'Connor, Fevzi Okumus, Margherita Pagani, Neeraj Pandey, Savvas Papagiannidis, Ilias O. Pappas, Nishith Pathak, Jan Pries-Heje, Ramakrishnan Raman, Nripendra P. Rana, Sven-Volker Rehm, Samuel Ribeiro-Navarrete, Alexander Richter, Frantz Rowe, Suprateek Sarker, Bernd Carsten Stahl, Manoj Kumar Tiwari, Wil van der Aalst, Viswanath Venkatesh, Giampaolo Viglia, Michael Wade, Paul Walton, Jochen Wirtz, and Ryan Wright (2023). "'So What If ChatGPT Wrote It?' Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy," *International Journal of Information Management* 71, no. 102642.
- Edwards, Benj. "Artist Receives First Known US Copyright Registration for Latent Diffusion AI Art." *Ars Technica*, 23 September 2023.
- Edwards, Dustin W. (2018). "Circulation Gatekeepers: Unbundling the Platform Politics of YouTube's Content ID," *Computers and Composition* 47, 61-74.
- Ellis, Lee. "T. S. Eliot Was Wrong." *New Yorker*, 18 February 2010.
- Ellison, Steven. "Who Owns DALL-E Images?" *FindLaw*, 29 August 2022.
- Factsheet on Copyright Act 2021 2022. (Intellectual Property Office of Singapore, 24 November 2022), available at <https://www.ipos.gov.sg/docs/default-source/resources-library/copyright/copyright-act-factsheet.pdf>.
- Geng, Yinuo (2023). "Comparing "Deepfake" Regulatory Regimes in the United States, the European Union, and China," *Georgetown Law Technology Review* 7, 157-178.
- Gervais, Daniel J. (2020). "The Machine as Author," *Iowa Law Review* 105, 2053-.
- Giusti, Serena, and Elisa Piras (eds) (2021). *Democracy and Fake News: Information Manipulation and Post-Truth Politics*. London: Routledge, 2021.
- Graw, Isabelle, and Ewa Lajer-Burcharth (eds) (2016). *Painting Beyond Itself: The Medium in the Post-Medium Condition*. London: Sternberg, 2016.
- Guadamuz, Andrés. "DALL-E Goes Commercial, But What About Copyright?" *TechnoLlama*, 25 July 2022.
- (2023). "A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs," <https://ssrn.com/abstract=4371204>.
- Harari, Yuval Noah. "AI Has Hacked the Operating System of Human Civilisation." *Economist*, 28 April 2023.

- Hayes, Carol (2023). "Generative Artificial Intelligence and Copyright: Both Sides of the Black Box," <https://ssrn.com/abstract=4517799>.
- Herings, P. Jean-Jacques, Ronald Peeters, and Michael S. Yang (2018). "Piracy on the Internet: Accommodate It or Fight It? A Dynamic Approach," *European Journal of Operational Research* 266, 328-339.
- Huber, Nick (2023). "Rapid Advances in AI Set to Upend Intellectual Property." *Financial Times*, 14 June 2023.
- Jayakumar, Shashi, Benjamin Ang, and Nur Diyanah Anwar (eds) (2021). *Disinformation and Fake News*. Singapore: Palgrave Macmillan, 2021.
- Koestler, Arthur (1964). *The Act of Creation*. London: Hutchinson.
- Kucherbaev, Pavel, Alessandro Bozzon, and Geert-Jan Houben (2018). "Human-Aided Bots," *IEEE Internet Computing* 22, no. 6, 36-43.
- Lemley, Mark A. (2023). "How Generative AI Turns Copyright Upside Down," https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4517702.
- Lemley, Mark A., and Bryan Casey (2021). "Fair Learning," *Texas Law Review* 99, no. 4.
- Li, Mingjie, Zichi Wang, and Xinpeng Zhang (2023). "An Effective Framework for Intellectual Property Protection of NLG Models," *Symmetry* 15, no. 1287.
- Lim, Jeffrey. "A.I. & Copyright – Did Singapore’s Copyright Act 2021 Solve Copyright Problems in the Training of A.I.?" *Legal500*, 25 August 2023.
- Maguire, Laurie (2020). *The Rhetoric of the Page*. Oxford: Oxford University Press.
- Martínez, Gonzalo, Lauren Watson, Pedro Reviriego, José Alberto Hernández, Marc Juárez, and Rik Sarkar (2023). "Towards Understanding the Interplay of Generative Artificial Intelligence and the Internet," arXiv 2306.06130.
- Menéndez, Javier. "AI-Generated Artwork Is Blowing Up The Economics of Art." *Medium*, 4 February 2023.
- Menn, Joseph (2003). *All the Rage: The Rise and Fall of Shawn Fanning's Napster*. New York: Crown.
- Mims, Christopher (2023). "AI Tech Enables Industrial-Scale Intellectual-Property Theft, Say Critics." *Wall Street Journal*, 4 February 2023.
- Morris, Meredith Ringel (2023). "Scientists' Perspectives on the Potential for Generative AI in their Fields," arXiv 2304.01420 [cs.CV].
- Nordemann, Axel (1999). "Germany." In *Copyright and Photographs: An International Survey*, eds. Ysolde Gendreau, Axel Nordemann and Rainer Oesch. The Hague: Kluwer, 135-.
- Nova Productions v. Mazooma Games* [2007] EWCA Civ [2007] 219.
- O’Leary, Daniel E. (2013). "Artificial Intelligence and Big Data," *IEEE Intelligent Systems* 28, 96-99.
- "The "A.P." News." (1894). *New York Times*, 10 July 1894.
- Padmanabhan, Arjun, and Tanner Wadsworth (forthcoming). "A Common Law Theory of Ownership for AI-Created Properties," *Journal of the Patent and Trademark Office Society*, <https://ssrn.com/abstract=4411194>.
- Phelan, Ryan N., and Matthew Carey. "ChatGPT and Intellectual Property (IP) Related Topics." *IP Litigator*, May/June 2023, 8.
- Prassl, Jeremias (2018). *Humans as a Service: The Promise and Perils of Work in the Gig Economy*. Oxford: Oxford University Press.
- Rao, Rahul. "AI-Generated Data Can Poison Future AI Models." *Scientific American*, 28 July 2023.
- Reisner, Alex. "Revealed: The Authors Whose Pirated Books Are Powering Generative AI." *The Atlantic*, 19 August 2023a.
- . "These 183,000 Books Are Fueling the Biggest Fight in Publishing and Tech." *Atlanci*, 25 September 2023b.

- Research, Goldman Sachs 2023. Generative AI Could Raise Global GDP by 7% (Goldman Sachs, 5 April 2023), available at <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>.
- Rethinking Database Rights and Data Ownership in an AI World 2020. (Singapore Academy of Law: Law Reform Committee, July 2020).
- Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence 2020. (World Intellectual Property Organisation, Geneva, 21 May 2020), available at www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_2_ge_20/wipo_ip_ai_2_ge_20_1_rev.pdf.
- Roberts, Huw, Josh Cowls, Jessica Morley, Mariarosaria Taddeo, Vincent Wang, and Luciano Floridi (2021). "The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation," *AI & Society* 36, 59-77.
- Romm, Tony (2020). "Tech Giants Led by Amazon, Facebook, and Google Spent Nearly Half a Billion on Lobbying over the Past Decade, New Data Shows." *Washington Post*, 22 January 2020.
- Samuelson, Pamela (1986). "Allocating Ownership Rights in Computer-Generated Works," *University of Pittsburgh Law Review* 47, 1185-.
- Schade, Michael 2023. Will OpenAI Claim Copyright over What Outputs I Generate with the API? (OpenAI, 2023), available at <https://help.openai.com/en/articles/5008634-will-openai-claim-copyright-over-what-outputs-i-generate-with-the-api>.
- Schoppert, Peter. "The Books Used to Train LLMs." *AI and Copyright (Substack)*, 11 March 2023.
- Seabrook, John. "The Case for and Against Ed Sheeran." *New Yorker*, 5 June 2023.
- Second Request for Reconsideration for Refusal to Register Théâtre D'opéra Spatial (SR # 1-11743923581; Correspondence ID: 1-5T5320R) 2023. (United States Copyright Office: Copyright Review Board, 5 September 2023), available at <https://fingfx.thomsonreuters.com/gfx/legaldocs/byprqkqxe/AI%20COPYRIGHT%20REGISTRATION%20decision.pdf>.
- Séjourné, Stéphane 2020. Draft Report on Intellectual Property Rights for the Development of Artificial Intelligence Technologies (European Parliament, Committee on Legal Affairs, Brussels, 24 April 2020), available at www.europarl.europa.eu/doceo/document/JURI-PR-650527_EN.pdf.
- Seng, Daniel (2014). "The State of the Discordant Union: An Empirical Analysis of DMCA Takedown Notices," *Virginia Journal of Law & Technology* 18, 369-473.
- Silberling, Amanda. "Science Fiction Publishers Are Being Flooded with AI-Generated Stories." *TechCrunch*, 22 February 2023.
- Sun, Yuchen, Tianpeng Liu, Panhe Hu, Qing Liao, Shaojing Fu, Nenghai Yu, Deke Guo, Yongxiang Liu, and Li Liu (2023). "Deep Intellectual Property Protection: A Survey," arXiv 2304.14613v2.
- Tan, David (2017). "Fair Use and Transformative Play in the Digital Age." In *Research Handbook on Intellectual Property in Media and Entertainment*, eds. Megan Richardson and Sam Ricketson. Cheltenham: Edward Elgar, 102-131.
- . "The Best Things in Life are Not for Free: Copyright and Generative AI Learning." *Singapore Law Gazette*, April 2023.
- (2024a). "Generative AI and Copyright: Infringement and Fair Use (Part 1)," *SAL Practitioner* [2024], ___-.
- (2024b). "Generative AI and Copyright: Infringement and Fair Use (Part 2)," *SAL Practitioner* [2024], ___-.
- Tan, David, and Wee Liang Tan (2022). "AI, Author, Amanuensis," *Journal of Intellectual Property Studies* 5, 1-32.
- Torrance, Andrew W., and Bill Tomlinson (forthcoming). "Training Is Everything: Artificial Intelligence, Copyright, and Fair Training," *Dickinson Law Review* <https://ssrn.com/abstract=4437680>.
- Zikopoulos, Paul, Dirk deRoos, Krishnan Parasuraman, Thomas Deutsch, James Giles, and David Corrigan (2012). *Harness the Power of Big Data: The IBM Big Data Platform*. New York: McGraw Hill.

