

# THE AI-COPYRIGHT TRAP

CARYS J. CRAIG\*

[*Forthcoming*, 100 CHICAGO-KENT LAW REVIEW (2025), *AI Disrupting Law Symposium*]

## ABSTRACT

*As AI tools proliferate, policy makers are increasingly being called upon to protect creators and the cultural industries from the extractive, exploitative, and even existential threats posed by generative AI. In their haste to act, however, they risk running headlong into the Copyright Trap: the mistaken conviction that copyright law is the best tool to support human creators and culture in our new technological reality (when in fact it is likely to do more harm than good). It is a trap in the sense that it may satisfy the wants of a small group of powerful stakeholders, but it will harm the interests of the more vulnerable actors who are, perhaps, most drawn to it. Once entered, it will also prove practically impossible to escape. I identify three routes in to the copyright trap in current AI debates: first is the “if value, then (property) right” fallacy; second is the idea that unauthorized copying is inherently wrongful; and third is the resurrection of the starving artist trope to justify copyright’s expansion. Ultimately, this article urges AI critics to sidestep the copyright trap, resisting the lure of its proprietary logic in favor of more appropriate routes towards addressing the risks and harms of generative AI.*

## INTRODUCTION

Artificial Intelligence (AI) is dramatically reshaping our cultural landscape. Today’s AI is generating outputs—from written text to graphic art and videos, and from music to software code—that have all the external hallmarks of human creative expression. Facially, these outputs are indistinguishable from works of human authorship; culturally and legally, however, they are something else entirely.

---

\* LL.B (Hons), LL.M., S.J.D, Professor of Law and Associate Dean (Research & Institutional Relations), Osgoode Hall Law School, York University. My thanks to Edward Lee and the editors for the opportunity to contribute to this Special Issue. Thanks also to the participants of the Symposium and, in particular, to Mark Lemley and Christa Laser for their comments. For helpful discussions as the paper took shape, I am grateful to Alina Trapova, Luke McDonagh, Patrick Goold, Tobias Schonwetter, Robert Brauneis, Meredith Jacob, Michael Carroll, and Bitu Amani, amongst many others. Thanks also to Tianchu Gao (Osgoode, JD Class of 2024) for her excellent research assistance. All mistakes are my own.

Novel as this may seem, modern AI technology is thereby raising age-old questions about the socio-economic structures of cultural production, demanding that we interrogate anew the appropriate role of copyright law in incentivizing creative activity and mediating the impacts of technological change on artists, authors, and the markets for their works. As AI tools proliferate and pervade every corner of our culture, policy makers are being called upon with increasing urgency to protect creators and the cultural industries from the extractive, exploitative, and even existential threats posed by generative AI. In the haste to act, however, and amidst the steady onslaught of AI hype and moral panic, there is a real risk that they will run headlong into the copyright trap.

This essay will explain the nature of the copyright trap as I see it, with a view to cautioning AI critics against the pragmatic deployment of copyright concepts and narratives to counteract the rapid and under-regulated acceleration of AI. It begins in Part 1 with some discussion of the challenge posed by AI to copyright law and policy, and, in particular, the growing controversy around generative AI systems. I then turn, in Part 2, to identify what lures us into the copyright trap: the perceived promise that copyright law can readily rein in AI and thereby protect human creators and the market for human-authored works in our new technological reality. It is a trap in the sense that, if it works, it will satisfy the wants of a small group of powerful stakeholders who therefore seek to set it in place and strive to maintain it—but also because, in doing so, it is poised to harm the interests of the more vulnerable actors who are, perhaps, most naturally drawn to it. And finally, it is a trap because, once entered, it will prove practically impossible to escape.

This Part goes on to identify and explore three routes by which we risk running ourselves in the copyright trap in current AI debates: first is the tenacious commitment to the idea that a private, proprietary right must attach to every culturally or socially valuable text (the “if value, then right” fallacy); second is the insistence that unauthorized copying is an inherent wrong, and so the individual “copy” should be the relevant object of analysis and control even in the digital context; and third, is the widely-held belief that the allocation of private copyright control holds the answer to creators’ economic struggles, empowering them to secure fair returns and future livelihoods. Each of these routes is already well trodden in the historical trajectory of copyright law and policy debates, but I will revisit them briefly here, if only to signpost the points at which they may once again lead us astray.

There are many good reasons to be concerned about the rise of generative AI and the threats that it presents to creators, our culture, and the public interest. There are also good reasons, however, to be wary of running into a copyright trap as we set about rewriting laws that will shape and steer technological change for the foreseeable future. The essay concludes by urging AI’s most important and perceptive critics to sidestep the copyright trap—resisting its lure in favour of more apt and direct routes towards reducing the risks and harms that they rightly perceive.

## 1. THE AI-COPYRIGHT CHALLENGE

“The rise of the machines is here,” it has been said, “but they do not come as conquerors, they come as *creators*.”<sup>1</sup> The examples are by now well known. Trained on vast quantities of text data scraped from the Internet, Chat-GPT spits out prose, in response to basic prompts, that readily fulfils the demands of most writing tasks, adopting styles and parroting common parlance in a way that can fool the most discerning reader into believing the words were uttered by another person, not constructed by algorithmic code. Trained on a massive dataset of images and textual descriptions, Midjourney can do the same for visual arts, turning basic textual prompts into fascinating graphic images faster than any artist’s paintbrush could move across canvas. Music generating tools like Mubert AI and Google’s MusicLM can compose songs and musical loops from basic text inputs, while OpenAI’s recently released Sora app turns text prompts into strikingly photorealistic videos in the blink of an eye. IBM’s Watsonx Code Assistant is essentially code trained on code to generate new code. Once the distant imaginings of philosophers and science fiction writers, then, authored artifacts have become capable of generating new artifacts that appear, at least, to be authored. And so, with that, we have entered an age of mass synthetic creativity.<sup>2</sup>

It is hardly surprising that the law of copyright—which allocates exclusive rights over original works of literature and software, music, visual arts, drama and film, with the goal of encouraging their creation and dissemination—has been launched into a fresh existential crisis the likes of which we haven’t seen since the arrival of the World Wide Web. Indeed, some now predict that AI will be the death of copyright; others imagine that copyright will spell the death of AI.<sup>3</sup> Both things could be true, though it seems more likely that neither will come to pass. In the meantime, however, the damage that could be wrought by each upon the other should not be underestimated.

Much of the initial literature on generative AI asked whether its increasing sophistication should cause us to rethink embedded assumptions about the meaning of authorship. In an earlier work, Ian Kerr and I argued that the very notion of AI-authorship rests on a category mistake—an error about the *ontology* of authorship.<sup>4</sup> We cautioned

---

<sup>1</sup> Andres Guadamuz, *Artificial Intelligence and Copyright* 5 WIPO Magazine 14 (2017) at 17, [https://www.wipo.int/wipo\\_magazine/en/2017/05/article\\_0003.html](https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html) [emphasis added].

<sup>2</sup> Dan Burk, *Cheap Creativity and What it Will Do*, 57 Georgia L. Rev. 1669, 1679 et seq (2023).

<sup>3</sup> See e.g. Louis Menand, *Is A.I. the Death of I.P.?*, The New Yorker (15 January 2024), <https://www.newyorker.com/magazine/2024/01/22/who-owns-this-sentence-a-history-of-copyrights-and-wrongs-david-bellos-alexandre-montagu-book-review>. Cf Pamela Samuelson, *Generative AI Meets Copyright* (CITRIS and the Banatao Institute, 26 April 2023), <https://www.youtube.com/watch?v=6sDGIrVO6mo> (arguing that “Copyright law is the only law that’s already in existence that could bring generative AI systems to their knees.... Copyright law is an existential threat to progress in this field.”)

<sup>4</sup> Carys Craig & Ian Kerr, *The Death of the AI Author*, 52 Ottawa L. Rev 31 (2020). See also, e.g., Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 47 U. PITT. L. REV. 1185 (1985), James Grimmelman, *There’s no such thing as a computer-authored work—And it’s a good thing, too*, 39 Columbia J. Law & Arts 403 (2016); Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, 34

against the kind of romanticized conceptions of both authorship and “artificial intelligence” that would lead us to mistake the computational and predicative processes of large language models with the relational and communicative act of authorship, properly understood as a social and dialogic practice (and therefore an essentially human endeavour). It follows, we argued, that AI-generated outputs are *de facto* unauthored and, in the absence of a statutorily enacted legal fiction, uncopyrightable *per se*.

In a later work, I advanced the same understanding of authorship as a fundamentally communicative act to argue that training AI is not a copyright infringing activity; just as the AI is categorically incapable of authoring original works of expression, it is incapable of receiving, reading, or enjoying them as such.<sup>5</sup> The machine is not an audience or addressee of any meaning or message communicated by a work, and so, I argued, its use of a work cannot implicate the copyright interests that flow from authorship.<sup>6</sup> In the training process, a work of authorship within a dataset is simply translated into statistics—its meaning turned into math. The tokenized information extracted from the work is not the expressive element of the work over which a copyright owner may lay claim. I have also insisted that, even if it this process were to *prima facie* infringe the copyright owner’s right, it would typically constitute fair use or, in Canada, fair dealing for the purpose of research or private study.<sup>7</sup> Either approach leads to the conclusion that training AI on copyright-protected works is a non-infringing act for which the copyright owner’s consent is not required—and this is, I believe, an essential starting point for any appropriate regulatory response to AI technologies and the risks they pose to our socio-cultural ecosystem.

---

BERKELEY TECH. L.J. 343 (2019); Dan L. Burk, *Thirty-Six Views of Copyright Authorship*, by Jackson Pollock, 58 HOUS. L. REV. (2020).

<sup>5</sup> Carys Craig, *The AI-Copyright Challenge: Tech-Neutrality, Authorship, and the Public Interest in Ryan Abbott* (ed.) RESEARCH HANDBOOK ON INTELLECTUAL PROPERTY AND ARTIFICIAL INTELLIGENCE (Edward Elgar Press, 2022). Cp.

<sup>6</sup> Cp. Oren Bracha, *The Work of Copyright in the Age of Machine Production* (September 24, 2023). U OF TEXAS LAW, LEGAL STUDIES RESEARCH PAPER, <https://ssrn.com/abstract=4581738> or <http://dx.doi.org/10.2139/ssrn.4581738>.; Maurizio Borghi and Stavroula Karapapa, *Non-display Uses of Copyright Works: Google Books and Beyond*, QUEEN MARY JOURNAL OF INTELLECTUAL PROPERTY 1 (2011), 21 (44-45). But see James Grimmelman, *Copyright for Literate Robots*, 101 IOWA LAW REVIEW (2016).

<sup>7</sup> Carys Craig, *AI and Copyright* in Florian Martin-Bariteau & Teresa Scassa (eds), ARTIFICIAL INTELLIGENCE AND THE LAW IN CANADA (LexisNexis, 2021); Sean M. Fiil-Flynn, et al, *Legal reform to enhance global text and data mining research: Outdated copyright laws around the world hinder research*, 378(6623) SCIENCE 951-53 (2 December 2022); Carys Craig et al, *Joint Submission of IP Scholars, Consultation on a Modern Copyright Framework for Artificial Intelligence* Sept 26, 2021), <<https://digitalcommons.osgoode.yorku.ca/cgi/viewcontent.cgi?article=1226&context=reports>>. For the argument that non-expressive uses for the purpose of text and data mining typically constitute lawful fair use, see e.g., Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J. COPYRIGHT Soc’y U.S.A. 291 (2019); Michael W Carroll, *Copyright and the Progress of Science: Why Text and Data Mining is Lawful*, 53 UC DAVIS L. REV. 893 (2019); Pamela Samuelson, *Text and Data Mining of In-Copyright Works: Is It Legal?* 64:11 COMMUNICATIONS OF THE ACM 20 (2021); Mark A. Lemley and Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743 (2021); Matthew Sag, *Copyright Safety for Generative AI*, HOUS. L. REV. (forthcoming).

In what follows, I mean to mount a principled defence of these claims in combination: *Copyright law should neither incentivize and reward the use of generative AI nor obstruct its training and development.* Rather than directly grappling once again with the doctrinal questions of originality and authorship or infringement and fair dealing, however, I will advance a more general argument against copyright overreach: copyright law is neither apposite nor equipped to govern the way that generative AI is developed, trained, deployed, or enjoyed. Insisting that it should do so, and imagining that it is up to the task, could do far more harm than good.

Unfortunately, however, it seems clear that copyright law (or a contorted version thereof) is increasingly being invoked as a regulatory response to the harms of AI.

### A. *The Copyright Policymaking Context: The Canadian Example*

National policy discourse around rapidly evolving AI was initially less concerned with the doctrinal niceties of copyright law and its application to generated outputs and more concerned with cultivating domestic AI industries to support the innovation economy, positioning national economies at the forefront of this next technical revolution. In 2017, Canada became the first country launch a national AI strategy, for example, investing significant funds and resources into the development, commercialization, and adoption of AI technology.<sup>8</sup> As such, when the Committee charged with conducting Canada's 2019 *Copyright Act Review* turned its mind to AI, its concern was with "help[ing] Canada's promising future in artificial intelligence become reality."<sup>9</sup> The Committee had been warned that copyright law had the potential to become an obstacle to AI's development. As lawyer Maya Madeiros explained in her testimony before the Committee:

AI learns to think by reading, listening and viewing data, which can include copyrighted works such as images, video, text and other data....

The training process can involve reproductions of the training data, and these can be temporary reproductions to extract features of the data that can be discarded after the training process. ... It is unclear whether the use of copyrighted works for training an AI system is considered copyright infringement if the ... copyright

---

<sup>8</sup> See Ana Brandusescu, *Artificial Intelligence policy and funding in Canada: Public Investments, Private Interests* CENTRE FOR INTERDISCIPLINARY RESEARCH ON MONTREAL, MCGILL UNIVERSITY (March 2021), [https://www.mcgill.ca/centre-montreal/files/centre-montreal/aipolicyandfunding\\_report\\_updated\\_mar5.pdf](https://www.mcgill.ca/centre-montreal/files/centre-montreal/aipolicyandfunding_report_updated_mar5.pdf). See also CIFAR, "CIFAR Pan-Canadian Artificial Intelligence Strategy", <https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy>. For an overview of Canada's multi-stage AI strategy see *The State of Artificial Intelligence Research in Canada*, HILLNOTES (March 8, 2023), <https://hillnotes.ca/2023/03/08/the-state-of-artificial-intelligence-research-in-canada/#:~:text=The%20Pan%2DCanadian%20Artificial%20Intelligence,developing%20a%20Canadian%20AI%20community>.

<sup>9</sup> Statutory Review of the Copyright Act: Report of the Standing Committee on Industry, Science and Technology, 87 (June 2019, 42 Parl. 1<sup>st</sup> Sess.), <<https://www.ourcommons.ca/Content/Committee/421/INDU/Reports/RP10537003/indurp16/indurp16-e.pdf>>.

owner's permission is not obtained. This uncertainty exists even if the initial training is done for research purposes....This uncertainty can limit the data that is used by AI innovators to train the AI system. The quality of the dataset will impact the quality of the resulting trained algorithm.<sup>10</sup>

The problem posed by copyright law to the training of AI systems and the quality of their outputs is by now well known. High profile litigation and media-covered controversies have drawn wide-spread attention to the essential role that pre-existing works play in the process of “teaching” an AI to “think.”<sup>11</sup> What is interesting to note, however, is the ease with which the 2019 Committee was persuaded that AI had an inconvenient copyright problem—one that could and should be statutorily solved. Its report simply recommended that “the Government of Canada introduce legislation to amend the *Copyright Act* to facilitate the use of a work or other subject-matter for the purpose of informational analysis.”<sup>12</sup>

Also striking in retrospect is that there were apparently “very few witnesses who argued against an exception for informational analysis.” Only one witness argued that such uses should be licensed by collective societies, advancing what has now become a common refrain: There “is no justification for turning authors and publishers into unpaid suppliers to technology developers—in other words, for requiring a sector that operates on very thin margins to subsidize a sector that can well afford to pay a fair price to its suppliers.”<sup>13</sup>

The Committee's recommendation to add an exception for “informational analysis” (now more commonly referred to as “text and data mining”) was not acted upon. Instead, two years later, the Canadian Government launched a public consultation specifically seeking input on the matter of copyright and AI.<sup>14</sup> The submissions it received addressing the use of copyright-protected works in AI training were later summarized in the following terms:

Stakeholders from the technology sector, scholars, and user groups generally argued for an exception making clear that the use of works in text and data mining activities (TDM) does not require additional authorization from rightsholders. Creative industries were of the view that a new exception is not desirable, as it

---

<sup>10</sup> *Id.* at 86.

<sup>11</sup> For a running list of copyright infringement lawsuits filed against AI companies, see <https://chatgptiseatingtheworld.com/2023/12/27/master-list-of-lawsuits-v-ai-chatgpt-openai-microsoft-meta-midjourney-other-ai-cos/>.

<sup>12</sup> *Id.* at 87 (introducing Recommendation 23).

<sup>13</sup> *Id.* at note 270, quoting Brush Education, *Brief Submitted to INDU* (5 September 2018), <https://www.ourcommons.ca/Content/Committee/421/INDU/Brief/BR10008276/br-external/BrushEducationInc-e.pdf>.

<sup>14</sup> *Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things* (July 16, 2021), <https://ised-isde.canada.ca/site/strategic-policy-sector/en/marketplace-framework-policy/copyright-policy/consultation-modern-copyright-framework-artificial-intelligence-and-internet-things>.

would preclude rights holders from receiving fair compensation for the use of their works in TDM activities.<sup>15</sup>

What seemed to have been a relatively uncontroversial proposition in 2018-19—that copyright ought not to obstruct the development of AI by preventing the lawful and optimal training of AI systems—had, by 2021, become the controversial assertion of a select group of stakeholders and experts; one that was uniformly opposed by stakeholders in the creative industries, who were suddenly and overwhelmingly concerned with ensuring “fair compensation” for rightsholders.

The political ground continued to tilt. Two years later, presumably unsatisfied with the previous responses, Canada’s Government launched yet another public consultation, now specifically on *Copyright in the Age of Generative AI*.<sup>16</sup> While acknowledging that the actual copyright issues remained unchanged, this consultation paper stressed that Canadians should review these considerations in light of their intervening “experience with generative AI” and expressly “acknowledge[d] that some stakeholders have raised concerns about the impacts of AI on creators and creative industries going beyond copyright *per se*.”<sup>17</sup> It explained that the Government’s aim was to balance two objectives: supporting innovation and investment in AI and emerging technologies, on one hand, and, on the other, “support[ing] Canada’s creative industries and preserv[ing] the incentive to create and invest provided by the rights set out in the...Copyright Act, including to be adequately remunerated for the use of their works.”<sup>18</sup>

I would be remiss not to point out that Canada’s *Copyright Act*<sup>19</sup> contains no such general right for authors—never mind the creative industries—to be “adequately remunerated.”<sup>20</sup> Copyright simply establishes limited exclusive rights to perform

---

<sup>15</sup> Consultation Paper: Copyright in the Age of Generative AI, 5-6 (December 2023), <https://ised-isde.canada.ca/site/strategic-policy-sector/sites/default/files/documents/2023-12/2023-consultation-paper-en.pdf>. My responses to the Canadian Consultation Paper are contained in two submissions: S. Flynn, L. Guibault, C. Handke, J. Vallbé, M. Palmedo, C. Craig, M. Geist & J.P. Quintais, *Submission to Canadian Government Consultation on a Modern Copyright Framework for AI and the Internet of Things* (Sept. 17, 2021), <https://ssrn.com/abstract=3952238>; and Carys Craig, Bitu Amani, Sara Bannerman, Céline Castets-Renard, Pascale Chapdelaine, Lucie Guibault, Gregory Hagen, Cameron Hutchison, Ariel Katz, Alexandra Mogyoros, Graham Reynolds, Anthony D. Rosborough, Teresa Scassa & Myra Tawfik, *Submission by IP Scholars Copyright and Artificial Intelligence*, [https://www.uwindsor.ca/law/sites/uwindsor.ca.law/files/final\\_ai\\_submission\\_canadian\\_ip\\_scholars.pdf](https://www.uwindsor.ca/law/sites/uwindsor.ca.law/files/final_ai_submission_canadian_ip_scholars.pdf).

<sup>16</sup> <https://ised-isde.canada.ca/site/strategic-policy-sector/en/marketplace-framework-policy/consultation-copyright-age-generative-artificial-intelligence>.

<sup>17</sup> Consultation Paper: Copyright in the Age of Generative AI, 6 (December 2023), <https://ised-isde.canada.ca/site/strategic-policy-sector/sites/default/files/documents/2023-12/2023-consultation-paper-en.pdf>

<sup>18</sup> *Id.* at 7.

<sup>19</sup> *Copyright Act*, R.S.C. 1985, c. C-42, s. 29. See Craig et al., *Submission by IP Scholars Copyright and Artificial Intelligence*,

<sup>20</sup> A right to equitable remuneration exists only for performers and sound recording makers (who are not authors) in respect of the publication of sound recordings: *Copyright Act*, s. 19. Under section 32.3, this right does not apply in respect of an act that does not infringe copyright. The only other right to remuneration arises under the limited blank audio recording media levy under section 81.

specified actions in relation to protected works, which owners may exchange for whatever value the market happens to ascribe to them. More importantly, however, the rights and interests in the balance, as articulated in the consultation paper, include only *industries'* interests in incentives to innovate, create and invest, and *rightsholders'* supposed rights to adequate remuneration. Nowhere in this so-called balance was any mention made of the public side of the copyright balance—the public's interest in the dissemination of works, for example, or users' rights to make fair and lawful uses of protected works, or the importance of the public domain (in which facts and information—i.e. *data*—reside). The articulation of the government's overarching policy approach thus overstated the scope of copyright holders' entitlements, focused on industries' interests rather than the interests of individuals or the public, and overlooked the users' rights that are central to the copyright balance as repeatedly confirmed by Canada's Supreme Court.<sup>21</sup>

Canada, often torn between its US neighbors and its European colonial roots, is always an interesting jurisdiction to watch when it comes to copyright policy reform. Cross-border trade relations (and economic pragmatism) often encourage or require coherence with US laws, while the historical and jurisprudential through-lines lead back to the United Kingdom. The influence of Quebec and French language requirements in Ottawa, however, also seems to feed affinities with policy approaches in France and continental Europe. For now, while the US position on the copyright legalities of TDM remains to be resolved through litigation focused on transformative fair use principles (which may not overlap neatly with Canada's somewhat more restrictive fair dealing doctrine), all signs point to the rising influence of the European approach in Canada.

In Europe, where copyright tends to be less explicitly utilitarian in its focus and more concerned with protecting owners and *le droit d'auteur*, the contestation over copyright and AI is seemingly settled (on paper at least) by Articles 3 and 4 of the *Digital Single Market Directive*.<sup>22</sup> Article 3 creates an exception for TDM for scientific research available only to research organizations and cultural heritage institutions, while Article 4 requires member states to provide a generally available exception for “reproductions and extractions of lawfully accessible works...for the purposes of text and data mining,” but allows owners to “opt out” by expressly reserving their rights in an “appropriate manner.” Importantly, by adopting these specific exceptions, the EU legislature has confirmed that TDM and (by implication) AI training processes are copyright-relevant activities.<sup>23</sup> And, as is typically the case in European copyright laws, limits and exceptions to the copyright owners' rights control are specific and narrowly drawn.

---

<sup>21</sup> See e.g., *Théberge v. Galerie d'Art du Petit Champlain Inc.*, 2002 S.C.C. 34 at ¶ 11–12, 30. See also *CCH v. Law Society of Upper Can.*, 2004 S.C.C. 13 at ¶ 23.

<sup>22</sup> 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, Official Journal of the European Communities 2019 L 130, 92

<sup>23</sup> Martin Senftleben, *Generative AI and Author Remuneration*, 54 *IIC* 1535, 1542 (2023), <https://doi.org/10.1007/s40319-023-01399-4>.



According to the so-called Brussels Effect, market forces alone are often sufficient to ensure that European Union rules govern the global operations of multilateral corporations and ultimately set norm for the international stage.<sup>24</sup> Certainly, the normative baseline seems to have shifted, in Canada’s copyright policy discourse, to assume, as a starting point, that authors and owners have a copyright entitlement to control the use of their works for AI training, such that the policy problem becomes how to operationalize or enforce that right and, almost immediately, how to monetize it. Canada’s latest consultation questionnaire did not ask, for example, whether remuneration would be appropriate for rightsholders whose works were used in the training of AI; rather it asked, “what level of remuneration would be appropriate for the use of a given work in TDM activities?”. It did not ask whether it should include an exception to permit TDM activities, rather it asked, “If the Government were to amend the Act to clarify the scope of permissible TDM activities, what should be its scope and safeguards?”<sup>25</sup>

It remains to be seen what regulatory response (if any) this consultation will entail, but it seems likely that copyright and its mechanisms of control will soon be called upon to play a larger role in restraining the freedom of AI developers to train their models on copyright materials in Canada—and that this will be hailed by some interested stakeholders as a victory for Canada’s creative industries and the creators they (purport to) represent.<sup>26</sup>

This short detour into the recent state of Canadian copyright policymaking sketches a trajectory in the transnational policy discourse around generative AI that is, I believe, representative of a larger narrative arc in the AI-copyright story. Whereas the initial excitement about the affordances of this emerging technology led policymakers and the public to marvel over the potential significance of the things it could produce, mounting concern over the consequences of those outputs for the health of our cultural industries has redirected attention to the developers of AI tools and the informational inputs on which they depend. More specifically, the growing fear that AI’s mass synthetic creations will effectively substitute in the market for human-authored works—reducing the demand for (and so the economic value of) works of artists and creative professionals, thereby threatening their livelihoods and the viability of the creative industries—makes tempting the idea that this automated competition could be constrained or cut off by copyright law at the source. For those who would like to see generative AI halted or at least retarded in its tracks, copyright offers a ready obstacle to throw in its path. And, at the international level, for nations whose initial AI industry-leading ambitions have

---

<sup>24</sup> Anu Bradford, *The Brussels Effect; How the European Union Rules the World* (Oxford UP, 2020).

<sup>25</sup> Consultation Paper: Copyright in the Age of Generative AI, 6 (December 2023), section 2.1.3, <https://ised-isde.canada.ca/site/strategic-policy-sector/en/marketplace-framework-policy/consultation-paper-consultation-copyright-age-generative-artificial-intelligence#s213>.

<sup>26</sup> The submissions are available online, <https://ised-isde.canada.ca/site/strategic-policy-sector/en/marketplace-framework-policy/copyright-policy/submissions-consultation-copyright-age-generative-artificial-intelligence>.

faltered in the face of global competition, economic pragmatism may now point in the direction of protecting domestic cultural industries against foreign tech industry giants (or at least extracting some toll from the latter).<sup>27</sup>

Certainly, it seems, the general willingness to grant the developers and owners of these technological systems free use of vast swathes of our human-created culture has diminished—and with it, apparently, the political will to enact exceptions to allow for AI systems to be freely trained on copyright works. Indeed, the impetus is now decidedly in the opposite direction—towards expanding and fortifying the boundaries of copyright’s exclusive domain.

### *B. Invocations of Copyright Control in the AI Political Discourse*

At the highest level, it might be observed, enthusiasm for AI and its affordances has clearly soured somewhat. Even AI’s greatest shills, the original hawkers of AI hype, have performed an incredible turnabout to caution that AI is growing too powerful, that it presents potentially existential threats for humankind and society as we know it, and they have therefore famously called for greater regulation and a pause in AI research and development.<sup>28</sup> AI critics have suggested, in turn, that such performative protestations were clearly calculated to achieve strategic goals: reinforcing the mythic notion of AI as capable of agency, consciousness, and independent action and thereby feeding into the AI hype; distracting us from the very real and current harms being perpetrated by today’s more mundane but still socially malignant AI systems; and soliciting regulations that are less likely to curtail AI development than they are to concretize the advantages of current leaders in the field.<sup>29</sup> These critical observations are important for our purposes because they reveal a pattern that has yet to be adequately acknowledged by critics in the copyright realm: Notwithstanding appearances, calling for the intervention of law (including copyright law) and the imposition of legal limits may be more self-serving and -preserving than it is self-restraining or -sacrificial. It follows that, when AI critics call for such regulations, they may inadvertently be playing into the hands of the powerful actors they mean to rein in.

The strategic sounding of such alarm bells works, however, because it resonates with wider and well-placed fears about the societal costs of AI and its unbridled arrival into every aspect of our lives. Concerns about black-box algorithmic decision-making by the state and employers, for example, combine with economic anxieties around

---

<sup>27</sup> My thanks to Luke McDonagh for proposing this explanation of shifting domestic AI-copyright policies over time.

<sup>28</sup> *Pause Giant AI Experiments: An Open Letter* (March 22, 2023), <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. See also Future of Life Institute, *Policymaking in the Pause: What can Policymakers do now to Combat Risks from Advanced AI Systems* (April 12, 2023), <https://futureoflife.org/document/policymaking-in-the-pause/>.

<sup>29</sup> See e.g. *Statement from the listed authors of Stochastic Parrots on the ‘AI pause’ letter* (March 31, 2023), [https://www.dair-institute.org/blog/letter-statement-March2023/?utm\\_source=pocket\\_saves](https://www.dair-institute.org/blog/letter-statement-March2023/?utm_source=pocket_saves).

automation and the displacement of skilled human labour; these are compounded by concerns about environmental costs and the ecological impact of generative AI, which merge, in turn, with fears about deep-fakes and AI-augmented disinformation; these only add to the general perception of the unprecedented risks that generative AI poses to everything from local journalism to classroom education, and from the viability of our professional organizations and cultural institutions to the sustainability of human creativity at large. There are, in other words, plenty of reasons to worry!

But when it comes to the AI-copyright challenge, what worries me is that the positions I previously propounded—that copyright law is not implicated in and should not obstruct AI training or development—have become increasingly controversial, against this backdrop, amongst both copyright commentators and AI critics (often those whose political positions are otherwise most aligned with my own).<sup>30</sup> As the backlash against AI and its tech giants has grown, it threatens to sweep away claims about user rights and the appropriate limits of copyright as naïve or misguided at best—as ultimately serving the interests of big tech and corporate gain while undermining the rights of (human) artists and content creators. The so-called copyleft movement has typically enticed the anti-capitalists (the leftists, the humanists, the feminists) with its call to resist corporate control and ownership over knowledge and the cultural domain; but now, many such critical voices bundle *disrespect for copyright* into the harms done and risks posed by generative AI.<sup>31</sup> And so, predictably, renewed respect for copyright—its expansion and heightened enforcement—emerges as one potential solution to the AI problem.

By way of example, critiques of AI as a fundamentally “extractive” technology, compellingly advanced by Kate Crawford, point to the way in which AI is premised on extraction—the extraction of natural resources, of human labor, and of data.<sup>32</sup> This important critical insight encompasses the way that data is gathered and processed in the training of AI systems—indeed, the metaphor of data-*mining* neatly captures the analogy drawn by Crawford and others between the multiplicity of extractive practices, both

---

<sup>30</sup> For recent examples of copyright scholars calling for recognition of authors and/or copyright owners right to control the use of their works in AI training data, see e.g. Benjamin L.W. Sobel, Artificial Intelligence's Fair Use Crisis, 41 COLUM. J.L. & ARTS 45, 93 (2017); Martin Senftleben, *Generative AI and Author Remuneration*, IIC 54, 1535–1560 (2023), <https://doi.org/10.1007/s40319-023-01399-4>; Christophe Geiger & Vincenzo Iaia, *The Forgotten Creator: Towards a Statutory Remuneration Right for Machine Learning of Generative AI*, 52 COMPUTER LAW & SECURITY REVIEW 105925 (2024); Frank Pasquale & Haochen Sun, *Consent and Compensation: Resolving Generative AI's Copyright Crisis*, 110 U. VA. L. REV. ONLINE (forthcoming, 2024), <https://papers.ssrn.com/abstract=4826695>; Robert Brauneis *Copyright And The Training Of Human Authors And Generative Machines* (on file with the author).

<sup>31</sup> See e.g. Frank Pasquale and Haochen Sun, *Consent and Compensation: Resolving Generative AI's Copyright Crisis*, 110 U. VA. L. REV. ONLINE, 32 (forthcoming, 2024), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4826695](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4826695) (“The narrative of the copyleft, which argues that big content owners exploit users, must be supplemented by another story: big technology firms exploiting labor without adequate (and, often, any) compensation. A #CreateDontScrape movement has capitalized on this sentiment, adopting the copyleft's rhetoric of distributional justice and democratization toward a very different end”).

<sup>32</sup> Kate Crawford, *ATLAS OF AI* (Yale UP, 2021).

physical and intangible, on which AI depends. The critical concept of extraction can therefore extend to describe the way in which valuable human intellectual labour expended in the creation of original works is caught up and notionally refined into mere “data” for training purposes. As Crawford explains, the rhetorical transformation of human activities into “data” (“a bloodless word”) conveniently recasts them as natural resources (the “new oil”!) and implies (within a colonial frame<sup>33</sup>) that all this data is simply there for the taking.<sup>34</sup> As Crawford sees it,

The practices of data extraction and training dataset construction are premised on a commercialized capture of what was previously part of the commons.... an extraction of knowledge-value from public goods. ...[W]e see a handful of privately owned companies that now have enormous power to extract insights and profits from those sources. The new AI gold rush consists of enclosing different fields of human knowing, feeling, and action—every time of available data—all caught up in an expansionist logic of never-ending collection....a pillaging of public space.<sup>35</sup>

This critique of AI’s extractive logic might seem to offer a compelling basis on which to call for a limit—or even an end—to the use and datafication of copyrightable works of human expression in service of tech companies’ profits. Indeed, in a recent article with Jason Schulz, Crawford and Schulz observe:

Copyright law was developed by eighteenth-century capitalists to intertwine art with commerce. In the twenty-first century, it is being used by technology companies to allow them to exploit all the works of human creativity that are digitized and online.<sup>36</sup>

Whereas copyright control served the capitalists of earlier times, the argument goes, today’s capitalists rely on copyright’s legally prescribed *limits* (such as fair use). If we want to put an end to corporate exploitation and the commercialization of human endeavour for industrial gain, it might seem to follow that we need to change copyright law – or, more specifically, that we need to change or limit fair use. They go on to ask, “What expectations of consent, credit, or compensation should human creators have going forward, when their online work is routinely incorporated into training sets?” The suggestion is that we should focus on equity and creativity, thinking more deeply about “who has a say in how creations can be used, and who should profit.”<sup>37</sup> Now these are important considerations to be sure; but my concern is that copyright law (with new or strengthened private power to control use of works for data analysis) seems to be the

---

<sup>33</sup> Luke Stark and Anna Lauren Hoffman, *Data is the New What?*, 4(1) *Journal of Cultural Analytics* (2019), <https://doi.org/10.22148/16.036>, cited by Crawford, *ATLAS OF AI* at 113.

<sup>34</sup> Kate Crawford, *ATLAS OF AI*, 113.

<sup>35</sup> Kate Crawford, *ATLAS OF AI*, 120-21.

<sup>36</sup> Kate Crawford and Jason Schultz. “Generative AI Is a Crisis for Copyright Law.” *Issues in Science and Technology* (): 79–80. <https://doi.org/10.58875/GUYG6120>.

<sup>37</sup> Kate Crawford and Jason Schultz. “Generative AI Is a Crisis for Copyright Law.” *Issues in Science and Technology* (): 79–80. <https://doi.org/10.58875/GUYG6120>

imagined route into shifting these norms. Copyright's essential boundaries, carefully erected over time to safeguard fair uses and the public domain from encroaching private control, have become contested territory once again.

The so called “3Cs” of “Consent, Credit and Compensation” are getting a lot of air time these days. They were a frequent refrain in the recent US copyright hearings on AI.<sup>38</sup> Certainly, they have a nice ring to them and not only thanks to the alliteration; they resonate individually, but compound in combination, because each concept has earned its obvious pedigree in the political discourses of feminism, anticapitalism, equality, and labour rights movements. In combination, the 3Cs’ Rule itself was developed by the Cultural Intellectual Property Rights Initiative, inspired by *The Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization*, to guide best practices for the use and distribution of traditional knowledge (TK) and cultural expressions (TCE) in the context of the fashion and textile industry.<sup>39</sup> It might therefore seem like a compelling way to articulate the rightful demands of weaker parties in response to widespread cultural misappropriation. But surely, and especially given this original motivation and the settler colonial context that necessitated the development of this 3Cs rule for best practices, the use of the same phrase to capture the presumed entitlement of *every* author, creator, or copyright holder should raise some qualms (if not alarm).

It is one thing to insist upon the free, prior, and informed consent of crafts persons, Indigenous peoples, or local communities to the use of their TCE. It is quite another to suggest that every individual author should have the right to grant—or withhold—consent to the use of their work as a source of data. It is one thing to insist on the acknowledgment of the source community and inspiration for TK and TCE, but quite another to demand that the name of every copyright owner of every text in a dataset of billions of texts is collected and acknowledged in connection with every AI model and its outputs. And, of course, it is quite a different thing to insist upon compensating Indigenous communities and benefit-sharing, in recognition of their contribution to the survival and transmission of traditional knowledge and cultural expressions over generations, than it is to demand individual rightsholder remuneration through collective societies for every digital text or image scraped into a training dataset.

It might be argued (though I am not sure how convincingly) that the appropriation or repurposing of the 3Cs in the context of AI is nonetheless apt given the power imbalances at play or the inherent unfairness at issue in the battle between everyday

---

<sup>38</sup> See *Copyright and Artificial Intelligence*, <https://copyright.gov/ai/>. See also James Love, *We Need Smart Intellectual Property Laws for Artificial Intelligence* (August 7, 2023), <https://www.scientificamerican.com/article/we-need-smart-intellectual-property-laws-for-artificial-intelligence/>.

<sup>39</sup> See <https://www.culturalintellectualproperty.com/mission#:~:text=The%203Cs%20stand%20for%3A%20Consent,monetary%20or%20non%2Dmonetar>

creators and powerful AI-industry actors. Frank Pasquale (another leading critic of the AI and tech industries)<sup>40</sup> and IP scholar Haochen Sun observe:

For many turn-of-the-millennium advocates of an open Internet, copyright was a menace, constantly threatening to stifle innovation. By contrast, many artists and activists now see it as one of the few tools left to demand accountability from an extraordinarily concentrated and powerful technology industry.<sup>41</sup>

Taking “the principles of consent and compensation” as their “loadstar,” Pasquale and Sun therefore propose a legislative scheme combining control (opt-outs) and remuneration for copyright owners to fix the “broken social contract between technologists and creatives.”<sup>42</sup> Feminist technology law scholar Amanda Levendowski also invokes the 3Cs when proposing that we might use copyright in this context as a “a really robust hammer to hit some nails...for justice-oriented solutions.”<sup>43</sup> Copyright may not be the best tool to address the injustices wrought by AI, I take her to mean, but it is both powerful and available.<sup>44</sup> More specifically, her suggestion is that the fair use inquiry should consider not only the public benefit served by a technology, which has historically created room for fair use to protect new and productive technological tools,<sup>45</sup> but must also “grapple with the countervailing public harms.” With that in mind, she asks:

[I]s there a place in copyright law to force courts...to come to a fairer resolution than fair use necessarily requires? [One] that may...put some credit, put some compensation, put some consent back where authors and artists and journalists want it, not necessarily where copyright requires it.<sup>46</sup>

In this way, the pursuit of the 3Cs is intended to push back at power, employing the blunt tool of copyright control, but reaching beyond what copyright actually requires by narrowing the scope of what fair use permits (in the name of greater fairness).

---

<sup>40</sup> See Frank Pasquale, *THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION* (Harvard University Press, 2016).

<sup>41</sup> Frank Pasquale and Haochen Sun, *Consent and Compensation: Resolving Generative AI's Copyright Crisis*, 110 U. VA. L. REV. ONLINE, 32 (forthcoming, 2024), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4826695](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4826695).

<sup>42</sup> *Id.* 50-51. But note that, for Pasquale and Sun, the proposal is not simply strategic or political; rather, it seems rooted in the normative endorsement of an author's claim to control their work as a matter of right.

<sup>43</sup> See *Generative AI and Copyright Conference, Panel 2: The Future of Generative AI and Copyright Policy*, Colorado Law School (Oct 9, 2023), <https://www.youtube.com/watch?v=c6vmVuCwZtA> (at 44:38-45:26)

<sup>44</sup> See e.g., Amanda Levendowski, *Using Copyright to Combat Revenge Porn*, 3 NYU J. INTELL. PROP. & ENT. L. 422 (2013-2014) (espousing a similar strategy in a different context).

<sup>45</sup> The reference is presumably to cases such as *Google LLC v. Oracle America, Inc.*, 593 U.S. \_\_\_\_ (2021) or *Authors Guild v. Google* 804 F.3d 202 (2d Cir. 2015) wherein the public benefits of the technology weighed in favour of fair use with respect to the transformative purpose and character of the use.

<sup>46</sup> See *Generative AI and Copyright Conference, Panel 2: The Future of Generative AI and Copyright Policy*, Colorado Law School (Oct 9, 2023), <https://www.youtube.com/watch?v=c6vmVuCwZtA> (at 46:24-46:45).

Even if one were to endorse the harnessing of expanded copyright control and redeployment of the 3Cs rule, however, their realization would have entirely different and unprecedented practical ramifications in this AI context. Moreover, as James Love explains, “[e]ach C has its own practical challenges that run counter to the most favorable text and data mining exceptions.”<sup>47</sup>

An obvious one relates to sheer scale. Even the first release of Stable Diffusion required a data set of 2.3 billion images.<sup>48</sup> Anyone who has tried to identify the owner of even a single work and secure rights to its use knows that finding rightsholders and obtaining consent for every text in an LLM’s dataset would be a practical impossibility. The 3Cs mean that creators must be also given credit, however, which means identifying not only the owners of the copyright but the creator (author) of every work. The proposal is so implausible that the only reasonable assumption is that it is designed to obstruct—the whole point being to prevent the practice from taking place, or at least to significantly slow down the development of these generative AI technologies.<sup>49</sup>

But even if getting consent and giving credit were achievable, what would it achieve? To have one’s name included in a database of billions of names distantly associated with and obliquely credited for an indiscernible, unknowable fraction of a fraction of the generated outputs of a particular AI tool? And if compensation were then to be required (presumably by payment to the copyright owner but perhaps also to the creator, as the political discourse seems to assume), it is hard to imagine a transfer of any significant value to individuals whose works happen to be amongst the billions of works in a dataset from which an (unprotected) output is generated. Undertaking transparent and appropriate distribution of what would presumably be millions of *micro*-payments by collective societies (who are keen to get in on the act for obvious reasons) would presumably produce burdensome and enormously disproportionate transactions costs. Ultimately, the individual compensation proposal based on the inclusion of copyright-protected works in AI training data seems almost as pointless as it does improbable.<sup>50</sup>

---

<sup>47</sup> James Love, *We Need Smart Intellectual Property Laws for Artificial Intelligence*, SCIENTIFIC AMERICAN (August 7, 2023), <https://www.scientificamerican.com/article/we-need-smart-intellectual-property-laws-for-artificial-intelligence/>

<sup>48</sup> *Id.*

<sup>49</sup> Cf. Frank Pasquale et al, *Generating AI: A Historical, Cultural, and Political Analysis of Generative AI: Submission to the Office of the Secretary-General*, <https://dataethics.eu/wp-content/uploads/2023/09/Generating-AI.pdf> (“There is a chance that the growing corporate battles over copyright laws might create barriers that will slow down these technologies. Generative AI uses copyrighted material with no credit, no compensation, and no control for the people who created that material. We should encourage government to rewrite the rules of corporate and intellectual property for Generative AI development.”) See also Frank Pasquale and Haochen Sun, *Consent and Compensation: Resolving Generative AI’s Copyright Crisis*, 110 U. VA. L. REV. ONLINE, 45-50 (forthcoming, 2024), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4826695](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4826695).

<sup>50</sup> Cf. Martin Senftleben, *Generative AI and Author Remuneration*, 54 IIC 1535, 1542 (2023), <https://doi.org/10.1007/s40319-023-01399-4>, 1546-49 (acknowledging the practical and legal obstacles to implementing individual rightsholders’ remuneration protocols in respect of AI-training data “in the real world”).

In a recent podcast, Timnit Gebru, a computer scientist and expert on AI ethics, dismissed the “it’s too big to do that” argument against curating and documenting datasets, arguing “where else can you do that? Can you sell me a food item at a restaurant and be like: Eat this food, I don’t know what it’s made of, there’s some sugar, I know that there’s some flour, whatever else you’re on your own.”<sup>51</sup> The point is quite compelling. There are, after all, many fields of activity upon which we impose burdensome disclosure requirements for the protection of individuals and the public. (The response to objections might reasonably be: if it’s not worth it, don’t do it.) But the analogy looks more tenuous when we recall that such disclosures (of ingredients, to take Gebru’s example) are required to protect the users or consumers of the final product—not the economic or intellectual property rights of the original suppliers. There may be good reason to impose burdensome transparency and disclosure obligations on AI developers, but author attribution is not one of them. Demanding disclosure of a massive database of authors’ or copyright owners’ names, even if it were possible, would not directly address any of the public harms posed by generative AI (and may exacerbate many of them, especially when one recalls that any and all original texts, images, photographs, videos, transcripts, scraped from the Web are authored works for copyright purposes).<sup>52</sup> Indeed, requiring mere credit for inclusion a mass of inputs bears no real relation to the problem at all.

Gebru’s is an extremely important critical voice in the AI discourse; her work has been absolutely vital to revealing and explaining the inequities, abuses, harms, and biases produced by AI systems and obscured by AI hype.<sup>53</sup> In a recent piece on *AI Art and its Impact on Artists*, Gebru and her co-authors convincingly argued, first, that “image generators are not artists.”<sup>54</sup> Describing as misguided and harmful the idea that these artifacts have agency, that they’re somehow “inspired” by the data they are trained on, the authors write:

Art is a form of communication: it communicates.

In contrast, the outputs of artifacts like image generators are not framed for enjoyment...the image generator has no understanding of the perspective of the audience or the experience that the output is intended to communicate to [the] audience.

---

<sup>51</sup> Paris Marx with Timnit Gebru, *AI Hype Distracted Us From Real Problems*, TECH WONT SAVE US, <https://podcasts.apple.com/ca/podcast/ai-hype-distracted-us-from-real-problems-w-timnit-gebru/id1507621076?i=1000642067936>.

<sup>52</sup> Consider, e.g., the privacy harms that such disclosures could present, recalling that the ‘works’ included in datasets are not simply novels and painting, for example, but posts on Reddit and photos from Flickr, whose “authors” are just people going about their everyday onlife activities.

<sup>53</sup> Timnit Gebru, *Race and Gender* in Markus D. Dubber, Frank Pasquale, Sunit Das (eds.) THE OXFORD HANDBOOK OF ETHICS OF AI, 251-69 (2020), [doi:10.1093/oxfordhb/9780190067397.013.16](https://doi.org/10.1093/oxfordhb/9780190067397.013.16)

<sup>54</sup> Harry H. Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru, *AI Art and its Impact on Artists* (2023) in AAAI/ACM CONFERENCE ON AI, ETHICS, AND SOCIETY (August 08-10, 2023), <https://doi.org/10.1145/3600211.3604681>.



I could not agree more. Indeed, this is precisely the point that Ian Kerr and I set out to make when calling for the death of the mythic “AI author.”<sup>55</sup> But while I argue that the training of AI is, by the same token, not use of a work as a communicative expression, Gebru and her coauthors gesture to the harms done to working artists—the “increasing barriers and job losses for creatives because of image generators.” Pulling equity into frame, the authors point to disproportionate harms for the most marginalized artists, as well as the “reputational damage and trauma” that can flow from the use of artists’ works without their consent, and the reinforcement of biases and stereotyping more broadly.

The piece is a powerful indictment of insidious encroachment of AI generated art. As it notes, “while some of the harms discussed...overlap with the rights protected by US copyright law, others [do] not.”<sup>56</sup> Observing that “image generators represent a dynamic between artists and large-scale companies appropriating their work that has previously not been examined in US copyright law,” the authors conclude that there are “gaps in the law that do not take into account the social and economic harm to artists.”<sup>57</sup> Amongst these, the ways in which image generators “facilitate[] informational harms” or “normalize appropriation of art at the input stages” are identified as problems “beyond the scope of fair use considerations.”<sup>58</sup> How art generators might be prompted to mimic an artist’s “style,” for example, is also beyond the carefully circumscribed scope of the derivative work right or moral rights’ protections. The authors ultimately “advocate for regulation that prevents organizations from using people’s content to train image generators without their consent.”<sup>59</sup>

Once again, then, the proposed solution is effectively to override the crucial limits of copyright law—limits that have traditionally restrained corporate power to protect the public interest—in order to respond to the harms being endured by artists facing the chilling impacts of generative AI.

Gebru’s work, like that of Crawford and other critical scholars working in this space—Emily Bender, Alex Hanna, Shannon Vallor—calls for an approach to the regulation of AI technologies informed by feminist insights, one that recognizes “how technology interacts with power.”<sup>60</sup> Again, I wholeheartedly agree. But to this I want to add, still from critical feminist perspective, that *we must also be cognisant of how copyright interacts with power*. If we want to safeguard the human creative endeavour, to advance equality and inclusion, to strive for representation and participation in the cultural sphere, to create regulatory structures that support the responsible and ethical development of AI, then strengthening copyright law or supplementing it with copyright-*plus* requirements is, I believe, simply the wrong way to go. In doing so, we risk losing sight of what matters here and why, instead reinscribing the logics of exclusivity, private

---

<sup>55</sup> Carys Craig & Ian Kerr, *The Death of the AI Author*, 52 *Ottawa L. Rev.* 31 (2020).

<sup>56</sup> Jiang et al, *AI Art and its Impact on Artists*, 369.

<sup>57</sup> *Id.*

<sup>58</sup> *Id.* at 369-70.

<sup>59</sup> *Id.* at 371.

<sup>60</sup> *Id.* at 372.

property, and corporate control over information and cultural expression. The responsibilities and ethical imperatives that must be invoked to inform AI policy should not borrow from the proprietary logic of individual entitlement and alienability on which copyright depends.

In the fight against the social, cultural, and economic harms of generative AI, I am sorry to say that copyright is entering the fray as a *false friend*. At best, the perceived copyright problem underlying the inequities of AI is a diversion—a red herring. Worse, I fear, invoking copyright as a counterweight to corporate power in the AI industry is going to be counterproductive in advancing the critics’ important political agenda.

## II. THE COPYRIGHT TRAP

I began by cautioning AI critics against the copyright trap: the mistake of assuming that that copyright law is the best tool with to support creators and cultural production (when in fact it is likely to do more harm than good). While Part I.A. identified a new preoccupation with copyright control in the Canadian AI policy discourse, Part I.B. identified the same tendency towards the reinvigoration of copyright norms (and the copyright-adjacent norms of consent, credit, and compensation) in the writings of a few of the most insightful critical scholars working in the field today. The reality, however, is that the public discourse around generative AI is absolutely replete with such allusions to authorial entitlement, demands that copyright be respected, and accusations of widespread infringement in need of an immediate judicial remedy or legislative response. If the empowerment and reward of human creators is the bait in the copyright trap (as Gebru notes, “artists are galvanizing”),<sup>61</sup> then it is an easy trap to set and one whose effectiveness has been demonstrated time and again throughout copyright’s history—form the great Literary Property Debates at the turn of the 18<sup>th</sup> century to the copyright wars of the Napster era at the turn of the 21<sup>st</sup>.

I also suggested at the outset that there are a few well-worn paths into this old copyright trap, and it is to these that I now turn.

### A. *If Value Then (Property) Right*

One way into the copyright trap is to assume that everything that has value must be privately owned. The notion that intellectual property rights should attach to any intangible thing of value—“if value/then right”—is a common but unfortunate fallacy.<sup>62</sup>

---

<sup>61</sup> Paris Marx with Timnit Gebru, *AI Hype Distracted Us From Real Problems*, TECH WONT SAVE US, <https://podcasts.apple.com/ca/podcast/ai-hype-distracted-us-from-real-problems-w-timnit-gebru/id1507621076?i=1000642067936>.

<sup>62</sup> See Alfred C. Yen, *Brief Thoughts About If Value/Then Right*, 99 BOS. U. L. REV. 2479, 2480 (“That principle, which the U.S. Copyright Act does not embrace, expresses the intuition that “wherever value is received, a legal duty to pay arises, regardless of whether imposing that legal duty serves public welfare”, citing Wendy Gordon). See also Rochelle Cooper Dreyfuss, *Expressive Genericity: Trademarks as*

As scholars have warned repeatedly over the years, it is a damaging default to assume that intellectual property simply expands to enclose the latest valuable innovation.<sup>63</sup> This entails the unnecessary swelling of our IP system and the continual encroachment of IP claims into the public domain. In the context of generative AI, the fallacy can produce the misguided assumption that AI-generated works should be copyrightable unless we can wholly deny their value, economic or otherwise. More pertinent for our purposes here, however, it can also support the assumption that anyone who *reaps value* from a work without permission has *taken* something to which they necessarily have no right. Closely related to the idea that value entails property is the similarly mistaken conviction that freeriding on another's efforts is tantamount to theft.<sup>64</sup> As Wendy Gordon has explained, while there is

an obvious moral attractiveness to the idea that it is unjust for an entity to reap where it has not sown...when taken literally, as a standalone prohibition on free riding, the...claim is drastically overbroad. A culture could not exist if all free riding were prohibited within it.<sup>65</sup>

The core claim of the artistic community's anti-AI activism is that AI image-generation involves "a kind of *theft*"—that they "steal artists' creative labour."<sup>66</sup> Similarly, in the controversy around the Books3 dataset and Chat-GPT, authors complained of "the biggest act of copyright theft in history."<sup>67</sup> As one commentator put it, "never before have so few been in a position to steal so much from so many."<sup>68</sup> The language of theft, of course, implies that property has been stolen; and with this, we run headlong into centuries of philosophical, legal, and metaphysical debates about the nature of both

---

*Language in the Pepsi Generation*, 65 NOTRE DAME L REV 397, 405–06 (1990) (questioning the idea that relationship between value and ownership justifies granting trademark rights); Wendy J. Gordon, *On Owning Information: Intellectual Property and the Restitutionary Impulse*, 78 VA. L. REV. 149, 178–80, 244 (1992). Perhaps the most powerful critique of this tautological reasoning is still that of Felix Cohen, *Transcendental Nonsense and the Functional Approach*, 35 COLUM. L. REV. 809, 815 (1935).

<sup>63</sup> See, e.g., David Lange, *Recognizing the Public Domain*, 44 L. & CONTEMP. PROBS. 147 (1981); James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 L. & CONTEMP. PROBS. 33 (2003); James Boyle, *THE PUBLIC DOMAIN* (2008). See generally Pamela Samuelson, *Enriching Discourse on Public Domains*, 55 DUKE L.J. 783 (2006); Carys Craig, *The Canadian Public Domain: What, Where and to What End?*, 7 CAN. J. LAW & TECH. 221 (2010).

<sup>64</sup> See Mark A. Lemley, *Property, intellectual property, and free riding*, 83 TEX L. REV. 1031 (2004). [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=582602](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=582602) (disputing the appropriateness of a "property-free riding paradigm").

<sup>65</sup> Wendy J. Gordon, *On Owning Information: Intellectual Property and the Restitutionary Impulse*, 78 VA. L. REV. 149 (1992). 167.

<sup>66</sup> See Trystan S. Goetze, *AI Art is Theft: Labour, Extraction, and Exploitation, Or, On the Dangers of Stochastic Pollocks*, arXiv preprint <https://arxiv.org/abs/2401.06178>.

<sup>67</sup> Kelly Burke, 'Biggest act of copyright theft in history': thousands of Australian books allegedly used to train AI model, GUARDIAN, (Sept 28, 2023) <https://www.theguardian.com/australia-news/2023/sep/28/australian-books-training-ai-books3-stolen-pirated>.

<sup>68</sup> James B. Meigs, *Is Ai Just Theft Under Another Name?*, *Science, Nature & Technology* (March 2024), <https://www.commentary.org/articles/james-meigs/ai-openai-chatgpt-copyright-theft/>.

copyright's subject matter and the entitlement that copyright secures.<sup>69</sup> A dominant thread in these debates is a Lockean one, whether implied or explicit, that pulls together the assumption that authors have a natural property right over the fruits of their intellectual labour with the idea that AI companies are benefitting from authors' works, reaping what they have not sown, and thereby misappropriating ("stealing") the authors' property.

I have written at length elsewhere, as has Gordon and many others, about the limits and risks of a Lockean theory of copyright law.<sup>70</sup> Not least amongst these is its tendency to turn limited copyright claims into absolute proprietary dominion over works, privileging the individual's right to exclude over the interests of users and the public domain. That is precisely what we see when accusations of theft are wielded in a way that disregards the limits of copyright owner's claims over the information contained in their works, for example, or their artistic "style" (neither of which copyright protects, for good reason). We also see this kind of property absolutism in the effective denial of fair dealing or fair use—as though the copyright owner's entitlement naturally extends beyond what fair use permits (such that even lawful fair uses are recast as trespass—an unlawful encroachment onto the copyright owner's claimed property).

Another concern with a property-rights approach is its reliance on the physical analogue in the intangible realm. Appropriation of a physical thing is very different from the unauthorized use of an intangible artifact: being non-rivalrous, informational works can be infinitely reproduced and reused without depriving anyone of the enjoyment of the work or the information it contains. It follows that foreclosing their use by others comes at a cost.<sup>71</sup> Authors whose works are included in datasets are not thereby denied enjoyment or use of their works. Including an author's book in training data is nothing like "taking the milk out of [their] cereal bowl."<sup>72</sup> The idea that their works have been "stolen" from them misrepresents the nature of the harms that flow from the use of their work in AI-training processes while disregarding—or rendering presumptively unfair, unjust, or unethical—any potential benefits that such uses might afford.<sup>73</sup>

---

<sup>69</sup> See also Paris Marx with Timnit Gebru, *AI Hype Distracted Us From Real Problems*, TECH WONT SAVE US, <https://podcasts.apple.com/ca/podcast/ai-hype-distracted-us-from-real-problems-w-timnit-gebru/id1507621076?i=1000642067936> (Gebru stating, "Their entire business model is predicated on theft and not compensating people for stuff.")

<sup>70</sup> Carys Craig, *Locke, Labour, and Limiting the Author's Right: A Warning Against a Lockean Approach to Copyright Law*, 28 QUEEN'S LAW JOURNAL 1 (2002).

<sup>71</sup> Mark A. Lemley, *Property, intellectual property, and free riding*, 83 TEX L. REV. 1031 (2004), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=582602](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=582602)

<sup>72</sup> See Valeire Ouellet, Sylvene Gilchrist, Shaki Sutharsan, *CBC News Analysis finds thousands of Canadian authors, books, in controversial dataset used to train AI* (Dec 7, 2023), <https://www.cbc.ca/news/canada/canadian-authors-books3-ai-dataset-1.7050243> (quoting Canadian author Hayden Taylor and his concerns "about copyright violations of his work": "In the last 35 years that I've been a writer, almost all of my income has been derived from royalties, It's literally taking the milk out of my cereal bowl.")

<sup>73</sup> The controversial Books3 data set of "stolen" books, for example, was created by Shawn Presser with the stated objective of opening up data to allow for the development of quality AI tools by more than just a handful of wealthy corporations. To recast this access-oriented goal as mere theft is to miss the context for

Consider, for example, that the use of copyright materials is vital to training AI models that are not generating new works as such, as well as other kinds of data processing and informational analysis the purpose or effect of which is not to produce potentially competing outputs but to identify patterns, predict probable outcomes, or diagnose problems. As James Love explains, we have to be careful if “the rules that apply to recorded music or art also carry over to the scientific papers and data used for medical research and development.”<sup>74</sup> (They do.) An approach to generative AI that relies, for its regulatory force, upon the presumed rights of copyright owners to prohibit the use of their works as sources of information and objects of analysis, creates a private realm of exclusion far greater than any copyright claim should afford. In doing so, it encloses elements of the public domain and forecloses an unknowable number of potentially beneficial uses.

Relatedly, the use of terms like “datamining” and “extraction” to describe informational analysis in the AI policy discourse are problematic for the same reason: they rely upon a physical analogy that implies rivalry between competing uses/users and, in turn, a kind of taking away, depriving, removing, and degrading of the original thing, space, or place. Moreover, in post-colonial contexts, especially in the Global South, these associations are particularly potent.<sup>75</sup> Such rhetorical portrayals of the data analysis or machine learning processes vital for training AI clearly come loaded with political weight. Notwithstanding their apparent anti-corporate credentials, at least when used to call out the exploitative activities of big tech, such terminology also feeds a kind of copyright-expansionist logic that ultimately serves corporate interests under the guise of seeking fairness. Meaningful resistance to corporate control over information cannot invoke, as its premise, data ownership as a matter of proprietary entitlement. After all, what is owned can be alienated—it can and will be bought and sold, licensed or assigned, by those with the resources to lay claim to it and the power to transact.

Regarded from within the property-freeriding paradigm, it may reasonably appear as though “ethical and responsible” AI development requires respect for the rightsholders’ exclusive enjoyment of their valuable data as intellectual property. But if we approach the question instead with our eye on the public’s interest and protection of the public domain from private appropriation and corporate control, our regulatory focus may settle on ensuring that the training materials necessary for the development of ethical and responsible AI systems are *open and available* for all to use.

---

and implications of the act. See Kate Knibbs, *The Battle Over Books3 Could Change AI Forever*, WIRED (Sept 4, 2023), <https://www.wired.com/story/battle-over-books3/>.

<sup>74</sup> James Love, James Love, *We Need Smart Intellectual Property Laws for Artificial Intelligence*, SCIENTIFIC AMERICAN (August 7, 2023), <https://www.scientificamerican.com/article/we-need-smart-intellectual-property-laws-for-artificial-intelligence/>.

<sup>75</sup> See generally Caroline Ncube, Desmond Oriakhogba, Issaac Rutenberg, & Tobias Schonwetter, ARTIFICIAL INTELLIGENCE AND THE LAW IN AFRICA [Pin?] (LexisNexis 2024).

### *B. Copying is Wrong*

Another common route into the copyright trap is the assumption that copying is inherently wrongful, with the consequence that any unauthorized copying is regarded as unethical and condemned as unlawful. The first objection to this assumption must be that there is nothing inherently wrong with copying. Indeed, as Abraham Drassinower has explained, “the assumption that copying is wrongful is a radically mistaken way to approach copyright law.”<sup>76</sup> Copyright makes the reproduction of a work an exclusive right of the author or owner, to be sure, but the exclusive right to copy extends only to that which is original expression within the work and that which is not otherwise a fair or permitted use. The doctrinal contours of copyright are drawn to ensure that anyone is free to copy another’s “ideas” so long as the expression of those ideas is not substantially reproduced. Notably, an author’s “style” is within this realm of abstract ideas that are free to be copied, along with genres, systems, methods, facts and information, and common stock elements. These are not unfortunate restrictions on copyright’s reach but, as decades of jurisprudence has confirmed, essential limits that ensure the sustainability of the copyright system and prevent it from defeating its own ends.<sup>77</sup> Indeed, the boundaries of copyright recognize and confirm the essential nature and desirability of copying in furtherance of learning, downstream creativity, and progress of the arts and sciences.

A second objection might seem more technical in nature. The reality is that copyright’s focus on copying is an ill fit for the digital age in which copying is easy and virtually costless, and almost every consumptive activity in relation to a work involves at least background digital copying. This fundamentally changes the significance of copying in the modern cultural marketplace. Many commentators have suggested that copyright’s policy focus should therefore shift away from the mere technicality of making copies to the more dynamic realities of creative expressive, distribution, and access. This approach underlies Matthew Sag’s argument, for example, that copies made for “non-expressive” or “non-consumptive” uses purposes should not infringe copyright;<sup>78</sup> Cheryl Foong adds to this “non-disseminative” purposes, emphasizing the significance of encouraging dissemination as an animating purpose of the copyright system.<sup>79</sup> Alan Strowel argues, as have I, that copying should infringe copyright only if one is using the work “as a work,” with a communicative dimension. In the context of AI, a misplaced fixation on copying directs our attention to digital copies in training data sets that are never enjoyed or consumed by public audiences or human recipients; the digital copying of content is a necessary but merely incidental precursor to the machine learning activity itself. Such

---

<sup>76</sup> Abraham Drassinower, *WHAT’S WRONG WITH COPYING?*, 2 (Harvard University Press, 2015)

<sup>77</sup> See Jessica Litman, *The Public Domain*, 39 EMORY L. J. 965, 967 (1990).

<sup>78</sup> Matthew Sag, ‘*Copyright and Copy-Reliant Technology*’ (2009) 103 NORTHWESTERN UNIVERSITY LAW REVIEW 1607; Matthew Sag, ‘The New Legal Landscape for Text Mining and Machine Learning’ (2019) 66 *Journal of the Copyright Society of the USA* 291.

<sup>79</sup> See Cheryl Foong, *Immaterial Copying in the Age of Access*, (2022) 44(9) EUROPEAN INTELLECTUAL PROPERTY REVIEW; see also Bingbin Lu, ‘*Reconstructing Copyright from ‘Copy-Centric’ to ‘Dissemination-Centric’ in the Digital Age*’ (2013) 39(4) JOURNAL OF INFORMATION SCIENCE 479.



copies are immaterial both literally and figuratively—their technical existence as “copies” ought not to be what shapes and confines the future development of AI technologies.

In Canada, unfortunately, the Supreme Court can be taken to have ruled that every more-than-transitory copy counts as a copy for infringement purposes, even if only incidental to another permitted purpose.<sup>80</sup> I have argued elsewhere that such an approach contradicts a principle of substantive technological neutrality, upsetting the copyright balance in the digital age.<sup>81</sup> For the majority of the Court, however, each and every digital copy implicates copyright, while concerns with technological neutrality enter the picture only at the point of attributing economic value to the copies (for the purposes of tariff-setting).<sup>82</sup> With respect, the absurdity of such an approach is nicely captured in this observation by Aaron Perzanowku and Jason Schultz:

Although copies have existed since the origins of copyright law, we are shifting quickly into a post-copy world, one where digital works exist as data flows and rarely reside in a material object for more than a transitory period of time, where copies blink into and out of existence on a nearly constant basis. In such a world, expecting consumers, rightsholders, or regulators to keep tabs on individual copies is as useful as demanding that fish track the movements of particular drops of water.<sup>83</sup>

Perhaps, as Michael Madison has suggested,<sup>84</sup> we need to stop thinking of the vast volume of texts, images, and expressive works that we have amassed in this information era as massive accumulations or storehouses of individuated artifacts and begin to think of them instead as an ocean of knowledge. Perhaps we could then turn to the question of governing this ocean as a knowledge commons—a vast expanse of shared intellectual and cultural resources in whose vitality and sustainability we all have an interest; but within which none of us can stake an individual claim *qua* right to exclude. In this picture, we can no more demand the exclusive right to control the use of individual digital copies in technological processes that we can extract our tomato juice from the sea.<sup>85</sup>

---

<sup>80</sup> Can. Broad. Corp. (CBC) v. SODRAC 2003, Inc., 2015 S.C.C. 57.

<sup>81</sup> Carys Craig, *The AI-Copyright Challenge: Tech-Neutrality, Authorship, and the Public Interest in Ryan Abbott* (ed.) RESEARCH HANDBOOK ON INTELLECTUAL PROPERTY AND ARTIFICIAL INTELLIGENCE (Edward Elgar Press, 2022). See also Can. Broad. Corp. (CBC) v. SODRAC 2003, Inc., 2015 S.C.C. 57, para. 117, 164 (Abella J. (diss): “Each broadcast-incidental copy is not a separate reproduction of the work under the *Act* simply because the technical imperatives of effecting a broadcast require the presence of multiple copies. They do not, as a result, attract separate royalties. To conclude otherwise is to doom both technological neutrality and the ability of copyright law to preserve the delicate balance between the rights of copyright holders and the public’s interest in the dissemination of creative works”).

<sup>82</sup> Can. Broad. Corp. (CBC) v. SODRAC 2003, Inc., 2015 S.C.C. 57, para. 55.

<sup>83</sup> Aaron Perzanowski and Jason Schultz, *Legislating Digital Exhaustion* (2014) 29 BERKELEY TECHNOLOGY LAW JOURNAL 1535, describing a ‘post-copy world’.

<sup>84</sup> [In conversation with the author.]

<sup>85</sup> Cp. Robert Nozick, ANARCHY, STATE AND UTOPIA (1974) (famously asking, as a challenge to the Lockean logic of labour and appropriation, “If I own a can of tomato juice and spill it in the sea so that its molecules (made radioactive, so I can check this) mingle evenly throughout the sea, do I thereby come to own the sea, or have I foolishly dissipated my tomato juice?”)

Our regulatory attention could turn, then, to ensuring the health and sustainability of the cultural and information ecosystem—which should include a renewed concern with encouraging of human creative expression, fostering a dynamic social dialogue, and preserving a participatory and inclusive culture—and not on establishing and enforcing protections against the mere making of millions of unseen incidental copies.

### *C. Copyright Serves the Struggling Artist*

Finally, we risk running headlong into the copyright trap when we assume that the allocation of private copyright control holds the answer to creators' economic struggles, empowering authors and artists to secure fair returns and future livelihoods.

The public discourse and debate around generative AI is brimming with references to the economic plight of the brilliant creators on whom our culture depends. Two aspects of this political narrative bear mention. First, there is a troublesome (to my mind) hierarchy drawn between the *real artists* and everybody else. Even putting aside obvious objections to aesthetic elitism and cultural exclusion, as a practical matter this obfuscates the reality that copyright makes no such qualitative distinctions, attaching to virtually every more-than-minimally original piece of expression in almost every conceivable form. The copyright rules that we deploy to protect the authors of bestselling novels (the Margaret Atwoods of the world chagrined by their inclusion in the Books3 dataset) will apply equally to the authors of tweets, selfies, and Reddit reponses to AITA?. The copyright rule that applies to protect the comic strip of a talented graphic artist will apply equally to a photograph of a patient's wound taken in a doctor's office. And, as I've suggested, the rules that we create to protect these venerable novelists and artists from the threats posed by generative AI may—if we are not careful—similarly protect against building or sharing data sets to conduct informational analysis that could reveal the likelihood that a tumour is benign, say, or the patterns of systemic discrimination discernible in immigration tribunal decision-making.<sup>86</sup> Given the minimal standards, automatic nature, and sweeping scope of copyright protection, it should not be assumed that TDM and AI initiatives focused on, for example, “medicine, agriculture, or logistics” would be shielded from the chilling impacts of copyright control over training data.<sup>87</sup>

The trope of the starving artist has been dusted off and successfully leveraged by market incumbents with the arrival of each new paradigm-shifting technology since the

---

<sup>86</sup> See, e.g., Sean Rehaag, *Luck of the Draw III: Using AI to Examine Decision-Making in Federal Court Stays of Removal*, REFUGEE LAW LAB WORKING PAPER (11 January 2023), <https://ssrn.com/abstract=4322881>.

<sup>87</sup> Cf. Frank Pasquale and Haochen Sun, *Consent and Compensation: Resolving Generative AI's Copyright Crisis*, 110 U. VA. L. REV. ONLINE, 49-50 (forthcoming, 2024), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4826695](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4826695) (suggesting that copyright-induced challenges to generative AI may encourage “[t]hose now working on perfecting AI-generated music, movies, and novels [to] turn their considerable talents to advancing computation in less copyright-intensive areas, such as medicine, agriculture and logistics”).



printing press. Unfortunately, the reality is that copyright ill-serves the artists and creators that it purports to save—this is the second problem with the narrative: The very parties nobly insisting upon the fair treatment of artists and authors are often the ones taking ownership of copyright from creators, asserting control over their works, and extracting the bulk of royalty payments before (sometimes) passing along any remaining benefits. Notably, Pasquale and Sun assume, for the purposes of their argument, that authors own the copyright in their works, explaining that their “proposal is agnostic as to whether authors, copyright owners, or both should be able to deny consent to use their works in AI.”<sup>88</sup> Unfortunately, this is not merely “a detail to be worked out,” as they suggest, but a central facet of the way that copyright works (or doesn’t) and the interests it serves (or fails to serve). As Rebecca Giblin demonstrates, the assumption that copyright will get authors paid has “never really held good.” But “[b]e that as it may,” she continues, “the idea that it gets authors paid is a core article of faith in support of copyright.”<sup>89</sup> The politically prevalent assertion that strengthened copyright protections will translate into compensation for professional creators is yet another well-worn fallacy—or conviction of blind faith—that leads us directly into the copyright trap.

In the fast-moving debate over generative AI, the predictable dynamics of our capitalist political economy are already apparent. The publishers, the recording industry representatives, and the copyright collectives are lining up, cap in hand, to paint a picture of the worthy, struggling artists they purport to serve and on behalf of whom they demand compensation. In theory, the more that users pay for the use of creators’ works, the more these intermediaries will pay creators for them—but we have spent decades expanding copyright while creators get poorer.<sup>90</sup> As Cory Doctorow explains, “creative workers have been conditioned—by big media and tech companies—to reflexively turn to copyright as the cure-all for every pathology....[But] turning every part of the creative process into ‘IP’ hasn’t made creators better off.” In fact, he points out, labor unions have a much better record of getting economic benefits to creators than have expanded copyright protections, which creators are then left to bargain away in the free market.<sup>91</sup> The recent slew of corporate agreements licensing massive datasets of published works for use in AI trainings sets is a sign of how this is likely to unfold: Whether one is a freelance author who published in the *New York Times* or a user who posted on Reddit,<sup>92</sup> it seems wrong

---

<sup>88</sup> Pasquale & Sun, *Consent and Compensation: Resolving Generative AI’s Copyright Crisis*, 110 U. VA. L. REV. ONLINE, 20-21 (forthcoming, 2024), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4826695](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4826695).

<sup>89</sup> Giblin, *A New Copyright Bargain? Reclaiming Lost Culture and Getting Authors Paid*, 41(3) THE COLUMBIA JOURNAL OF LAW & THE ARTS 369, 389 (2019), <https://doi.org/10.7916/jla.v41i3.2019>.

<sup>90</sup> Cory Doctorow, *Copyright won’t solve creators’ Generative AI problem: The machine-learning monkey’s paw*, MEDIUM (Feb 9, 2023).

<sup>91</sup> See, e.g., Angela Watercutter, *The Hollywood Strikes Stopped AI From Taking Your Job. But For How Long?* WIRED (Dec 25, 2023), <https://www.wired.com/story/hollywood-saved-your-job-from-ai-2023-will-it-last/>.

<sup>92</sup> See, e.g. Anna Tong, Echo Wang and Martin Coulter, *Exclusive: Reddit in AI content Licensing deal with GoodI, Reuters* (Feb 21, 2024), <https://www.reuters.com/technology/reddit-ai-content-licensing-deal-with-google-sources-say-2024-02-22/>.

to imagine that the inclusion of one's works in the dataset will bring about any personal windfall. Instead, boilerplate contracts and terms of service will dictate who reaps the benefits. As Louis Menard recently wrote:

Whatever happens, the existential threats of A.I. will not be addressed by copyright law. What we're looking at right now is a struggle over money. Licensing agreements, copyright protections, employment contracts—it's all going to result in a fantastically complex regulatory regime in which the legal fiction of information "ownership" gives some parties a bigger piece of the action than other parties.<sup>93</sup>

The position of market incumbents and established intermediaries in the AI copyright debates is reminiscent of the Stationers of old, when the proliferation of the printing presses first threatened to break up the publishers' monopoly—and thus copyright was born.<sup>94</sup> It should come as no surprise, in this context, that the loudest calls for exclusive rights over machine learning come from the companies and not the creators. In this latest installment of the literary property debates, industry demands for creators' compensation in the name of fairness should be met, I would suggest, with demands for financial transparency, corporate accountability, and the fair distribution of any royalties that might be collected. And before establishing any collective licensing model or opt-in/-out schemes in the name of compensating creators, policymakers must be implored to calculate the benefits that would actually flow to these creators. As I suggested above, given the vast numbers of works involved in producing AI outputs that will, themselves, have relatively little economic value, it seems highly unlikely that a pro-rata micro-share of license fees for copies contained in massive training data-sets could possibly make a significant difference in the economic lives of artists and authors, and certainly not enough to save them from the substitutionary effects of synthetic creativity at large. A significant percentage of the accumulated royalties could, however, nicely line the pockets of established intermediaries—but to whose benefit and at what cost?

## CONCLUSION

I have suggested, in this essay, that AI critics are being steered towards a copyright trap. The trap entices by deploying copyright's age-old but persistent entitlement narratives combined with compelling accounts of the obvious importance of human creators and creativity. But these lures ultimately lead us to reinscribe a proprietary logic that props up the privatization of knowledge and the consolidation of corporate control

---

<sup>93</sup> Menard, *Is A.I. The Death of I.P.?* THE NEW YORKER (Jan 15, 2024), <https://www.newyorker.com/magazine/2024/01/22/who-owns-this-sentence-a-history-of-copyrights-and-wrongs-david-bellos-alexandre-montagu-book-review>.

<sup>94</sup> See generally, e.g., L. Ray Patterson, COPYRIGHT IN HISTORICAL PERSPECTIVE (1968); Ronan Deazley, RETHINKING COPYRIGHT: HISTORY, THEORY, LANGUAGE (2006).

over information resources. Paradoxically but predictably, the retrenchment of copyright control over this next paradigm-shifting technology is poised to serve the interests of the media, cultural, and tech industries by parcelling out intellectual properties and letting market logic and private-ordering prevail. Calls for credit, consent, and compensation, even if met, are likely only to cast a legitimizing cloak over continued practices of commodification, alienation, and exploitation.

There are many good reasons to be concerned about the rise of generative AI – nothing here should be taken to mean that I am blind to the threats it poses or the harms it may wreak on our shared culture, our equality, and our lived experiences. Unfortunately, there are also many good reasons to be concerned about copyright’s growing prevalence in the policy discourse around AI’s regulation. Insisting that copyright protects an exclusive right to use materials for text and data mining practices (whether for informational analysis or machine learning to train generative AI models) is likely to do more harm than good. As many others have explained, imposing copyright constraints will certainly limit competition in the AI industry, creating cost-prohibitive barriers to quality data and ensuring that only the most powerful players have the means to build the best AI tools<sup>95</sup> (provoking all of the usual monopoly concerns that accompany this kind of market reality but arguably on a greater scale than ever before). It will not, however, prevent the continued development and widespread use of generative AI.

Moreover, deploying private copyright control in this way is likely to produce reliance on poor or incomplete data sets, which will affect the quality of the AI tools that are developed and sent out into our world, most likely in a way that will compound the risk of bias and exacerbate stereotyping in AI outputs and applications.<sup>96</sup> More broadly, given the way that generative AI models work (not as sources of meaning but essentially as probability calculators) and their likely importance in shaping the cultural landscape of the future, reliance on copyright permissions and licensing arrangements to determine the AI tools that are made and the materials on which they are trained will almost certainly *exclude* significant swaths of society from both the inputs and outputs of AI. We have to ask ourselves what kinds of texts, from what geographical and cultural contexts, are most likely to be compiled into valuable databases and offered up by major platforms and intermediaries to the highest bidder under the watchful eyes of copyright lawyers. If only certain texts from certain culturally privileged sources make it into the licensed dataset, the result is likely to be the reinforcement of cultural and knowledge hierarchies and the gradual erasure of voices from the margins and counter-hegemonic or divergent worldviews. As Michal Shur-Ofry has explained, the technical traits of generative AI

---

<sup>95</sup> As Shawn Presser, the creator of the “stolen” Books3 dataset explains, “If you really want to knock Books3 offline, fine. Just [know that] the world that you[re choosing is one where only billion-dollar corporations are able to create these large language models.” Quoted in Kate Knibbs, *The Battle Over Books3 Could Change AI Forever*, WIRED (Sept 4, 2023), <https://www.wired.com/story/battle-over-books3/>

<sup>96</sup> See generally Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem*, 93 WASH. L. REV. 579 (2018). See also Benjamin L.W. Sobel, *Artificial Intelligence’s Fair Use Crisis*, 41 COLUM. J.L. & ARTS 45, 96 (2017)

already mean that its outputs will tend towards the dominant, likely reflecting “a relatively narrow, mainstream view, prioritizing the popular and conventional over diverse contents and narratives.”<sup>97</sup> Perhaps, then, if the political goal is to push for equality, participation, and representation in the AI age, critics’ demands should focus not on exclusivity but *inclusivity*.<sup>98</sup> If we want to encourage the development of ethical and responsible AI, maybe we should be asking what kind of material and training data must be *included* in the inputs and outputs of AI to advance that goal. Certainly, relying on copyright and the market to dictate what is in and what is out is unlikely to advance a public interest or equality-oriented agenda.

Leading critics of AI have articulated important critiques about the inequalities, exploitation, and private profiteering that characterise the swelling AI industry; but these are also characteristics of the copyright industries, sustained by copyright law and its powerful political narratives. There is cause for concern, then, that the most compelling advocates for fairness, equality, and the public interest in the AI debates are running into an old copyright trap. It is up to those of us who recognize the risk to sound the warning before our policymakers, purporting to respond to these concerns, rewrite laws that will strengthen copyright structures of corporate control while limiting users’ rights and public benefits.

If copyright is not the solution, however, it might reasonably be asked: what is? The first step to answering that question—to producing a purposively sound prescription and evidence-based prognosis, is to correctly diagnose the problem. If, as I have argued, the problem is *not* that AI models are being trained on copyright works without their owners’ consent, then requiring copyright owners’ consent and/or compensation for the use of their work in AI-training datasets is *not* the appropriate solution. Of course, treating an imagined condition based on a misdiagnosis of the problem can have disastrous results. That is the crux of the argument advanced here. If the only real copyright problem is that the outputs of generative AI may be substantially similar to specific human-authored and copyright-protected works, then copyright law as we know it already provides the solution.<sup>99</sup> For other kinds of problems posed by the rapid rise of generative AI, of which there are many, the law may—or ought to—offer other solutions: Privacy harms may be addressed by improved privacy laws;<sup>100</sup> gross income inequalities

---

<sup>97</sup> See Michal Shur-Ofry, *Multiplicity as an AI Governance Principle*, 8-9 (May 10, 2023), <http://dx.doi.org/10.2139/ssrn.4444354>.

<sup>98</sup> See *id.*, proposing an AI-governance principle of “multiplicity by design.” See generally, Severine Dusollier, *Inclusivity in Intellectual Property*, in G. Dinwoodie (ed.), *INTELLECTUAL PROPERTY AND GENERAL LEGAL PRINCIPLES – IS IP A LEX SPECIALIS ?*, 101-118 (2015).

<sup>99</sup> See Carys Craig, AI and Copyright in Florian Martin-Bariteau & Teresa Scassa (eds), *ARTIFICIAL INTELLIGENCE AND THE LAW IN CANADA*, 33-35 (LexisNexis, 2021).

<sup>100</sup> See e.g. Daniel J. Solove, Daniel J. and Woodrow Hartzog, *The Great Scrape: The Clash Between Scraping and Privacy* (July 03, 2024), <https://ssrn.com/abstract=4884485> or <http://dx.doi.org/10.2139/ssrn.4884485>. Solove and Hartzog’s arguments against a consent model to regulate privacy in AI training data resonate with my own concerns about copyright control and consent: “Most consent in privacy laws is fictional. Such an approach would

may be better addressed by more targeted tax laws; corporate power and disproportionate profit may be better addressed by anti-trust and competition laws; the misuse of Indigenous knowledges could be better addressed by improved protections for TK/TCE; the continuing exploitation of knowledge producers in the Global South could be better addressed by contextual systems of knowledge governance and data sovereignty;<sup>101</sup> the underfunding of the arts could be better addressed by improved cultural policies and funding schemes; creative (and other) workers' rights could be better addressed by bolstered labor laws; the plight of struggling artists (and others) could be more effectively tackled by establishing Universal Basic Income, and so on....

Almost two decades ago, reflecting on regulatory dilemmas presented by the arrival of the Internet, Jessica Litman observed:

[T]he threat and promise of the Internet has induced those of us who are copyright lawyers to an act of breathtaking hubris. We define a set of rules that we say ought to be the basic copyright rules of the road, and then we construe those rules to govern every single way that information coded in electrons can move from one computer to another.<sup>102</sup>

As we contemplate both the threat and promise of AI technologies, I mean to caution against a similar kind of instinctive copyright overreach—copyright law is neither apposite nor equipped to govern the development of AI. Insisting that it should do so, and imagining that it is somehow up to the task, could be a course-setting and costly mistake.

---

subject individuals to data gathering and use on a massive scale, wrapping it in a farcical veneer of legitimacy.” *Id.* 54-55).

<sup>101</sup> See e.g. *Draft Cradle Principles on Knowledge Governance*, InfoJustice (Mar 5, 2024), <https://infojustice.org/archives/45627>.

<sup>102</sup> Jessica D. Litman, *DIGITAL COPYRIGHT* (2<sup>ND</sup> ED), 31 (Prometheus Books, 2006).