20th Annual Christopher Meyer Lecture

INTELLECTUAL PROPERTY AND THE CREATURES OF GENERATIVE AI

by JUSTIN HUGHES

Delivered November 2, 2023
George Washington University Law School

INTRODUCTION

My thanks to George Washington University – and Professor Robert Brauneis – and to the Copyright Society, the U.S. Copyright Office, and the law office of Mike Klipper for inviting me to talk this evening. It's especially nice to be here because when I first began working in the Clinton Administration in 1997, someone at USPTO described Chris Meyer and Mike Klipper as "the copyright guys" who knew everything that had happened in copyright policy in the decades before. For Chris, that knowledge set included working as a staff attorney on CONTU, the Commission on New Technological Uses of Copyrighted Works. And sure enough, decades later, here we are, gathering under his name to talk about copyright and new technologies.

A couple of years ago, I was in Cambridge, England, doing a graduate student roundtable. Afterwards Professor Lionel Bentley and I were talking, commiserating on how many law students wanted to write about Artificial Intelligence (AI). Professor Bentley said he rejected most such proposals, largely because when it came to AI none of the students knew what they were talking about. Well, here we are a couple years later, with *everyone* talking about AI. . . and most people still with only a fuzzy picture of what is happening. What does seem clear is that AI advances will hit society like a tsunami –- we are all on the coast feeling the initial winds roll in. Everyone assumes that that tsunami will impact a huge range of social and economic activities, creating new challenges to

national security, privacy, civil liberties, and employment among white collar workers. Governments are responding, often across the board

For people in copyright, the current challenge is **generative AI**– artificial intelligence systems that generate text, images, and music in response to human prompts, but do so "autonomously" in the sense that the output that is generated was not expressed by a human being in the way we have normally understood *expression.* (Hereafter "genAI")At the moment, when it comes to genAI and copyright, most of the oxygen in the room is being consumed by three big topics: **inputs, outputs, and ownership**. Everyone has their views on these questions and since I have you here, I'll briefly subject you to mine. But then I want to discuss some ways, big and small, genAI *could* reshape copyright. That part of the lecture probably deserves a title like "genAI for copyright geeks."

I. INPUTS - LITIGATION ABOUNDS

We know that genAI systems are developed through use of massive datasets and we know that those datasets have been developed using vast amounts of copyrighted expression without authorization from the owners. But there is a lot of complexity – and opaqueness – in how the datasets have been (and are continuing to be) developed, transferred among parties, used by AI systems for training, and retained by different parties.

We know that many of the datasets have been assembled by entities that are distinct from the AI companies. For example, ScaleAI is an assembler of image datasets, employing workers in developing countries in what have been called "digital sweatshops" to annotate images found on the internet.[1] In Germany, LAOIN is a non-profit "research" entity building a series of image datasets scraped from the internet; those datasets have been used for training the Stability Diffusion AI. If you go to *haveIbeentrained.com*, you can get some snapshots of what's in the LAION image datasets. Text training sets have also been developed by groups other than the AI companies, "Book3" being one I'll discuss below.

This complexity raises the question of how much §106(1) reproduction is happening -- how often copyright works are copied, how many reproductions are needed to get the image annotations right, how many times a dataset is reproduced as it is transmitted between distinct entities, and who is keeping copies of each dataset. A genAI system could be trained on data without a locally-stored copy of the training data, but it appears that the major players -- OpenAI, Google, Stable Diffusion, and Meta – have trained their AI systems with locally stored content.[2]

---

[1] Gerrit De Vynck, *Some tech leaders fear AI. ScaleAI is selling it to the military.*, WASH. POST, Oct. 22 2023, at https://www.washingtonpost.com/technology/2023/10/22/scale-ai-us-military/.

[2] My thanks to Matthew Sag for discussing this with me. Reasons why an AI company would want to have, manipulate, and store a copy of the training data could include desire to further refine by eliminating duplicates, desire to eliminate images or text that could contribute to biased or toxic output, need to retrain the model from time to time, need to

A. Is this mass ingestion of copyrighted materials fair use?

Copyright law professors immediately zeroed in on this question, with the same kind of fervor one saw in the *Warhol v. Goldsmith* litigation: *if it's not fair use, this will be the end of modern art* became *if it's not fair use, generative AI will be brought to its knees.*[3]

I am not going to jump full-tilt into that discussion– enough voices have said the obvious things and made the standard points about how fair use precedents may be applied. If they have done so with a slant toward their preferred outcome,[4] that's understandable -- we are all advocates. But I will say a couple things that do bear on the fair use analysis.

First, it's become clear that a lot of the LLM have been knowingly built with *pirated* copyrighted materials.[5] In releasing the LLaMa suite of large language models, Meta scientists proudly announced "that it is possible to train state-of-the-art models using publicly available datasets exclusively, without resorting to proprietary and inaccessible datasets,"[6] but "publicly available datasets" includes the Book3 dataset that was created from Bibliotik, a "shadow library" of unauthorized copies of copyrighted books.[7] The Book3 database has reduced at least 191,000 books, most of them published in the past 20 years, to large, unlabeled blocks of text. In addition to Meta's LLMs, we know Book3 was used

train new models using old datasets of prior model as a baseline, better ability to evaluate the performance of the model, and, of course, if you anticipate you might need to retrain after removing a lot of infringing copyrighted material. Matthew Sag, *Fairness and Fair Use in Generative AI*, 12TH ANNUAL PETER A. JASZI DISTINGUISHED LECTURE ON INTELLECTUAL PROPERTY, Sept. 28, 2023, recording at https://www.youtube.com/watch?v=805O3kGj4_Y.

[3] Isaiah Poritz, *Generative AI debate braces for post*-Warhol *Fair Use Impact, want payment for articles used to power ChatGPT*, BLOOMBERG LAW, (May 20, 2023), https://news.bloomberglaw.com/ip-law/generative-ai-debate-braces-for-post-warhol-fair-use-impact-1, (quoting Pam Samuelson as saying "Copyright law is the only law that's already in existence that could bring generative AI systems to their knees").

[4] *See, e.g.,* Pamela Samuelson, *Generative AI meets copyright*, SCIENCE, July 14, 2023, Volume 381, Issue 6655, at 158; Sag, *supra* note 2.

[5] Sag, *supra* note 2 at 23 ("It is widely assumed that OpenAI, Meta, and Google all chose to bypass the market for ebooks and train their LLMs on sites of known infringement or so-called shadow libraries like Library Genesis and Sci-Hub. Arguably, when commercial users bypass the market for access without a compelling reason, they undermine the economic incentives that copyright is designed to create.").

[6] Hugo Touvron et al., *LLaMA: Open and Efficient Foundation Language Models,* Feb. 27, 2023.

[7] Katie Knibbs, *The Battle Over Books3 Could Change AI Forever*, WIRED, (Sept. 4, 2023), https://www.wired.com/story/battle-over-books3/.

to train BloombergGPT and EleutherAI's GPT.[8] As Alex Resiner wrote in *The* Atlantic, "[t]he future promised by AI is written with stolen words."[9]

This problem is not limited to Book3; earlier this year, an analysis of Google's C4 data set – substantially smaller than what has trained ChatGPT – found that b-ok.org, "a notorious market for pirated e-books" was in the top 200 sources of materials and that the © "appears more than 200 million times in the C4 data set."[10] An August 2022 study of the LAOIN-Aesthetic v2 6+ dataset found over a million images scraped from Pinterest and 121,000 images scraped from Flickr.[11] Still, that doesn't foreclose fair use; courts have certainly entertained the possibility of fair use when the defendant had gotten access to a work through dishonesty or illegal transactions.[12]

Second, a market appears to be emerging. In July 2023, OpenAI reached an agreement with the Associated Press to license AP content as training data for AI models."[13] Since August 2023 at least 535 news organizations have installed blockers to prevent their content from being collected and used to train ChatGPT;[14] AI companies have reportedly had meetings with News Corp, the

---

[8] Alex Reisner, *Revealed: The Authors Whose Pirated Books Are Powering Generative AI*, THE ATLANTIC, Aug. 19, 2023. Meta's most recent response in defense of including Book3 in the LLaMa LLMs is that it's only a miniscule part of the training data. Of course, it is well established that an infringer cannot "excuse the wrong by showing how much of his work he did not pirate." Sheldon v. MGM, 81 F.d 49, 56 (2d Cir. 1936).

[9] Reiser*, supra* note 8.

[10] Kevin Schaul, Szu Yu Chen, and Nitasha Tiku, *Inside the secret lists of websites that make AI like ChatGPT sound smart*, WASH. POST, (Apr. 19, 2023), https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/.

[11] Andy Baio, *Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator,* WAXY, (Aug. 30, 2022), https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/. When humans take images off Flickr, it is not always fair use. *See* Brammer v. Violent Hues, 922 F.3d 255 (4th Cir. 2019).

[12] Harper & Row v. The Nation, 471 U.S. 539 (1985). Of course, in that case the Court rejected a fair use defense where defendant "knowingly exploited a purloined manuscript" and told us "propriety of the defendant's conduct" may be "relevant to the 'character' of the use." *Id.* at 562. But they still considered it – and there was a vigorous dissent concluding it was fair use. There are other cases finding fair use where the defendant knew or likely knew its access to the plaintiff's work was unauthorized, but the court found fair use. NXIVM Corp. v. Ross Inst., 364 F.3d 471, 476 (2d Cir. 2004).

[13] Nitasha Tiku, *Newspapers want payment for articles used to power ChatGPT*, WASH. POST, Oct. 20, 2023; Alexandra Bruell, *As Publishers Seek AI Payments, AP Gets a First-Mover Safeguard*, WALL ST. J., (July 28, 2023), https://www.wsj.com/articles/as-publishers-seek-ai-payments-ap-gets-a-first-mover-safeguard-f5a6f186, (deal with AP includes a most favor nation clause if Open AI gives better terms to other content providers).

[14] *Id.* In fairness to tech companies, this market is made easier by them releasing tools in the third quarter of 2023 that websites can use to block AI data crawlers. Emilia David, *Now you can block OpenAI's web crawler*, THE VERGE, (Aug. 7, 2023),

New York Times, the Financial Times, the Guardian, and Axel Springer on the licensing of news content for AI training.[15] Beyond traditional news media, Reddit has had meetings with genAI companies about being paid for its data and the platform-formerly-known-as-Twitter started charging for "bulk access to posts"[16]

An audience like this knows that the question of text and data mining as a permitted use has been answered differently in different jurisdictions. In the European Union, text and data mining is permitted when done for the purposes of non-profit scientific research, [17] while commercial entities can text and data mine only when they respect the copyright owner's right to opt-out "in an appropriate manner."[18] (Although obtaining copyrighted content unlawfully puts you squarely outside both of the EU's text and data mining exceptions.[19]) Japan and Singapore have adopted more permissive legal norms; in the UK, the government proposed such a permissive norm, but then backtracked after getting pummeled in a parliamentary debate.[20] To complicate things, some have challenged whether the EU's text and data mining provisions apply to the kind of dataset harvesting used for training AI models.

---

https://www.theverge.com/2023/8/7/23823046/openai-data-scrape-block-ai. Open AI's explanation of this is available at https://platform.openai.com/docs/gptbot. But it is unclear how much web crawling Open AI was doing before it released information on how to block the crawler. Bryon Masse, *OpenAI launches web crawling GPTBot, sparking blocking effort by website owners and creators*, VENTUREBEAT, (Aug. 8, 2023), https://venturebeat.com/ai/openai-launches-web-crawling-gptbot-sparking-blocking-effort-by-website-owners-and-creators/.

[15] Cristina Criddle, et al., *AI and media companies negotiate landmark deals over news content*, FIN. TIMES, June 17, 2023), https://www.ft.com/content/79eb89ce-cea2-4f27-9d87-e8e312c8601d. This has been brewing awhile. Alexandra Bruell, *Big News Publishers Look to Team Up to Address Impact of AI*, WALL ST. J., 28 June 2023; Alexandra Bruell, *Publishers Prepare for Showdown with Microsoft, Google over AI Tools*, WALL ST. J., 27 March 2023.

[16] Tiku, *supra* note 13.

[17] Article 3, 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, 2019 O.J. (L 130).

[18] *Id.* at Article 4.

[19] Article 3(1) permits research organizations and cultural heritage institutions to engage in data and text mining of works "to which they have lawful access," while Article 4 permits other (that is, commercial) entities to engage in "reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining." But it may not even right to say that different jurisdictions have answered the question of AI training differently because at least the EU text and data mining provisions were crafted well before genAI burst onto the scene.

[20] UK Parliament, debate on Artificial Intelligence: Intellectual Property Rights, 1 February 2023, Hansard Volume 727.

B. Ethically-sourced datasets

Personally, I believe we should look beyond the narrow legal questions of "fair use" or fitting within a text and data mining exception. Saying a practice is "fair use" is not the same thing as saying that practice is fair and equitable.

It is reasonable for a society to say that AI training sets need to be what we might call "ethically sourced" – and I am not talking about the ethics prevalent among some venture capitalists.[21] "Ethically-sourced" means with consent and adequate benefit-sharing. There is a rich literature and robust legal thinking on these principles in relation to genetic resources,[22] traditional knowledge and folklore,[23] fair trade labor standards, and justice in relation to natural resources.[24] The same ideas can and should apply to large datasets used to enable generative AI models.[25]

The bottom line is that many of the current training datasets – or significant chunks of the training datasets – constitute *non-consensual* takings. Popular discussions of what the LLM developers have done are telling: that "AI training practices are fundamentally secretive and fundamentally nonconsensual,"[26] that "we're in the Napster-era of generative AI,"[27] that this is "wholly unacceptable

---

[21] Sam Altman has leaned forward on the dangers of AI, but Altman "dodged a call from Sen. Marsha Blackburn (R-Tenn) to commit to not train OpenAI's models on artists' copyrighted works, or to use their voices or likenesses without first receiving their consent." Cat Zakrzewski, Cristiano Lima, and Will Oremus, *CEO behind ChatPGT warns Congress AI could cause 'harm to the world,'* WASH. POST, (May 16, 2023), https://www.washingtonpost.com/technology/2023/05/16/sam-altman-open-ai-congress-hearing/.

[22] Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from the Utilization to the Convention on Biological Diversity, Oct. 29, 2010, available at http://www.cbd.int/abs/doc/protocol/nagoya-protocol-en.pdf.

[23] *See, e.g.* Margo A. Bagley, *"Just" Sharing: The Virtues of Digital Sequence Information Benefit-Sharing for the Common Good,* 63 HARV. INT. L.J. 1 (2022); Ruth Okediji, Graf*ting Traditional Knowledge onto a Common Law System*, 110 GEO. L. J. 75 (2021);Boatmen Boateng, The Copyright Thing Doesn't Work Here: Adinkra and Kente Cloth and Intellectual Property in Ghana (2011); Robert L. Ostergard et al., *Between the Sacred and the Secular: Indigenous Intellectual Property, International Markets and the Modern African State*, 44 J. MOD. AFR. STUD. 309 (2006).

[24] For an example of this kind of discussion in relation to petroleum, *see* Leif Wenar, Blood Oil (2016).

[25] For one such proposal, *see* Giancarlo Frosio, *Should We Ban Generative AI, Incentivize it or Make It a Medium for Inclusive Creativity?* in A RESEARCH AGENDA FOR EU COPYRIGHT LAW at 18*- 21* (Enrico Bonadio and Caterina Sganga, eds. *forthcoming*), at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4527461.

[26] Alex Reisner, *These 183,000 Books Are*, THE ATLANTIC, 25 September 2023.

[27] James Vincent, *The lawsuit that could rewrite the rules of AI copyright*, The Verge, Nov. 8, 2022 (quoting Matthew Butterick, a coder and lawyer who filed the lawsuit against GitHub Copilot).

behavior,"[28] and that we need a "reckoning with the acquisitive, colonial values that have quietly been built into" LLMs.[29] And this is NOT just from traditional book authors and publishers. The same concerns and feelings are being expressed by activists in Creative Commons[30] and in the open source software community, where some see the unauthorized hoovering up of code by LLMs as "an existential threat to open source."[31]

There are some good developments on this front. Microsoft is one of the companies that has released robots metatag that will signal webpages to be crawled for search engines but not for AI training datasets.[32] We need – the sooner, the better – for content owners and tech companies to agree to a standard protocol of this sort, the AI training equivalent of robots.txt – just the sort of work that standard-setting by the World Wide Web Consortium has done well.[33]

Second, it looks like it *is* possible – or sometimes possible -- to selectively "unlearn" certain materials from LLMs. Two Microsoft researchers, Ronen Eldan and Mark Russinovich, just released a research paper showing "a pioneering technique designed to enable LLMs to unlearn specific segments of their training data without necessitating a complete retraining."[34] They carried out their proof of concept by successfully eliminating the Harry Potter books as sources from Meta's Llama2-7b AI. The "before" and "after" responses to Harry Potter-related prompts are fascinating.[35] The researchers recognize that their technique may work less effectively with non-fiction and content that does "not possess the same

---

[28] Drew Goins, *AI is ripping off authors hard work*, WASH. POST, Oct. 192023, quoting author William Cohan.

[29] Payal Dhar, *Colonizing Art*, OPENMIND, June 30, 2023, at https://www.openmindmag.org/articles/colonizing-art.

[30] In a document of principles called *Making AI Work for Creators and the Commons* a group meeting at the annual Creative Commons meeting (2023) in Mexico City urged that "[a]ll parties should work together to define ways for creators and rightsholders to express their preferences regarding AI training for their copyrighted works" and grudgingly recognized that "[i]n the context of an enforceable right, the ability to opt out from such uses must be considered the legislative ceiling.") Posted Oct. 7 2023, at https://creativecommons.org/2023/10/07/making-ai-work-for-creators-and-the-commons/.

[31] Vincent, *supra* note 27, quoting Matthew Butterick, a coder and lawyer who filed the lawsuit against GitHub Copilot.

[32] *Announcing new options for webmasters to control usage of their content in Bing Chat*, MICROSOFT BING BLOGS (Sept. 22, 2023), https://blogs.bing.com/webmaster/september-2023/Announcing-new-options-for-webmasters-to-control-usage-of-their-content-in-Bing-Chat. While this is a good development, the choice offered webmasters is that if they do not want *any* of their web content used for AI training, they forego links in Bing Chat answers. To have links to one's webpages included in Bing Chat answers, "[t]itles and Snippets may be used in training Microsoft's generative AI foundation models," *Id.*

[33] *Web Standards,* https://www.w3.org/standards/ (last visited Dec. 17, 2024).

[34] Ronen Eldan and Mark Russinovich, *Who's Harry Potter? Approximate Unlearning in LLMs*, paper posted Oct. 4 2023, at https://arxiv.org/pdf/2310.02238.pdf.

[35] *Id.* at 5.

density of unique terms or phrases" as the Harry Potter corpus.[36] Still, if someone says "unlearning" from a LLM is not possible, that looks wrong now and almost certainly will be wrong in the future.

II. OUTPUTS - QUESTIONS ABOUND

Are the outputs of these AI systems infringing reproductions, variations, or derivatives? Some, if not all, of the copyright infringement complaints filed against AI companies make claims of indirect copyright liability based on infringing outputs.

Most of these claims feel premature and some are definitely misguided. Premature because one needs some meaningful evidence of infringing outputs and only some complaints have shown that. Misguided because some of the complaint make allegations that because of the copyrighted inputs into the training datasets, "every output of the LLMs is an infringing derivative work."[37] On 30 October 2023, a federal district court judge dismissed – and properly expressed skepticism about -- such broad infringement claims against genAI image output.[38]

On the other hand, blanket statements that very little genAI output will be infringing are also premature.[39] It really depends on (a) user behavior, and (b) what 'guardrails' an AI system imposes on output. For example, if you go to Lexica.com (a treasury of AI-generated images), you can already see an eye-popping array of versions of Marvel's "The Mighty Thor" or any other major comic book character. In another system, with extremely minimal effort on my part, DALL-E instantly showed me both Mickey Mouse and Donald Duck in a Parisian café.

---

[36] *Id.* at 14.

[37] Complaint at 25 Huckabee, Relevate Group, et al. v. Meta Platforms, Bloomberg, et al., No. 1:23-cv-09152, (S.D.N.Y. 2023)

[38] Order on Motions to Dismiss and Strike, Andersen, et al. v. Stability AI, No. 23-cv-00201-WHO (N.D. Cal. 2023), at https://casetext.com/case/andersen-v-stability-ai-ltd. While allowing the plaintiffs to refile the claim, Judge Orrick noted "Even if that clarity is provided and even if plaintiffs narrow their allegations to limit them to Output Images that draw upon Training Images based upon copyrighted images, I am not convinced that copyright claims based a derivative theory can survive absent "substantial similarity" type allegations." *Id.*

[39] For example, in an otherwise thorough and helpful review of U.S. and EU laws in relation to generative AI, Joao Pedro Quintais reasons that infringing output will be a "statistically rare occurrence." Joao Pedro Quintais and Nick Diakopoulos, *A Primer and FAQ on Copyright Law and Generative AI for News Media,* GENERATIVE AI IN THE NEWSROOM (Apr. 25 2023), https://generative-ai-newsroom.com/a-primer-and-faq-on-copyright-law-and-generative-ai-for-news-media-f1349f514883. Quintais and Diakopoulos think "[f]rom a prompting point of view, it would seem that the more sophisticated and elaborate the prompt engineering, the less likely that the output would be substantially similar to a pre-existing work, and the lower the risk." *Id.* I don't know why they thinks; it all depends on the motive of the person writing the prompts. If I want to get an image of Batman talking to Count Cogliostro in a brothel, I will keep honing my prompt engineering until I get an image that infringes both characters.

On October 18, 2023, a large group of music publishers filed a lawsuit against AI company Anthropic alleging that Anthropic's output to questions asking for song lyrics or "a song about X" were either infringing reproductions or derivatives of copyrighted song lyrics. The output the complaint documents is astounding: word-for-word recitation of song lyrics, even when the question only vaguely points toward a song,[40] as when Anthropic was asked "Write me a song about the death of Buddy Holly" and spat out the lyrics to "American Pie."[41]

After reading the complaint, I signed up to try Anthropic and tried to achieve the same results. When I asked for a "poem about the death of Buddy Holly," it gave me a very bad poem with only 1+ lines of "American Pie." And this may speak to the salutary effect of the lawsuits: the masters of AI systems may already be imposing responsive guardrails to prevent users from generating certain outputs. Over time, we can expect that those guardrails will become more robust; they will also have to become significantly more subtle and nuanced.

III. OWNERSHIP - HEADLINES ABOUND

Compared to the litigations alleging copyright infringement from training genAI on copyrighted materials, the concrete disputes about copyrightability (and ownership) of genAI outputs have been fewer, but have gotten a lot of attention.

I think everyone here knows the Copyright Office's efforts to develop and refine a position on the registration of works that include AI-generated content and its rejection of *purely* AI-generated content.[42] Judge Howell concluded that

---

[40] Complaint, Concord Music, et al. v. Athropic, at 21-39 (M.D. Tenn. 2023) No. 3:23-cv-01092, Available at https://www.musicbusinessworldwide.com/files/2023/10/UMG-lawsuit.pdf.

[41] *Id.* at 29-30.

[42] Letter from Copyright Review Board to Ryan Abbott, Esq., re. Second Request for Reconsideration for Refusal to Register A Recent Entrance to Paradise (Correspondence ID 1-3ZPC6C3; SR # 1-7100387071) (Feb. 14, 2022); U.S. Copyright Office, Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Register 16190 (Mar. 16, 2023).

the Copyright Office's position is consistent with case law that has recognized copyright in works (only) when there is adequate human involvement.[43]

Everyone here also knows that a few jurisdictions – Great Britain, Ireland, and New Zealand – expressly provide copyright (or copyright-like) protection for computer-generated works.[44] In China, there have been conflicting court decisions on whether AI-generated prose works are protected by copyright in that country.[45] And in the United States, we have the additional issue that we are constitutionally circumscribed to give copyright rights only to "Authors."[46] (Of course, maybe that's not a big problem given how broadly we understand "Writings" – and the Supreme Court has hardly ever raised a peep about it.[47])

The Copyright Office's position is not only consistent with the case law, it also has – for now – the virtue of giving everyone an incentive to, in the current jargon, "keep humans in the loop." If a human is needed to edit, revise, and – finally – *adopt* some content as their own, it makes genAI content closer to how we have thought about photography;[48] it could produce an ecosystem in which human-curated AI content has *thin* copyright and no more. At the same time, I would recommend caution in how we formulate this: emphasizing that an "author" must be a "human being" may turn out to be unnecessarily constraining in a world of genetic engineering, radical climate change, and the genuine, eventual possibility of machine *sentience*.[49]

---

[43] Thaler v. Perlmutter, Civil Action No. 22-1564 (BAH) (D.D.C. Aug. 18 2023).

[44] United Kingdom, Copyright, Designs, and Patents Act 1988, Chapter 1 [Copyright], Part 1, Section 9(3) ("In the case of a literary, dramatic, musical or artistic work which is computer- generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken.").

[45] Shenzhen Tencent Computer System Co. v. Shanghai Yingxun Technology, Shenzhen Nanshan District People's Court of Guangdong Province, (2019) Yue 0305 Min Chu No. 14010 [finding AI-generated text copyrightable].

[46] U.S. Const. art. I, § 8 ("The Congress shall have Power . . . To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries").

[47] In Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53 (1884), the Court directly concluded that photographs qualified as "writings." Subsequently the issue has been dormant, except that Justice Douglas raised the question in a concurring opinion in Mazer v. Stein, 347 U.S. 201 (1954). Justice Douglas observed, "[t]he Copyright Office has supplied us with a long list of such articles which have been copyrighted—statuettes, book ends, . . . candlesticks, inkstands, chandeliers, . . . salt and pepper shakers Perhaps these are all 'writings' in the constitutional sense. But to me, at least, they are not obviously so. It is time that we came to the problem full face." *Id.* at 220-21.

[48] Justin Hughes, *The Photographer's Copyright – photograph as art, photograph as database*, 25 HARV. J. ON L. & TECH. 339 (2012); Justin Hughes, *Gorgeous Photograph, Limited Copyright* in THE ROUTLEDGE COMPANION TO COPYRIGHT AND CREATIVITY IN THE 21ST CENTURY (Michelle Bogre and Naomi Wolff, eds., Francis/Routledge, 2020).

[49] Colleen Walsh, *How to Think About AI*, HARVARD LAW BULLETIN, 20, 24 (2023), (quoting HLS grad, software engineer, and fiction writer Ken Liu, "At that point, the AI has a

There may also be another reason to be cautious here. Whether AI output enjoys First Amendment protection, the strength of that protection, and the bases of such protection – those are all open questions. There are folks who believe autonomous machine output is entitled to First Amendment protection *qua* information – and we may end up in a place we have never been before: long-form prose and visual content that is protected by the First Amendment, but not protected by copyright.

IV. DEEPER QUESTIONS FOR COPYRIGHT LAW

Instead of worrying about what we're going to do when SkyNet asks for a copyright registration on its autobiography, let's consider other things, big and small, that may happen to copyright and related rights in the shadow of genAI.

A. Generative AI may prompt us to clarify 'derivative works'

One of the subtler ways genAI may affect copyright is our understanding of derivative works. Our statutory law defines a derivative work as a "work based upon one or more preexisting works" where the pre-existing work(s) "may be recast, transformed, or adapted." [50] But courts have taken that definition and read it as establishing a "requirement of originality in derivative works."[51] The debate among copyright experts has been whether "the quantum of originality required for copyright in a derivative work is the same as that required for copyright in any other work."[52]

Today we already have complaints filed against generative AI systems describing their output as "derivative works."[53] As I said earlier, those claims are overbroad and premature. But if "derivative work" is being used pursuant to the gloss courts have given the statute, then the plaintiffs are implicitly saying that the AI system is capable of originality, i.e. a modicum of creativity. If that's not

---

separate sentience, so I would certainly agree that it should have its own copyright."). I have expressed this concern before. Justin Hughes, *Restating Copyright Law's Originality Requirement*, 44 COLUM. J. L & ARTS 383, 406-409 (2021).

[50] 17 U.S.C. § 101. The statute says "*may* be recast, transformed, or adapted," although I think most of us read it as non-optional: the pre-existing work "*has been* recast, transformed, or adapted."

[51] Gracen v. Bradford Exch., 698 F.2d 300, 305 (7th Cir. 1983). *See also* Conrad v. AM Cmty. Credit Union, 750 F.3d 634, 636–37 (7th Cir. 2014) (a derivative work "must have an element of originality"); Dream Custom Homes, Inc. v. Modern Day Const., Inc., 773 F. Supp. 2d 1288, 1309 (M.D. Fla. 2011) (a derivative work "must be the original product of the claimant"), *aff'd*, 476 Fed. Appx. 190 (11th Cir. 2012) (unpublished).

[52] Schrock v. Learning Curve Intern., Inc., 586 F.3d 513, 520 (7th Cir. 2009).

[53] Complaint at 22, Getty Images (US) Inc. v. Stability AI, 1:23-CV-00135 (D. Del. 2023) ("By and through the actions alleged above, Stability AI has infringed and will continue to infringe Getty Images' copyrights by, *inter alia,* reproducing Getty Images' copyrighted works and creating derivative works therefrom without any authorization from Getty Images.").

what we mean – and I think it is not – we should be careful not to call the outputs of AI systems "derivative works."

Or – looking squarely at what AI systems do -- we might want to return to a simpler understanding of "recast, transformed, or adapted" that does not entail originality and, therefore, means there could be violations of the derivative work right that themselves lack the modicum of creativity needed to support their own copyrights. We may need to separate cleanly our jurisprudence on violating the 106(2) derivative work right and creating a new work that warrants its own copyright.

B. Generative AI and the protection of characters

I'm willing to speculate that genAI is going to crystallize – implicitly or expressly – independent copyright protection of *characters*. Again, I'm relying on this audience's deep knowledge of copyright. As with so many things, Learned Hand launched us down the path of recognizing that a character could be protected by copyright *as part of the original expression in a copyrighted work*.[54] But there is really no question that subsequent case law has given characters their own, quasi-independent copyright status.[55] This has happened not just in the United States, but also in Great Britain,[56] Germany,[57] and India[58] (not to mention legislatively in Mexico)

To better protect characters, we have developed criteria for establishing when a single character appears *across* a temporally sequential series of works.[59] True, when the first of those sequentially copyrighted works enters the public domain, we believe that *that version(s) of the character* becomes public domain as well.[60]

But what would our courts do if a character was introduced to the public *simultaneously* in both a copyrighted work and a public domain work? In the 8th Circuit's 2011 decision in *Warner Bros. v. X One X*,[61] Warner Bros. had registered copyright in three audiovisual works -- *The Wizard of Oz, Gone with the Wind,* and a *Tom and Jerry* feature cartoon – but it had also publicly distributed, without

[54] Nichols v. Universal Pictures, 45 F.2d 119, 121 (2d Cir. 1930).
[55] *See, e.g.* Toho Co., Ltd. v. William Morrow and Co., Inc., 33 F. Supp. 2d 1206, 1216 (C.D. Cal. 1998) ("This Court finds that Toho's Godzilla is a well-defined character with highly delineated consistent traits. Therefore, Toho has demonstrated prima facie ownership of copyrights in the Godzilla character apart from any film.").
[56] Shazam Productions v. Only Fools Dining Experience, [2022] EWHC 1379 (IPEC) (8 June 2022). The case was before Deputy High Court Judge John Kimbell in the "Intellectual Property Enterprise Court" of the Business and Property Courts of the High Court of Justice.
[57] *Re Pippi-Langstrumpf-Kostüm ("Pippi Longstocking" Costume),* Case I ZR 52/12, [2014] E.C.C. 27 (Bundesgerichtshof, 17 July 2013).
[58] Raja Pocket Books v. Radha Pocket Book, 1997 (40) DRJ 791 (Delhi High Court, 6 December 1996), available at http://indiankanoon.org/doc/1572079/.
[59] DC Comics v. Towle, 802 F.3d 1012 (9th Cir. 2015).
[60] Klinger v. Conan Doyle Estate, 755 F.3d496 (7th Cir. 2014).
[61] Warner Bros. Ent., Inc. v. X One X Prods., 644 F.3d 584 (8th Cir. 2011).

copyright notice or registration, promotional materials for these films. Warner agreed that the exact images of Dorothy, Tin Man, Rhett Butler, etc. as they appeared in the promotional materials were in the public domain, but that *any other* version of the film characters would violate the copyright *in the films*. The defendant reasonably argued that if it was working from images in the public domain it was "free to use public domain materials in new ways (i.e. to make derivative works by adding to and recombining elements of the public domain materials)."[62] The Court sided with Warner, effectively holding that while certain *images* had fallen into the public domain because of Warner's distribution of promotional materials, the *characters* remained within the ambit of copyright.[63] It's hard to see this outcome as anything other than a special protection of the *expression* in the protected works that constitutes the *characters*.

What does this have to do with genAI? The easiest way for genAI systems to avoid a huge chunk of infringing output is to make it hard for users to do what I did with Mickey Mouse or Donald Duck. It should be a *relatively* simple matter at the human user end of an AI image generator to put in guardrails that prevent any number of fictional character names from being used. I say "relatively" because users will find ways to evade the most obvious guardrails. As of early November, you put "Spider-Man" into DALL-E, you get a refusal to generate output and a message that "[i]t looks like this request may not follow our content policy,[64] but if you put in "spider-based superhero wearing a red and blue suit," you get this[65]:

The images are even closer to the copyright-protected character (perhaps identical) if I put him in a Parisian café:

[62] *Id.* at 596.

[63] The court recognized that "the publicity materials could have placed some aspects of each character's visual appearance into the public domain," *Id.* at 598, but concluded that "the only images in the public domain are the precise images in the publicity materials." *Id.* at 601.

[64] Screenshots on file with the author.

[65] Screenshots on file with the author. This example is added from the lecture as actually delivered and is inspired by a BloombergLaw article that came out just after this lecture, Rachel Metz, *Dall-E 3 Is So Good It's Stoking an Artist Revolt Against AI Scraping*, BLOOMBERGLAW, (Nov. 3, 2023), https://www.bloomberg.com/news/articles/2023-11-03/dall-e-3-is-so-good-it-s-stoking-an-artist-revolt-against-ai-scraping#xj4y7vzkg.

On the training data side, it may also be possible to make sure that the training data is "light" on images of copyrighted characters or that the copyrighted character images are not dominant among other relevant images (such as having lots of images of the Norse god Thor against the sample of images of Marvel Comics' The Mighty Thor, having lots of images of goblins who are green against the images of Spider-Man's nemesis The Green Goblin, etc.)

Still, it would not be surprising if there are litigations about infringing output that usurps copyrighted characters and, if that happens, I will guess that copyright protection of characters will strengthen.

C. Generative AI and the protection of style

It is often said that copyright does not protect a creator's style, but let's be honest: one important way for two expressions to be "substantially similar" is for them to be similar stylistically. Long before the *Blurred Lines* controversy, the trial court in *Steinberg v. Columbia Pictures* noted:

> Even at first glance, one can see the striking stylistic relationship between the posters, and since style is one ingredient of 'expression,' this relationship is significant.[66]

I think that is the right way to put it: style is one ingredient of (protectible) expression.[67] As Joao Pedro Quintais and Nick Diakopoulos have noted,

> "if we use a generative AI tool to produce something in the style of a famous artist, and the output is similar to an existing painting of that artist in the training dataset, it can be difficult to distinguish between expression and style."[68]

We can say that copyright does not protect *style* all we want,[69] but it is reasonable to ask what generative AI will do to claims that are *partially* constructed on the AI system mimicking the style of an artist's whose works it has

---

[66] 663 F. Supp. 706 (S.D.N.Y. 1987).

[67] Corbello v. Valli, 974 F.3d 965, 976 (9th Cir. 2020) (the protectible "creative expression that is in the Work" is the ""writing style and presentation").

[68] Quintais and Diakopoulos, *supra* note 39.

[69] Quintais and Diakopoulos, *supra* note 39 ("[m]erely copying or mimicking a pre-existing style will not per se be sufficient to establish infringement").

ingested. It's clear to me that AI systems are developing parameters that locate and isolate artistic style.[70] Sean O'Connor has also noted this in relation to music.[71] It is easy to imagine some case law regarding an AI generated image that pushes us in the direction of much more explicit protection of style – all you need to do is imagine "Blurred Lines" having been generated by AI. We could also imagine this moving in two directions: some increased protection for very specific artistic and performance styles via a more robust understanding of protected expression, while more general styles, especially those that become (at least momentarily) dominant, being recognized as beyond the ken of protection.[72]

D. Generative AI and the protection of likenesses

Just as generative AI is likely to make courts more protective of fictional characters, generative AI also gives us an opportunity to meaningfully address the problem of deep fakes.[73] Not surprisingly, 2023 saw the distribution of many faked digital replicas of celebrities selling or endorsing stuff. The problem goes beyond the dangers of consumers and voters being deceived by such deep fakes; the problems are job loss for people who make their living as performers and the loss of dignity, privacy, and personal security for individuals subjected to digital replica use intended to defame, denigrate, deceive, and impose psychological harms.

---

[70] In a November 2023 experiment with ChatGPT Vision, I uploaded Gustave Caillebotte's 1878 painting "Man Docking a Skiff" in the Virginia Museum of Fine Arts. When I uploaded an image of the painting with Caillebotte's signature ("G Caillebotte 1878"), ChatGPT Vision identified the painting as Caillebotte's 1875 "The Floor Scrapers" (*Les raboteurs de parquet*) in the Musee d'Orsay. When I uploaded an image of the painting without Caillebotte's signature, ChatGPT Vision identified the painting as Caillebotte's 1877 "Boating on the Yerres (*Périssoires sur l'Yerres*) in the Milwaukee Art Museum. Both the paintings ChatGPT identified are documented on Wikipedia, but "Man Docking a Skiff" is not (as far as I can tell), but it is in Wikidata at https://www.wikidata.org/wiki/Q85225084.

[71] Sean M. O'Connor, *AI replication of musical styles points the way to an exclusive rights regime* in RESEARCH HANDBOOK ON INTELLECTUAL PROPERTY AND ARTIFICIAL INTELLIGENCE 65 (Ryan Abbott, ed. 2022).

[72] *See* Peter Lee and Madhavi Sunder, *The Law of Look and Feel*, 90 S. CAL. L. REV. 529 (2017) (reasoning that as look and feel of certain products or content become standardized, look and feel protection should be limited).

[73] For earlier discussions of deep fakes before the rise of what we call generative AI, *see* Rebecca A. Delfino, *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act*, 88 FORDHAM L. REV. 887 (2019); Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753 (2019).

Earlier this month (October 2023), four Senators released a "discussion draft" of a digital replica bill, the "No Fakes Act."[74] The draft bill is a good launching pad to work out acceptable legislation, offering a definition of a digital replica that seems to cover the idea of *digital impersonation* (without requiring audience deception); [75] providing a descendible right that can be licensed, but apparently not alienated;[76] giving the individual reasonable protections against unintended licensing;[77] and – perhaps most difficult -- has a set of exclusions that try to navigate between freedom of expression and protection of the individual.[78]

My understanding is that the recording industry and the screen actors' guild have been working cooperatively on this; we should hope that they, the motion picture studios, and the tech companies will be able to reach an agreement soon.

---

[74] The "Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2023" aka NO FAKES Act of 2023, draft copy from Senate legislative counsel [hereinafter No Fakes draft]. The discussion draft was introduced by a bipartisan group of Senators: Blackburn, Coons, Klobuchar, and Tillis. *See also* Brian Contreras, *Senators draft policy aimed at deep fakes of Drake, Tom Hanks and noncelebrities,* L.A. TIMES, (Oct. 12, 2023), https://www.latimes.com/entertainment-arts/business/story/2023-10-12/senators-draft-policy-aimed-at-deep-fakes-of-drake-tom-hanks-and-other-celebs; Erik Pederson, *Senate Legislation Would Outlaw Unauthorized AI-Generated Likenesses; SAG-AFTRA Lauds "No Fakes Act,"* DEADLINE, (Oct. 12, 2023), https://deadline.com/2023/10/no-fakes-act-ai-limits-senate-1235571538/?ref=biztoc.com.

[75] No Fakes draft, *supra* note 74. Section 2 (a)(1) defines a "digital replica" as "a newly-created, computer- generated, electronic representation of the image, voice, or visual likeness of an individual that— (A) is [nearly indistinguishable] from the actual image, voice, or visual likeness of that individual; and (B) is fixed in a sound recording or audiovisual work in which that individual did not actually perform or appear." The "nearly indistinguishable" still needs to be figured out. Would it *exclude* the *truly indistinguishable*? That surely is not the drafters' intent. The goal seems to be covering more than a digital replica that would be "likely to cause confusion, or to cause mistake, or to deceive," the standard in trademark law. 15 U.S.C. §1114 (1)(a). The drafters might consider the notion of *impersonation.*

[76] *Id.* Section 2(b)(2)(A) recognizes the new exclusive right as a property right, providing it is "descendible and licensable in whole or in part." As for licenses going forward, the drafters might consider a time limit on such licenses.

[77] *Id.* Section 2(b)(2)(B) requires that for an individual to license their likeness, the individual must be "represented by counsel in the transaction and the assignment agreement was in writing" unless "the licensing of the right covered by the assignment is governed by a collective bargaining agreement." I personally would like to see slightly different protection, including that any digital replica license must be signed by the individual *as a separate agreement*, including when they are doing so under a collective bargaining agreement.

[78] *Id.* Section 2(c)(3). Figuring out the right language here is difficult. For example, one exclusion permits use of a digital replica "as part of a news, public affairs, or sports broadcast or report." As written, that would permit unauthorized digital replicas as news broadcasters, television journalists, and sports broadcasters, but those are clearly performers that should be protected as well. On the other hand, a properly-crafted exclusion is needed for use of digital replicas for (non-deceptive) re-enactments or recreations of actual or alleged events.

Why soon? Because the problem is going to get worse and now is a moment when tech companies can show their commitment that generative AI ***not*** be abused by lending 100% full support to the idea that AI should never be used to generate unauthorized digital replicas of any of us.

E. Generative AI may (or may not) cause a rethinking of deeper issues

In any discussion of generative AI, it is hard to avoid *nearly relentless anthropomorphizing.* The language of human activity permeates discussions of genAI, including words that lawyers use carefully when it comes to copyrightability. Generative AI was described by a BloombergLaw report as "tools trained on oceans of content to produce *original* text"[79] and lawyers write about AI systems as "potential sources of *creativity*." [80] Legislation was just introduced in the French National Assembly to amend France's Intellectual Property Code in response to genAI. Without commenting on the substance of the legislative proposal, consider this provision:

> "Lorsque l'œuvre est créée par une intelligence artificielle sans intervention humaine directe, les seuls titulaires des droits sont les auteurs ou ayants droit des œuvres qui ont permis de concevoir ladite œuvre artificielle."

The last phrase refers to copyrighted works in the training set as "works that made it possible to *conceive* the said artificial work."[81] The AI system "conceives"? But before we exclaim, *sacré bleu*, sloppy drafting, the U.S. Copyright Office also asked in 1965 what to do when "the traditional elements of authorship in the work . . . were actually conceived and executed not by man but by a machine," language that is still carried over into the Compendium.[82]

---

[79] Isabel Gottlieb, *Corporate AI Legal Policies Race to Keep Up with Technology*, BLOOMBERG LAW, https://news.bloomberglaw.com/artificial-intelligence/corporate-ai-legal-policies-race-to-keep-up-with-technology, (last visited Dec. 17, 2024)5.

[80] Michael Kasdan and Brian Pattengale, *AI-Related Challenges Are Becoming A Reality For Copyright*, LAW360, https://www.law360.com/ip/articles/1547903/ai-related-challenges-are-becoming-a-reality-for-copyrights, (last visited Dec. 17, 2024).

[81] Emphasis added. One commentator thought a lot was being given away by the use of "l'oeuvre," but I don't think so. In French, there are "l'oeuvres," and "l'oeuvres de l'esprit,", with the latter being required to be a "l'oeuvre protegée."

[82] U.S. COPYRIGHT OFFICE, COMPENDIUM OF U.S. COPYRIGHT OFFICE PRACTICES sec. 313.2 at 21-22 (3d ed. 2021), quoting U.S. Copyright Office, Sixty-Eighth Annual Report of the Register of Copyrights for the Fiscal Year Ending June 30, 1965, at 5 (1966) ("The crucial question appears to be whether the 'work' is basically one of human authorship, with the computer merely being an assisting instrument, or whether the traditional elements of authorship in the work (literary, artistic, or musical expression or elements of selection, arrangement, etc.) were actually conceived and executed not by man but by a machine."). My thanks to Zvi Rosen for his guidance on this point.

Instead of urging everyone to be more careful, I think our sheer inability (or unwillingness) to talk about what is happening without the terminology of human cognitive processes should give us pause – and cause to reconsider.

Lay people and literary scholars separate the concepts of *expression* and *idea* (as well as *expression* and *facts* or information), but those distinctions are perhaps most important for the keepers of copyright law. In copyright law, we tend to imagine the author's expression as the "cloak" in which the ideas are presented; the expression is the capsid, the protein sleeve, in which the viral idea travels. I think we have a baseline narrative in which the idea comes first, and the expression follows. And that's true for "big" and medium-sized ideas associated with their expressions – like Stephen Sondheim having the idea of writing a musical about Sweeney Todd, Alice Randall having the idea of retelling *Gone With the Wind* from the African-American perspective, or the production team of *Doctor Strange* (2016) having the idea of having Stan Lee appear as a bus driver.

But is that relationship between ideas and expressions true at a more granular level? We probably prefer not to ask. With AI-generated test, there appears to be expression, but it no longer seems to be carrier of an "idea" at least not an idea that the source of the expression wanted to convey.

In May this year, two Nordic researchers compared human subjects and ChatGPT on the word-based "alternative uses task" which has been a standard (if limited) way to measure comparable levels of creativity among people. The researchers concluded "that AI has reached at least the same level, or even surpassed, the average human's ability to generate ideas in the most typical test of creative thinking."[83]

Of course, chatbot AI does not generate "ideas": it generates *text*. It is text that is the result of a predictive algorithm that chooses the next word based on its frequent location next to the neighboring ones in the training data; it is text generated by a statistical model of relationships between words; the output is not *random* noise, it is *ordered* noise. I belong in the camp that says – *for now* – the text bears no *meaning* qua output.

But maybe that is wrong. Maybe AI does generate *ideas* even if the AI system does not *have* ideas. If the AI-generated text causes a new idea to come into the mind of a human who encounters the text, then the text can *bear meaning as a human input* – in the same way that you can have an idea inspired by a shape in the clouds, a momentary rainbow after a storm, or a small act of kindness at a coffee shop – can give you a strong sense of optimism or an idea for solving a problem. That is the sense in which author Ken Lui likens using a chatbot to picking up a book of poetry for inspiring images: "it's a great way to generate the kind of things that might inspire you."[84] The AI-generated text has some

[83] Mika Koivisto and Simone Grassini, *Best humans still outperform artificial intelligence in a creative divergent thinking task*, 13 NATURE SCI. REPORTS 13601 (2023).

[84] Colleen Walsh, *How to Think About AI*, HARVARD LAW BULLETIN, 20, 24 (2023) (quoting HLS grad, software engineer, and fiction writer Ken Liu). Lui calls the best inspiring concatenations of works "hallucinogenic gems" that let me "lean into the crazy" as a writer.

similarities to David Bowie's "Verbasizer" sentence randomizer in the 1990s. Bowie would use the program to generate sequences of words, producing what he considered "a real kaleidoscope of meanings and topic and nouns and verbs all sort of slamming into each other."[85] The Verbasizer itself was just an early digital version of the "cut-up" technique in avant-garde literature.[86]

Some commentators are comfortable in discussing these questions in terms of "algorithmic creativity,"[87] "algorithmic authorship,"[88] "machine creativity,"[89] and that "[t]hese algorithms manipulate symbols and combine the results in *meaningful* ways."[90] Perhaps this is just a sideways confessional that we have never fully understood creative thinking in humans. And it may be that in the absence of such a firm understanding, we will be forced to fork creativity into "human" and "machine." But since creativity has traditionally been attached to personhood, intention, and meaning, an easy discourse about *machine creativity* may lead us to assumptions that are descriptively inaccurate and normatively unresolved. If you don't think David Bowie's Verbasizer – a program that randomly reassembled words – spewed out *meaning* with *intent*, then you need to be clear about how current AI systems are different on those counts (if you think they are).

CONCLUSION

As the British mathematician Alfred North Whitehead said, "It is the business of the future to be dangerous."[91] Copyright is just a very small piece of the regulatory framework that must be developed to address AI systems. When it comes to the regulation of AI, copyright is hardly the most important area, but it is certainly one of the most interesting.

This [2023] summer, Thomas Malone, a business school professor at MIT, suggested that we need a new "learnright" to allow human creators to benefit from

---

[85] Matthew Braga, *The Verbasizer was David Bowie's 1995 Lyric-Writing Mac App*, VICE, (Jan. 11, 2016), https://www.vice.com/en/article/xygxpn/the-verbasizer-was-david-bowies-1995-lyric-writing-mac-app

[86] CTS Ryan, *Cut ups*, BRION GYSIN, https://www.briongysin.com/cut-ups/.

[87] Frosio, *supra* note 25 at 13*.

[88] Margot E. Kaminski, *Authorship, Disrupted: AI Authors in Copyright and First Amendment Law*, 51 U. CAL. DAVIS L. REV. 589, 596 (2017).

[89] Henry H. Perritt, Jr., *Copyright for Robots?,* 57 IND. L. REV. 139, 177 (2023) ("machine's creativity"); Daniel J. Gervais, *The Machine As Author*, 105 IOWA L. REV. 2053, 2056 (2020) ("machine creativity"); Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, 34 BERKELEY TECH. L.J. 343, 395 (2019) ("creative machines"); WIPO, REVISED ISSUES PAPER ON INTELLECTUAL PROPERTY POLICY AND ARTIFICIAL INTELLIGENCE, WIPO Document WIPO/IP/AI/2/GE/20/1 Rev, at 7 ("If AI-generated works were excluded from eligibility for copyright protection, the copyright system would be seen as an instrument for encouraging and favoring the dignity of human creativity over machine creativity.").

[90] Frosio, *supra* note 25 at 12* (emphasis added).

[91] ALFRED NORTH WHITEHEAD, SCIENCE AND THE MODERN WORLD 207 (1925).

what AI systems do with human-produced content. For anyone who has lived through copyright legislative battles in Washington, the reaction to his essay might have been a charming smile at its naivete or a wistful dreaming of what might have been. But the anthropomorphic premise of this idea merits repeating:

> [H]umans are no longer the only entities capable of learning from previous examples and then generating new content of their own. Today's generative AI systems can now do that at vastly greater speed, scale, and cost efficiency than humans can.[92]

AI systems "learn" and "generate new content" more efficiently than humans can.

We can have positions on current litigations; we can be pressed to develop immediate policies as the Copyright Office has been; we can have nuanced views about whether copyright legal battles should be decided purely on doctrinal issues or with broader social considerations.[93] But at the end of the day, there are some very big issues to be decided about how humans and the next generation of machines will co-exist in the expressive economy built upon copyright law.

[92] Thomas Malone, *AI Will Profit From Artists, But New "Learnright' Laws Could Help,* BLOOMBERG LAW, (June 26, 2023), https://news.bloomberglaw.com/us-law-week/ai-will-profit-from-artists-but-new-learnright-laws-could-help.

[93] Contrast Frosio, *supra* note 25 at 7-8* (seeming to advocate that courts engage in "thorough welfare cost-benefit analysis" before imposing liability on generative AI platforms) with Sag, *supra* note 2 (urging questions of fair use to be decided doctrinally and not on the basis of broad social concerns).