To the Copyright Office: Comment on Artificial Intelligence & Copyright, Docket 2023-6

From Jeff Jarvis, Leonard Tow Professor of Journalism Innovation at the City University of New York's Craig Newmark Graduate School of Journalism

September 3, 2023

I am a journalist and journalism professor at the City University of New York. I write — speaking for myself — in reply to the Copyright Office's queries regarding AI, to bring one perspective from my field, as well as the context of history. I will warn that precedents set in regulating this technology could impinge on freedom of expression and quality of information for all. I also will share a proposal for an updated framework for copyright that I call creditright, which I developed in a project with the World Economic Forum at Davos.

First, some context from present practice and history in journalism. It is ironic that newspaper publishers would decry AI reading and learning from their text when journalists themselves read, learn from, rewrite, and repurpose each others' work in their publications every day. They do the same with sources and experts, without remuneration and often without credit. This is the time-honored tradition in the field.

The 1792 US Post Office Act provided for newspapers to send copies to each other for free for the express purpose of allowing them to copy each other, creating a de facto network of news in the new nation. In fact, many newspapers employed "scissors editors" — their actual job title — to cut out stories to reprint. As I recount in my book, *The Gutenberg Parenthesis: The Age of Print and Its Lessons for the Age of the Internet* (Bloomsbury Academic, 2023, 217), the only thing that would irritate publishers was if they were not credited.

As the Office well knows, the Copyright Act of 1790 covered only books, charts, and maps, and not newspapers or magazines. Not until 1909 did copyright law include newspapers, but even then, according to Will Slauter in *Who Owns the News?: A History of Copyright* (Stanford University Press, 2019), there was debate as to whether news articles, as opposed to literary features, were to be protected, for they were often anonymous, the product of business interest more than authorship. Thus the definition of authorship — whether by person, publication, or now machine — remains unsettled.

As to **Question 1**, regarding the benefits and risks of this technology (in the context of news), I have warned editors away from using generative AI to produce news stories. I covered the show-cause hearing for the attorney who infamously asked ChatGPT for citations for a federal court filing. I use that tale as an object lesson for news organizations (and search platforms) to keep large language models far away from any use involving the expectation of facts and

credibility. However, I do see many uses for AI in journalism and I worry that the larger technological field of artificial intelligence and machine learning could be swept up in regulation because of the misuse, misrepresentation, factual fallibility, and falling reputation of generative AI specifically.

AI is invaluable in translation, allowing both journalists and users to read news around the world. I have tested Google's upcoming product, NotebookLM; augmentative tools such as this, used to summarize and organize a writer's research, could be quite useful in improving journalists' work. In discussing the tool with the project's editorial director, author Steven Johnson, we saw another powerful use and possible business model for news: allowing readers to query and enter into dialogue with a publisher's content. Finally, I have speculated that generative AI could extend literacy, helping those who are intimidated by the act of writing to help tell — and illustrate — their own stories.

In reviewing media coverage of AI, I ask you to keep in mind that journalists and publishers see the internet and now artificial intelligence as competition. In an upcoming book, I assert that media are embroiled in a full-fledged moral panic over these technologies. The arrival of a machine that can produce no end of fluent prose commodifies the content media produce and robs writers of our special status. This is why I teach that journalists must understand that their value is not resident in the commodity they produce, content, but instead in qualities of authority, credibility, independence, service, and empathy.

As for **Question 8** on fair use, I am no lawyer, but it is hard to see how reading and learning from text and images to produce transformative works would not be fair use. I worry that if these activities — indeed, these rights — are restricted for the machine as an agent for users, precedent is set that could restrict use for us all. As a journalist, I fear that by restricting learning sets to viewing only free content, we will end up with a problem parallel to that created by the widespread use of paywalls in news: authoritative, fact-based reporting will be restricted to the privileged few who can and choose to pay for it, leaving too much of public discourse vulnerable to the misinformation, disinformation, and conspiracies available for free, without restriction.

I see another potential use for large language models: to provide researchers and scholars with a window on the presumptions, biases, myths, and misapprehensions reflected in the relationships of all the words analyzed by them — the words of those who had the power and privilege of publishing them. To restrict access skews that vision and potentially harms scholarly uses that have not yet been imagined.

The speculation in **Question 9**, about requiring affirmative permission for any copyrighted material to be used in training AI models, and in **Question 10,** regarding collective management organizations or legislatively establishing a compulsory licensing scheme, frightens me. AI

companies already offer a voluntary opt-out mechanism, in the model of robots.txt. As media [report](), many news organizations are availing themselves of that option. To legally require opt-in or licensing sets up unimaginable complications.

Such complication raises the barrier to entry for new and open-source competitors and the spectre of regulatory capture — as does discussion in the EU of restricting open-source AI models (**Question 25.1**). The best response to the rising power of the already-huge incumbent companies involved in AI is to open the door — not close it — to new competition and open development.

As for **Questions 18-21** on copyrightability, I would suggest a different framework for considering both the input and output of generative AI: as an intellectual, cultural, and informational commons, whose use and benefits we cannot not predict. Shouldn't policy encourage at least a period of development, research, and experimentation?

Finally, permit me to propose another framework for consideration of copyright in this new age in which connected technologies enable collaborative creation and communal distribution. In 2012, I led a series of discussions with multiple stakeholders — media executives, creative artists, policymakers — for a project with the World Economic Forum in Davos on rethinking intellectual property and the support of creativity in the digital age. In the safe space of the mountains, even entertainment executives would concede that copyright law could be considered outmoded and is due for reconsideration. The WEF report is available [here]().

Out of that work, I conceived of a framework I call "creditright," which I write about in *Geeks Bearing Gifts* (CUNY Journalism Press, 2014) and in *The Gutenberg Parenthesis* (221-2): "This is not the right to copy text but the right to receive credit for contributions to a chain of collaborative inspiration, creation, and recommendation of creative work. Creditright would permit the behaviors we want to encourage to be recognized and rewarded. Those behaviors might include inspiring a work, creating that work, remixing it, collaborating in it, performing it, promoting it. The rewards might be payment or merely credit as its own reward. I didn't mention blockchain; but the technology and its automated contracts could be useful to record credit and trigger rewards." I do not pretend that this is a fully thought-through solution, only one idea to spark discussion on alternatives for copyright.

The idea of creditright has some bearing on your **Questions 15-17** on transparency and recordkeeping — what might ledgers of credit in creation look like? — though I am trying to make a larger argument about the underpinnings of copyright. As I have come to learn, 1710's Statute of Anne was not formulated at the urging of — or to protect the rights of — authors, so much as it was in response to the demands of publishers and booksellers, to create a marketplace for creativity as a tradable asset. Said historian Peter Baldwin in *The Copyright Wars: Three*

Centuries of Trans-Atlantic Battle (Princeton University Press, 2016, 53-6): "The booksellers claimed to be supporting authors' just and natural right to property. But in fact their aim was to take for themselves what nature had supposedly granted their clients."

I write in my book that the metaphor of creativity as property — of art as artifact rather than an act — "might be appropriate for land, buildings, ships, and tangible possessions, but is it for such intangibles as creativity, inspiration, information, education, and art? Especially once electronics —from broadcast to digital — eliminated the scarcity of the printed page or the theater seat, one need ask whether property is still a valid metaphor for such a nonrivalrous good as culture."

Around the world, copyright law and doctrine are being mangled to suit the protectionist ends of those lobbying on behalf of incumbent publishers and producers, who remain flummoxed by the challenges and opportunities of technology, of both the internet and now artificial intelligence. In the context of journalism and news, Germany's *Leistungsschutzrecht* or ancillary copyright law, Spain's recently superseded link tax, Australia's News Media Bargaining Code, the proposed Journalism Competition and Preservation Act in the US, and lately Canada's C-18 Online News Act do nothing to protect the public's interest in informed discourse and, in Canada's case, will end up harming news consumers, journalists, and platforms alike as Facebook and Google are forced to take down links to news.

I urge the Copyright Office to continue its process of study as exemplified by this request for comments and not to rush into the frenzied discussion in media over artificial intelligence, large language models, and generative AI. It is too soon. Too little is known. Too much is at stake.