Testimony of Jeff Jarvis before the Senate Judiciary Subcommittee on Privacy, Technology, and the Law. January 10, 2024: Oversight of A.I.: The Future of Journalism

I have been a journalist for fifty years and a journalism professor for the last eighteen.*

I.   History

I would like to begin with three lessons on the history of news and copyright, which I learned researching my book, *The Gutenberg Parenthesis: The Age of Print and its Lessons for the Age of the Internet (Bloomsbury, 2023)*:

First, America's 1790 Copyright Act covered only charts, maps, and books. *The New York Times'* suit against OpenAI claims that, "Since our nation's founding, strong copyright protection has empowered those who gather and report news to secure the fruits of their labor and investment." In truth, newspapers were not covered in the statute until 1909 and even then, according to Will Slauter, author of *Who Owns the News: A History of Copyright* (Stanford, 2019), there was debate over whether to include news articles, for they were the products of the institution more than an author.

Second, the Post Office Act of 1792 allowed newspapers to exchange copies for free, enabling journalists with the literal title of "scissors editor" to copy and reprint each others' articles, with the explicit intent to create a network for news, and with it a nation.

Third, exactly a century ago, when print media faced their first competitor — radio — newspapers were hostile in their reception. Publishers strong-armed broadcasters into signing the 1933 Biltmore Agreement by threatening not to print program listings. The agreement limited radio to two news updates a day, without advertising; required radio to buy their news from newspapers' wire services; and even forbade on-air commentators from discussing any event until twelve hours afterwards — a so-called "hot news doctrine," which the Associated Press has

since tried to resurrect. Newspapers lobbied to keep radio reporters out of the Congressional press galleries. They also lobbied for radio to be regulated, carving an exception to the First Amendment's protections of freedom of expression and the press.

Publishers accused radio — just as they have since accused television and the internet and AI — of stealing "their" content, audience, and revenue, as if each had been granted them by royal privilege. In scholar Gwenyth Jackaway's words, publishers "warned that the values of democracy and the survival of our political system" would be endangered by radio. That sounds much like the sacred rhetoric in *The Times'* OpenAI suit: "Independent journalism is vital to our democracy. It is also increasingly rare and valuable."

To this day, journalists — whether on radio or at *The New York Times* — read, learn from, and repurpose facts and knowledge gained from the work of fellow journalists. Without that assured freedom, newspapers and news on television and radio and online could not function. The real question at hand is whether artificial intelligence should have the same right that journalists and we all have: the right to read, the right to learn, the right to use information once known. If it is deprived of such rights, what might we lose?

II. Opportunities

Rather than dwelling on a battle of old technology and titans versus new, I prefer to focus here on the good that might come from news collaborating with this new technology.

First, though, a caveat: I argue it is irresponsible to use large language models where facts matter, for we know that LLMs have no sense of fact; they only predict words. News companies, including CNET, G/O Media, and Gannett, have misstepped, using the technology to manufacture articles at scale, strewn with errors. I covered the show-cause hearing for a New York attorney who (like President Trump's former counsel, Michael Cohen) used an LLM to list

case citations. Federal District Judge P. Kevin Castel made clear that the problem was not the technology but its misuse by humans. Lawyers and journalists alike must exercise caution in using generative AI to do their work.

Having said that, AI presents many intriguing possibilities for news and media. For example:

AI has proven to be excellent at translation. News organizations could use it to present their news internationally.

Large language models are good at summarizing a limited corpus of text. This is what Google's NotebookLM does, helping writers organize their research.

AI can analyze more text than any one reporter. I brainstormed with an editor about having citizens record 100 school-board meetings so the technology could transcribe them and then answer questions about how many boards are discussing, say, banning books.

I am fascinated with the idea that AI could extend literacy, helping people who are intimidated by writing tell and illustrate their own stories.

A task force of academics from the Modern Language Association concluded AI in the classroom could help students with word play, analyzing writing styles, overcoming writers' block, and stimulating discussion.

AI also enables anyone to write computer code. As an AI executive told me in a podcast about AI that I cohost, "English majors are taking the world back... The hottest programming language on planet Earth right now is English."

Because LLMs are in essence a concordance of all available language online, I hope to see scholars examine them to study society's biases and clichés.

And I see opportunities for publishers to put large language models in front of their content to allow readers to enter into dialog with that content, asking their own questions and creating new subscription benefits. I know an entrepreneur who is building such a business.

Note that in Norway, the country's largest and most prestigious publisher, Schibsted, is leading the way to build a Norwegian-language large language model and is urging all publishers to contribute content. In the US, Aimee Reinhart, an executive student of mine at CUNY who works on AI at the Associated Press, is also studying the possibility of an LLM for the news industry.

III.   Risks

All these opportunities and more are put at risk if we fence off the open internet into private fortresses.

Common Crawl is a foundation that for sixteen years has archived the entire web: 250 billion pages, 10 petabytes of text made available to scholars for free, yielding 10,000 research papers. I am disturbed to learn that *The New York Times* has demanded that the entire history of its content — that which was freely available — be erased. Personally, when I learned that my books were included in the Books3 data set used to train large language models, I was delighted, for I write not only to make money but also to spread ideas.

What happens to our information ecosystem when all authoritative news retreats behind paywalls, available only to privileged citizens and giant corporations able to pay for it? What happens to our democracy when all that is left out in public for free — to inform both citizens and machines — is propaganda, disinformation, conspiracies, spam, and lies? I well understand the economic plight of my industry, for I direct a Center for Entrepreneurial Journalism. But I

also say we must have a discussion about journalism's moral obligation to an informed society and about the right not only to speak but to learn.

IV.   Copyright

And we need to talk about reimaging copyright in this age of change, starting with a discussion about generative AI as fair and transformative use. When the Copyright Office sought opinions on artificial intelligence and copyright (Docket 2023-6), I responded with concern about an idea the Office raised of establishing compulsory licensing schemes for training data. Technology companies already offer simple opt-out mechanisms (see: robots.TXT).

Copyright at its origin in the Statute of Anne of 1710 was enacted not to protect creators, as is commonly asserted. Instead, it was passed at the demand of booksellers and publishers to establish a marketplace for creativity as a tradeable asset. Our concepts of creativity-as-content and content-as-property have their roots in copyright.

Now along come machines — large language models and generative AI — that manufacture endless content. University of Maryland Professor Matthew Kirschenbaum warns of what he calls "the Textpocalypse." Artificial intelligence commodifies the idea of content, even devalues it. I welcome this. For I hope it might drive journalists to understand that their value is not in manufacturing the commodity, content. Instead, they must see journalism as a service to help citizens inform public discourse and improve their communities.

In 2012, I led a series of discussions with multiple stakeholders — media executives, creative artists, policymakers — for a project with the World Economic Forum on rethinking intellectual property and the support of creativity in the digital age. In the safe space of Davos, even media executives would concede that copyright is outmoded. Out of this work, I conceived of a framework I call "creditright," which I've written is "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." It is just one idea, intended to spark discussion.

Publishers constantly try to extend copyright's restrictions in their favor, arguing that platforms owe them the advertising revenue they lost when their customers fled for better, competitive deals online. This began in 2013 with German publishers lobbying for a *Leistungsschutzrecht*, or ancillary copyright, which inspired further protectionist legislation, including Spain's link tax, articles 15 and 17 of the EU's Copyright Directive, Australia's News Media Bargaining Code, and most recently Canada's Bill C-18, which requires large platforms — namely Google and Facebook — to negotiate with publishers for the right to link to their news. To gain an exemption from the law, Google agreed to pay about $75 million to publishers — generous, but hardly enough to save the industry. Meta decided instead to take down links to news rather than being forced to pay to link. That is Meta's right under Canada's Charter of Rights and Freedoms, for compelled speech is not free speech.

In this process, lobbyists for Canada's publishers insisted that their headlines were valuable while Meta's links were not. The nonmarket intervention of C-18 sided with the publishers. But as it turned out, when those links disappeared, Facebook lost no traffic while publishers lost up to a third of theirs. The market spoke: Links are valuable. Legislation to restrict linking would break the internet for all.

I fear that the proposed Journalism Competition and Preservation Act (JCPA) and the California Journalism Protection Act (CJPA) could have similar effect here. As a journalist, I

must say that I am offended to see publishers lobby for protectionist legislation, trading on the political capital earned through journalism. The news should remain independent of — not beholden to — the public officials it covers. I worry that publishers will attempt to extend copyright to their benefit not only with search and social platforms but now with AI companies, disadvantaging new and small competitors in an act of regulatory capture.

V. Support for innovation

The answer for both technology and journalism is to support innovation. That means enabling open-source development, encouraging both AI models and data — such as that offered by Common Crawl — to be shared freely.

Rather than protecting the big, old newspaper chains — many of them now controlled by hedge funds, which will not invest or innovate in news — it is better to nurture new competition. Take, for example, the 450 members of the New Jersey News Commons, which I helped start a decade ago at Montclair State University; and the 475 members of the Local Independent Online News Publishers; the 425 members of the Institute for Nonprofit News; and the 4,000 members of the News Product Alliance, which I also helped start at CUNY. This is where innovation in news is occurring: bottom-up, grass-roots efforts emergent from communities.

There are many movements to rebuild journalism. I helped develop one: a degree program called Engagement Journalism. Others include Solutions Journalism, Constructive Journalism, Reparative Journalism, Dialog Journalism, and Collaborative Journalism. What they share is an ethic of first listening to communities and their needs.

In my upcoming book, *The Web We Weave*, I ask technologists, scholars, media, users, and governments to enter into covenants of mutual obligation for the future of the internet and, by extension, AI.

There I propose that you, as government, promise first to protect the rights of speech and assembly made possible by the internet. Base decisions that affect internet rights on rational proof of harms, not protectionism for threatened industries and not media's moral panic. Do not splinter the internet along national borders. And encourage and enable new competition and openness rather than entrenching incumbent interests through regulatory capture.

In short, I seek a Hippocratic Oath for the internet: First, do no harm.

* Professional biography: Jeff Jarvis holds the Leonard Tow Chair in Journalism Innovation and direct the Tow-Knight Center for Entrepreneurial Journalism at the City University of New York's Craig Newmark Graduate School of Journalism. Previously, he was the president and creative director of Advance.net, online arm of Advance Publications; creator and founding editor of Entertainment Weekly at Time Inc.; Sunday editor and associate publisher of the New York Daily News; TV critic for TV Guide and People; and a columnist, reporter, and editor on the San Francisco Examiner and Chicago Tribune. He is the author of six books: *The Web We Weave: Why We Must Reclaim the Internet from Moguls, Misanthropes* (fall 2024 from Basic Books), *The Gutenberg Parenthesis: The Age of Print and Its Lessons for the Age of the Internet* (Bloomsbury Academic 2023); *Magazine* (Bloomsbury 2023); *Geeks Bearing Gifts: Imagining New Futures for News* (CUNY Journalism Press 2015); *Public Parts* (Simon & Schuster 2011), and *What Would Google Do?* (Harper, 2009). He is cohost of the podcasts This Week in Google and AI Inside.