Exhibit 1

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC. COPYRIGHT INFRINGEMENT LITIGATION This Document Relates to: *All Cases*	Case No.25-md-3143 (SHS) (OTW)

[PROPOSED] BRIEF OF AMICUS CURIAE
GREGORY LEPPERT IN SUPPORT OF DEFENDANTS

Matthew S. Warren
Erika H. Warren
WARREN LLP
2261 Market Street, No. 606
San Francisco, California, 94114
+1 (415) 895-2940
+1 (415) 895-2964 facsimile
25-3143@cases.warrenllp.com

Attorneys for [Proposed] Amicus Curiae Gregory Leppert

TABLE OF CONTENTS

TABLE OF AUTHORITIES ii

STATEMENT OF IDENTITY AND INTEREST OF AMICUS CURIAE 1

INTRODUCTION 2

ARGUMENT 3

I. Libraries Are Essential to American Expression 3

II. Libraries Face a Critical Time in American History 5

III. Libraries Rely on Robust General-Purpose Models 7

A. Libraries Struggle to Meet Their Goals of Preservation and Access 7

B. AI Supports Preservation and Access At Scale 10

IV. Still, Consider the Library 14

CONCLUSION 15

TABLE OF AUTHORITIES

Cases *Page*

Authors Guild v. Google, Inc.,
804 F.3d 202 (2d Cir. 2015) 3, 14

Google LLC v. Oracle America, Inc.,
593 U.S. 1 (2021) 3, 14

Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.,
499 U.S. 340 (1991) 3

Statutes and Rules

U.S. Const. art. I, § 8, cl. 8 3

STATEMENT OF IDENTITY AND INTEREST OF AMICUS CURIAE

Gregory Leppert is the Executive Director of the Institutional Data Initiative (IDI) at the Harvard Law School Library, a research initiative that works with libraries, universities, and governments to refine and publish their collections as datasets for computational uses. His work at IDI is animated by a desire to affirm the role of public-interest institutions as stewards of knowledge in a future shaped by AI. IDI is supported by a consortium of funders, including Jane Street, Meta, Microsoft, and OpenAI.

Mr. Leppert is also the Chief Technologist of the Berkman Klein Center for Internet and Society at Harvard University, where he enables research on transformative technologies as they intersect with society. He was previously the Director of Product and Research at the Harvard Law School Library's Library Innovation Lab (LIL), where he led strategy and implementation of tools and research for improving library technologies.

Through his work on LIL's Caselaw Access Project, Mr. Leppert helped the Harvard Law School Library as it digitized and made accessible 360 years of U.S. case law, creating a first-of-its kind dataset that has democratized access to law. Through LIL's Perma.cc platform, Mr. Leppert helped a consortium of libraries archive the web so that citations to web content—especially those in legal filings such as this one—are properly preserved, even if the original webpages change or disappear.

In leading IDI, Mr. Leppert has worked with the Boston Public Library to build AI systems that parse and make accessible 1.47 million pages from the library's newspaper collections; with the University of Michigan Library to use general-purpose AI reasoning models to extract key decision points from nearly 200 years of the university's board proceedings; and

with Harvard Library and Google to release nearly one million books scanned from Harvard Library's collections.

As a lifelong artist and musician who once made his living as a graphic designer and full-time touring musician, Mr. Leppert has also personally relied on the markets and creative freedoms that copyright and fair use enable. Together, these experiences give him a direct interest in how copyright law can continue to support creative markets while enabling libraries and other public-interest institutions to use new technologies to preserve and expand access to knowledge.

Mr. Leppert submits this brief solely in his individual capacity, and provides his institutional affiliations for identification only. No party or its counsel authored this brief in whole or in part. No party or its counsel, and no person other than Mr. Leppert and his counsel, contributed money to fund or intended to fund the preparation or submission of this amicus curiae brief.

INTRODUCTION

This case involves training of general-purpose AI models on a vast corpus of data, including copyrighted works. Plaintiffs do not argue that the sole use of these models is infringement of their intellectual property. Instead, Plaintiffs argue that, because the models can be used to prepare other works and because those works could be infringing or similar to Plaintiffs' works, or otherwise compete with them, the general-purpose models comprise "machine[s] *capable* of flooding the market with competing works."[1] Mr. Leppert takes no position on whether any particular AI output would infringe any copyrighted work, or would be fair use. Instead, Mr. Leppert submits that general-purpose models like those at issue in this

[1] Dkt. No. 1982 at 27-28 (emphasis added); see also, e.g., Dkt. No. 2108 at 65 ("AI-generated outputs provide substitutive answers.")

action—due in large part to their broad training data—are capable of many uses that serve important public benefits which promote the fundamental goals of copyright.

Libraries, tasked as stewards of human knowledge and understanding, benefit from the availability of high quality general-purpose models in performing their duty to organize and preserve information. Libraries generally lack the resources to develop their own underlying models—indeed many struggle to staff basic functions, let alone hire specialized software developers—and yet they face a growing backlog of library materials to be processed and made accessible for future generations. General-purpose AI models, trained on a vast, beautifully variegated corpus of data, offer progress towards a future with libraries in it. Mr. Leppert respectfully submits that the Court should consider this among the relevant public benefits in applying the fair use factors.

ARGUMENT

I. Libraries Are Essential to American Expression

"[W]hile authors are undoubtedly important intended beneficiaries of copyright, the ultimate, primary intended beneficiary is the public, whose access to knowledge copyright seeks to advance by providing rewards for authorship."[2] Courts should "take into account the public benefits the copying will likely produce" including whether "those benefits, for example, relate[] to copyright's concern for the creative production of new expression."[3] "[T]he primary objective of copyright is not to reward the labor of authors, but to promote the Progress of Science and useful Arts."[4] Libraries[5] share this fundamental purpose of the Copyright Act. For example, the

[2] *Authors Guild v. Google, Inc*., 804 F.3d 202, 212 (2d Cir. 2015).
[3] *Google LLC v. Oracle America, Inc*., 593 U.S. 1, 35 (2021).
[4] *Feist Publications, Inc. v. Rural Tel. Serv. Co*., 499 U.S. 340, 349 (1991); U.S. Const. art. I, § 8, cl. 8.
[5] Mr. Leppert is aware that courts and parties across the country may take varied views on the meaning or scope of the term "library." As used in this brief, "library" refers to institutions that serve their communities through lending, promotion of literacy, and archival and preservation efforts as contemplated by 17 U.S.C. § 108.

United States Library of Congress states its mission is to provide "access to a rich, diverse and enduring source of knowledge to inform, inspire and engage you in your intellectual and creative endeavors."[6] Across other institutions, the language is remarkably consistent: Harvard Library's mission[7] is to "engage with our communities in the creation and sharing of new knowledge," MIT Libraries[8] seek to "advance knowledge by providing a trusted foundation for the generation, dissemination, use, creative engagement with, and preservation of information" and the New York Public Library[9] exists to "advance knowledge by providing free and open access to materials." The open availability of general-purpose AI models serves and supports an important public benefit in that it enables libraries to keep pace with the preservation and accessibility of human knowledge for future generations.

Libraries have formed part of American society since the time of its founding. What began as small, often private collections evolved into subscription-based lending institutions and, eventually, publicly funded institutions serving the general public.[10] One purpose of libraries is to support society's need for a common and diverse foundation of knowledge. The Founders spoke of the need for common knowledge as intuitive and innate. In his first address to a newly formed Congress, George Washington declared: "Knowledge is in every Country the surest basis

[6] *About the Library Overview*, Library of Congress, https://www.loc.gov/about/ [https://perma.cc/B394-2F76] (last visited Oct. 29, 2026).

[7] *About Harvard Library*, Harvard Library, https://library.harvard.edu/about [https://perma.cc/V7KU-RWKL] (last visited Oct. 29, 2026).

[8] *About us*, MIT Libraries, https://libraries.mit.edu/about/vision/vision-mission-values/ [https://perma.cc/QAU8-S9GA] (last visited Oct. 29, 2026).

[9] *NYPL's Mission Statement*, New York Public Library, https://www.nypl.org/about/mission [https://perma.cc/J3MA-2Q27] (last visited Oct. 29, 2026).

[10] Peggy Johnson, *Fundamentals of Collection Development and Management* 8-9 (4th ed. 2018), https://alastore.ala.org/sites/default/files/book_samples/9780838916414_sample.pdf [https://perma.cc/BH7Y-AAU7] (last visited Oct. 30, 2026).

of public happiness," as he advocated for government support of knowledge institutions.[11] John Adams went a step further, writing in the Massachusetts Constitution that "wisdom, and knowledge, as well as virtue, diffused generally among the body of the people" are "necessary for the preservation of their rights and liberties."[12] James Madison was even more blunt: "A popular Government, without popular information, or the means of acquiring it, is but a Prologue to a Farce or a Tragedy; or, perhaps both."[13]

II. Libraries Face a Critical Time in American History

Libraries provide an epistemic bedrock for the American experiment by preserving and making accessible what we know, and how we know it. This foundation affords us the opportunity to ground our national discourse and collective decision-making in a common knowledge: a shared inheritance of information, accessible regardless of stature or station, that grows with inquiry and experience. It fosters connections between ideas and experiences, expressed across languages and cultures, allowing us to discover that we have more in common than we might initially know.

Yet there are signs that our common knowledge may be faltering. In a pair of 2025 Pew Research Center surveys, eight in ten adults said Republican and Democratic voters cannot agree on basic facts,[14] and more than half said they find it difficult to determine what is true when

[11] George Washington, *George Washington to the United States Senate and House of Representatives* (Jan. 8, 1790), Founders Online, Nat'l Archives, https://founders.archives.gov/documents/Washington/05-04-02-0361 [https://perma.cc/JF58-9VNS] (last visited Oct. 30, 2026).

[12] Mass. Const. pt. 2, ch. V, § II, https://malegislature.gov/Laws/Constitution [https://perma.cc/6J9Z-44TT] (last visited Oct. 30, 2026).

[13] James Madison, *Letter from James Madison to W. T. Barry* (Aug. 4, 1822), in 9 The Writings of James Madison 103 (Gaillard Hunt ed., 1900-1910), https://press-pubs.uchicago.edu/founders/documents/v1ch18s35.html [https://perma.cc/U2AQ-NLEN].

[14] Elisa Shearer, *Most Americans say Republican and Democratic voters cannot agree on basic facts*, Pew Research Center (July 30, 2025), https://www.pewresearch.org/short-reads/2025/07/30/most-americans-say-republican-and-democratic-voters-cannot-agree-on-basic-facts/ [https://perma.cc/D5AD-9MQJ].

getting their news.[15] This divide extends to how Americans understand their own history. In 2024, 79% of Biden supporters said the legacy of slavery still affects the position of Black Americans today, compared with just 27% of Trump supporters.[16] The 1619 Project[17] and 1776 Commission,[18] two sharply different accounts of our nation's origins, reflect these differences. Through the 1619 Project, the New York Times Magazine "aim[ed] to reframe the country's history" around "the consequences of slavery and the contributions of black Americans." The Trump administration responded with the 1776 Commission, meant to counteract "scholarship [that] has vilified our Founders and our founding" and instead promote "a shared identity rooted in our founding principles."[19] The narratives of each project were ostensibly woven from the historical record, against which anyone should be able to test the projects' interpretations. Libraries have preserved this historical record for centuries, and continue to provide materials by which anyone can research these competing narratives and compare them to the primary source material. A robust and accessible historical record permits Americans to develop their own understanding, and to participate in public debate.

An emerging crisis of authenticity is further compounding our divisions. The rise of smartphones with powerful cameras, widely accessible editing tools, and social media platforms spurred calls of "fake news" from all corners. The wide availability of AI has further blurred the line between what content is real and what is imagined. While libraries stop short of arbitrating

---

[15] Luxuan Wang & Joanne Haner, *Many Americans say they often come across inaccurate news – and have a hard time knowing what's true*, Pew Research Center (Oct. 29, 2025), https://www.pewresearch.org/short-reads/2025/10/29/many-americans-say-they-often-come-across-inaccurate-news-and-have-a-hard-time-knowing-whats-true/ [https://perma.cc/5REX-TBET].

[16] *Cultural Issues and the 2024 Election*, Pew Research Center (June 6, 2024), https://www.pewresearch.org/politics/2024/06/06/cultural-issues-and-the-2024-election/ [https://perma.cc/9E2R-UAXT].

[17] Nikole Hannah-Jones, *The 1619 Project.*, The N.Y. Times Mag. (Aug. 14, 2019), https://www.nytimes.com/interactive/2019/08/14/magazine/1619-america-slavery.html (last visited Oct. 30, 2026).

[18] Exec. Order No. 13958, 85 Fed. Reg. 70951 (Nov. 2, 2020).

[19] *Id.*

what is true (and many materials within libraries contain untruths), they do strive to preserve the evidence needed to establish authenticity in the form of provenance chains that document the origins, ownership, and custodial histories of materials. These provenance chains help society—individually and collectively—decide how materials should weigh on truth, and allow us to build and create new ideas upon their contents. Without this provenance, the foundations of knowledge risk becoming mired in hallucinations and eroded by uncertainty until they can no longer support the weight of new inquiry.

III. Libraries Rely on Robust General-Purpose Models

A. Libraries Struggle to Meet Their Goals of Preservation and Access

For nearly four centuries, libraries in North America have studiously carried out their missions, collecting and preserving an ever-expanding landscape of knowledge. Each year, some 9,000 public library systems and their branches[20]—the classic neighborhood libraries of Americana—account for over two billion circulations, serving more than 150 million registered users and welcoming countless others who visit their halls and websites for other services.[21] Adding our academic, government, and other libraries, the total reaches 120,000 libraries, hundreds of millions of patrons, and collections numbering in the billions.[22]

At the same time that the historical record has become a cultural battleground, we have underinvested in preserving it. As Mr. Leppert has learned firsthand, the need for libraries' services far outstrips their operational capacity. Libraries are working tirelessly to preserve and

[20] *Public Libraries Survey (PLS)*, Institute of Museum and Library Services, https://www.imls.gov/research-evaluation/surveys/public-libraries-survey-pls [https://perma.cc/B7ZE-78BK] (last visited Oct. 29, 2025).

[21] Marisa Pelczar et al., *Public Libraries Survey Fiscal Year 2024: Data File Documentation and User's Guide*, Institute of Museum and Library Services (June 2026), https://www.imls.gov/sites/default/files/2026-06/PublicLibrariesSurvey_FiscalYear2024_DataDocumentationandUsersGuide.pdf [https://perma.cc/P3GQ-M8QQ].

[22] *Statistics -- Library*, American Library Association, https://www.ala.org/tools/atoz/statistics-library [https://perma.cc/B7R9-7JSQ] (last visited Oct. 30, 2026).

categorize their collections for future generations. This work includes digitizing physical collections to preserve their contents and extend public access beyond the library walls. Libraries across the nation are racing against time to scan the pages of books and newspapers, photograph artifacts, and 3D-scan other materials. Based on Mr. Leppert's work, he estimates that, at current rates, it will take libraries hundreds of years to digitize their existing materials, let alone new items added in the meantime. Over such a long horizon, materials will decay before they can be digitally preserved, and generations will pass without access to the knowledge they contain. We can only estimate because no one truly knows the full extent of library holdings or the wealth of knowledge that remains just out of reach—not even the libraries themselves. By the time digital databases arrived in the 1960s, libraries had been collecting materials for hundreds of years and cataloging them using physical systems, most famously the card catalog. Millions of these cards are still filed away in cabinets, waiting to be converted into machine-readable records before a full account can be made.

Cataloging backlogs continue to grow as many libraries struggle to process the steady stream of new items at the rate they arrive. In the case of large donations, where, for instance, a prominent figure bequeaths their papers, years can pass before the materials are fully indexed and described, and become generally findable within a library's systems. The result is vast swaths of knowledge and history that remain effectively invisible to humans and machines alike. Libraries are acutely aware of this, which is one reason they prioritize preservation. Preservation not only extends the usable life of materials, it buys time for society to invest in the work of extending access to them through processes like digitization. Libraries prioritize preservation because they can only do so much: by preserving as many materials as possible, they maximize the chance that new tools will be developed that reduce the cost of increasing access, and new technologies

will provide incentives to invest in creating and maintaining it. The Internet is one of those technologies. The ability to easily exchange digital information with parties across the world changed the nature and value of information access, leading Google to create Google Books, the single largest library digitization effort to date. Since 2004, the Google Books project has scanned tens of millions of books from library collections, indexing their contents to make them searchable while providing digital copies back to the libraries for their use and preservation.[23] This led to the creation of HathiTrust, a consortium of more than 290 libraries that make their digitized collections available for browsing, citation, and computational research, ushering in a new era for digital humanities research.[24]

Preservation is but one facet of the mission of libraries: their goal is not merely to hold knowledge but to animate and facilitate its use. These goals face significant barriers, and the barriers grow with scale. Access to knowledge was once limited by the cost and onerous process of physically acquiring materials; Harvard Library, the oldest library system in the U.S., began when John Harvard bequeathed his personal library of 400 volumes to the college that came to bear his name.[25] But the availability of materials has steadily increased, revealing new limits in the form of practical access: how can library patrons effectively search for and discover information when Harvard Library now contains not 400, but over 400 million physical items?[26]

---

[23] Haimin Lee, *15 years of Google Books*, Google (Oct. 17, 2019), https://blog.google/products-and-platforms/products/search/15-years-google-books/ [https://perma.cc/RL7Q-GPB5].

[24] Kevin O'Brien, *HathiTrust*, 99 J. Med. Libr. Ass'n 177, 177-78 (2011), https://pmc.ncbi.nlm.nih.gov/articles/PMC3066577/

[25] *About Harvard Library*, Harvard Library, https://library.harvard.edu/about [https://perma.cc/V7KU-RWKL] (last visited Oct. 29, 2026); Jennifer Tomase, *Tale of John Harvard's surviving book*, Harv. Gazette (Nov. 1, 2007), https://news.harvard.edu/gazette/story/2007/11/tale-of-john-harvards-surviving-book/ [https://perma.cc/U7MS-XPYK].

[26] Stephen Abrams, *2025 World Digital Preservation Day: Why Preserve?*, Harvard Library (Nov. 6, 2025), https://preservation.library.harvard.edu/news/2025/11/2025-world-digital-preservation-day-why-preserve [https://perma.cc/WY3L-GL34].

The Library of Congress—arguably the world's largest—adds 10,000 items every working day.[27] The need to catalog digital media has only accelerated this growth. U.S. academic libraries now provide access to nearly 4.5 billion digital and electronic collection items[28] in addition to web archiving efforts, such as the Library of Congress's 5.7 petabyte archive.[29] The number of American library patrons has also grown enormously, from an estimated 26 million registered public library borrowers in the 1930s to more than 150 million in 2023—from about 20% of the population to nearly half.[30] What began as hand-curated collections of neatly packed books has thus grown into towering piles of media, physical and virtual, climbing higher each day. The need to catalog collections, and quickly, grows more urgent.

B. AI Supports Preservation and Access At Scale

Libraries' relevance depends upon their ability to catalog diverse information at the scale and speed at which society produces it. It is no surprise that libraries are turning to AI to help organize, preserve, and provide access to their collections. AI is capable of revealing both explicit and latent relationships within immense volumes of information—connecting words to concepts to ideas to people—in ways that were not possible even five years ago. It can read decaying text from ancient manuscripts and correlate it with sentiments expressed in modern prose. It can bridge languages to find commonalities between seemingly disparate worldviews, and trace the historical record to allow deep inquiry into the origins of our ideologies.

---

[27] *Fascinating Facts*, Library of Congress, https://www.loc.gov/about/fascinating-facts [https://perma.cc/2YA2-U6TY] (last visited Oct. 29, 2026).

[28] Nat'l Ctr. for Educ. Statistics, *Table 8. Number of items in library collections and the number of times items are in circulation at Title IV degree-granting institutions, by level and control of institution, and type of library collection or circulation: United States, fiscal year 2023*, Integrated Postsecondary Education Data System (2023), https://nces.ed.gov/ipeds/search/viewtable?query=library&returnUrl=%2Fsearch&tableId=36549.

[29] Abbie Grotke et al., *Web Archiving at the Library After 25 Years*, Library of Congress Blogs (Jan. 12, 2026), https://blogs.loc.gov/thesignal/2026/01/web-archiving-at-the-library-after-25-years/ [https://perma.cc/7W8L-R37W].

[30] Carleton Bruns Joeckel, *Library Service*, U.S. Gov't Printing Office (1938), https://books.google.com/books/about/Library_service.html?id=OZohAAAAMAAJ [https://perma.cc/4U82-5V8G]

Beyond mere passive consumers of AI, libraries are active participants who are building and training their own tools on the foundation of general-purpose models. For example, the Boston Public Library and the Institutional Data Initiative at the Harvard Law School Library collaborated to train custom AI models, built on general-purpose base models, to turn historical newspaper scans into structured data by segmenting complex page layouts, extracting text, recovering reading order, and classifying individual page elements.[31] The University of Texas at Austin Libraries developed a workflow in which general-purpose large language models read images of CDs and LPs, extract bibliographic metadata, and prepare catalog records for librarian review.[32] The Theodore Roosevelt Presidential Library developed an AI system using general-purpose models to allow patrons to converse with a collection of roughly 300,000 archival records rather than simply browse or search them.[33] And researchers at Purdue University Libraries used general-purpose language models to transform more than 45,000 pages of historical U.S. Department of Agriculture plant records into a structured dataset for computational humanities research.[34]

General-purpose models are uniquely suited to support the missions of libraries. Because the predictions that AI models make are informed by what they encountered in their training

[31] Matteo Cargnelutti et al., *Institutional Newspapers Pipeline: Deriving Billions of High Quality Token from Historical Newspapers*, arXiv (Aug. 19, 2026), https://arxiv.org/abs/2608.18972 [https://perma.cc/QXH6-GPQ9].

[32] Hannah Moutran, *Artificial Intelligence-Assisted Music Cataloging Project*, Coalition for Networked Information (Jan. 14, 2026), https://www.cni.org/topics/information-access-retrieval/artificial-intelligence-assisted-music-cataloging-project [https://perma.cc/BLX8-9U3K].

[33] Pengce Wange et al., *The Living Library: Transforming Archival Collections into Conversational Knowledge Systems – Lessons from the Theodore Roosevelt Presidential Library*, arXiv (Sept. 8, 2026), https://arxiv.org/abs/2609.09368 [https://perma.cc/N9F4-5ZUS].

[34] Spencer Dean Stewart & Sankriti Sinha, *Retrieving Information from Unstructured Historical Sources Using Large Language Models*, 1 Computational Humanities Research 17 (2025), https://www.cambridge.org/core/journals/computational-humanities-research/article/retrieving-information-from-unstructured-historical-sources-using-large-language-models/845FDD146FF8A9283A32BFF0C3160F5F [https://perma.cc/B53L-BNUD].

data, models trained on a broader and more diverse range of inputs are better positioned to navigate the varied materials in library collections.[35] Many such uses are functional, and distant from any particular expression. For example, the layouts of newspapers have changed drastically over time. In 1769, paper was precious and the newspapers of that era were typeset by hand to within an inch of their lives, with narrow gutters between dense columns of text and relatively few illustrations to interrupt them.[36] By the 1880s, illustrations were commonplace, and newspapers had just begun to reproduce photographs, leading text to flow around increasingly intricate layouts and trace more complex paths around the page for the reader's eye to follow. 100 years later, paper was abundant and inexpensive, and page design was increasingly computerized, resulting in newspapers with luxuriously wide columns, large color photos pushing articles across pages, and elements of modern magazine design working their way into the daily news.[37] The limitations of a technology define a medium, and as printing evolved to remove constraints and enable creativity, so too did the visual language of newspapers.





[35] Shayne Longpre et al., *A Pretainer's Guide to Training Data: Measuring the Effects of Data Age, Domain Coverage, Quality, & Toxicity*, arXiv (May 22, 2023), https://arxiv.org/abs/2305.13169 [https://perma.cc/F5YP-U2VJ].

[36] Image of the Boston Chronicle, (July 31 - Aug. 3, 1769), Mass. Hist. Society, Digital Commonwealth, https://www.digitalcommonwealth.org/search/commonwealth-oai:mc87r695r [https://perma.cc/QW47-NNVK].

[37] Image of USA Today Cover Page (Feb. 5, 2009), Wikipedia, https://en.wikipedia.org/wiki/USA_Today [https://perma.cc/MQ69-A64V].

These differences, taken in aggregate, and not the content of any single article or page, significantly impact a model's effectiveness for library tasks.[38] A model trained only on data from newspapers published during the public domain era is thus unlikely to perform as well on tasks involving modern newspapers, such as recognizing certain words in their modern spellings or parsing the contents of pages to make them accessible for full-text search and screen-readers. If asked to parse and preserve a single article from a 2002 small-circulation newspaper, a model trained on a narrow window of early newspapers would be less likely to make correct predictions about how text flows from page to page, which images are connected to which text, and where one article ends and the next begins. In contrast, a general-purpose model which encountered a breadth of publications that span eras and the design language within them is more likely to do well on tasks involving newspapers from both the 1800s and the 2000s.

Even where libraries train their own AI models to perform library-specific functions, they rarely if ever develop any model from scratch due to the immense computational resources required to reach the capabilities of existing models, the cost of which regularly reaches beyond one hundred million dollars.[39] Instead, they typically draw upon an ecosystem of openly available models that have already been trained on a wide range of materials and thereby serve a wide range of needs. Libraries then extend the training of these models to tailor them to a specific task. Because of this, libraries depend as much upon the ability of developers in the broader model ecosystem to train on diverse materials as they do upon their own ability to augment models with additional training. Without a broader model ecosystem in which the cost

[38] Catherine Chen et al., *Are Layout-Infused Language Models Robust to Layout Distribution Shifts? A Case Study with Scientific Documents*, arXiv (June 1, 2023), https://arxiv.org/abs/2306.01058 [https://perma.cc/9ZFG-5BRY].

[39] Nestor Maslej et al., *The AI Index 2025 Annual Report*, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University (2025), https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf [https://perma.cc/A4SM-Q7B7].

of training capable base models can be shared rather than duplicated, most libraries would not be able to create models to suit their needs. The ability of libraries to use these tools therefore depends not only on access to AI models, but on whether those models can be trained on the broad swath of materials libraries serve.

A robust ecosystem of general-purpose models is thus integral to libraries' needs; limiting model training to a subset of materials—curated perhaps by cost, or commercial importance, or the practical availability of a licensing contract—would reduce the effectiveness of those models in preserving and making available materials outside of that distribution. Stretches of our shared cultural history could appear as dark ages in the computationally accessible historic record, excluded from the accessibility advancements and connected knowledge fabric that AI models enable. Precision could be limited to old materials whose copyrights have ended; newer content—especially physical media living in the era between the end of the public-domain and the beginning of digital media—could become as challenging to catalog and surface as the ornate illuminated manuscripts of yore. Providing Plaintiffs the absolute control they seek would serve to "limit, rather than expand, public knowledge."[40]

IV. Still, Consider the Library

Though some may argue that fears of library obsolescence are misguided, or overblown, or even purely hypothetical, so too are the horribles paraded by the Plaintiffs in this action. For example, Plaintiffs suggest that the training of general-purpose models threatens the market for their works because a motivated user could prompt a general-purpose model to regurgitate the whole of a particular work or fiction or old news story.[41] But we should not overlook the

[40] *Author's Guild,* 804 F.3d at 212; *see also Google v. Oracle*, 593 U.S. 1, 39 (2021) (permitting an absolute monopoly over copyrighted work would "interfere with, not further, copyright's basic creativity objectives.")

[41] *See*, *e.g.*, Dkt. No. 1982 at 13, Dkt. No. 2108 at 65.

existence of libraries, from which members of the public are free to enjoy published works such as the Plaintiffs' books and yesterday's newspapers in their entirety, if only libraries make it possible to find them among the stacks. More than 150 million Americans hold public library cards.[42] Those cards enable Americans to borrow and consume complete works, free of charge, from which to learn, explore, and develop a common knowledge. By contrast, cobbling together an infringing work through prompting a general-purpose AI model would likely take enormous effort, and could incur substantial cost to the user. In short, Americans who wish to read Game of Thrones do not have to devote hours to piecing it back together using AI. They need only visit their local library.

CONCLUSION

Mr. Leppert hopes that the Court will find this discussion helpful as it considers how to resolve the pending motions in this action.

Date: October 30, 2026

Respectfully submitted,

Matthew S. Warren (State Bar No. 4016275)
Erika H. Warren (*pro hac vice* pending)
WARREN LLP
2261 Market Street, No. 606
San Francisco, California, 94114
+1 (415) 895-2940
+1 (415) 895-2964 facsimile
25-3143@cases.warrenllp.com

Attorneys for [Proposed] Amicus Curiae Greg Leppert

[42] *Public Libraries Survey: Fiscal Year 2024 Data Documentation and User's Guide*, Institute of Museum and Library Services (2026), https://www.imls.gov/sites/default/files/2026-06/PublicLibrariesSurvey_FiscalYear2024_DataDocumentationandUsersGuide.pdf [https://perma.cc/TE6E-E6GH]

CERTIFICATE OF COMPLIANCE

Under Local Civil Rule 7.1(c), the above-named counsel certifies that this brief complies with the word-count limitation set by the Court's Order of August 7, 2026 (Dkt. No. 1663). As calculated by the word-count function of the word-processing program used to prepare this brief, the brief contains 4,491 words, excluding the caption, table of contents, table of authorities, signature block, and this certificate.