EXHIBIT A

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION This document Relates to: *All Cases*	No.: 1:25-md-03143 (SHS) (OTW) Hon. Sidney H. Stein

PROPOSED BRIEF OF AMICI CURIAE AI SAFETY ASIA, BRIAN J. A. BOYD, BUDDHISM & AI INITIATIVE, CARA PROJECT, INC., CHARLES CAMOSY, DAVID ATKINSON, EDTECH LAW CENTER PLLC, FUTURE OF LIFE INSTITUTE, IRREPLACEABLE, JOHN C. HAVENS, NOĒSIS COLLABORATIVE, PROGRESSIVE DEMOCRATS OF AMERICA, RADICALXCHANGE FOUNDATION, THE TECH OVERSIGHT PROJECT, WALTER SCHEIRER, AND YOUNG PEOPLE'S ALLIANCE IN SUPPORT OF PLAINTIFFS MOTION FOR SUMMARY JUDGMENT

PROPOSED AMICI CURIAE

AI Safety Asia, *Non-profit organization*

Brian J. A. Boyd, *Affiliated Scholar, Institute for Advanced Catholic Studies, Univ. of Southern California*

Buddhism & AI Initiative, *Volunteer project and collaborative network*

Cara Project, Inc.

Charles Camosy, *Professor, Catholic Univ. of America*

David Atkinson, *Fellow, Georgetown Univ. Law Center*

EdTech Law Center PLLC, *Consumer protection law firm*

Future of Life Institute, *Non-profit organization*

Irreplaceable, *Non-profit organization*

John C. Havens, *Author, Heartificial Intelligence: Embracing Our Humanity to Maximize Machines*

Noēsis Collaborative, *Non-profit organization*

Progressive Democrats of America, *Non-profit organization*

RadicalxChange Foundation, *Non-profit organization*

The Tech Oversight Project, *Non-profit organization*

Walter Scheirer, *Professor of Computer Science and Engineering, Univ. of Notre Dame*

Young People's Alliance, *Non-profit organization*

CORPORATE DISCLOSURE STATEMENT

The proposed Amici Curiae are individuals, non-profit organizations, and privately held companies, none of which have a parent corporation, and no publicly held corporation owns 10% or more of any of their stock.

STATEMENT REGARDING AUTHORSHIP AND FUNDING

No party's counsel authored this brief in whole or in part. No party, party's counsel, or other person, other than the amici and their counsel, have borne the costs of this brief.

TABLE OF CONTENTS

Page

PROPOSED AMICI CURIAE……..i

CORPORATE DISCLOSURE STATEMENT……ii

STATEMENT REGARDING AUTHORSHIP AND FUNDING……...ii

TABLE OF AUTHORITIES……v

INTEREST OF AMICI CURIAE…… 1

SUMMARY OF ARGUMENT……1

ARGUMENT……4

I. Defendants' use of copyrighted works does not satisfy the purpose Of copyright law and does not justify use……4

II. If Defendants' use of Plaintiffs' works is fair use, then so must Virtually all unauthorized and uncompensated uses by humans…….7

A. AI systems learning from patterns is not relevantly Different from how humans learn……9

B. As with humans, AI models cannot obtain facts without Also consuming the expression of works……..10

C. Humans, like AI systems, usually generate novel outputs And rarely say or write anything that matches what we've Seen, heard, and read verbatim……12

D. The inefficient learning of AI systems should not excuse further infringement……..13

III. Defendants' differences only reinforce that they should not be granted Fair use……..13

IV. There is no implied license to reproduce Plaintiffs' works ..……...14

A. Publicly available is not the same as public domain...…….15

B. Defendants' ravenous scraping is unlike traditional search Engines and undermines the internet's economics..……16

V. If the court affirms GenAI exceptionalism, it will undermine progress in science and the useful arts……..18

CONCLUSION……..18

TABLE OF AUTHORITIES

Page

Cases

ABKCO Music, Inc. v. Sagan (2d Cir. 2022) (S.D.N.Y. 2012)…….. 15

American Geophysical Union v. Texaco, Inc…….. 11

Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014)…….8

Carreyrou v. Anthropic PBC, No. 5:25-cv-10897-PCP (N.D. Cal. Feb. 26, 2026)……16

Kadrey v. Meta Platforms, Inc., No. 23-cv-03417-VC (N.D. Cal. Jun. 25, 2025) 788 F. Supp. 3d at 1050……3, 8, 9, 13

In Re: OpenAI, Inc., Copyright Infringement Litigation, No. 1:25-md-03143-SHS-OTW (S.D.N.Y. Sep. 4, 2026) …….. *passim*

Psihoyos v. Pearson Educ., Inc., 855 F. Supp. 2d 103……15

Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417 (1984)……..4

Publications

David Atkinson, *Unfair Learning: GenAI Exceptionalism and Copyright Law*, 25 Chi.-Kent J. Intell. Prop. 43 (2026)……17

Sam Altman, *The Gentle Singularity*, SAM ALTMAN BLOG (Jun. 10, 2025, 5:12 PM), https://blog.samaltman.com/the-gentle-singularity. ……4

Geoff Blumfiel, *AI solved one of math's hardest problems. Humanity learned nothing (so far)*, NPR (Sep. 22, 2026, 7:08 AM), https://www.npr.org/2026/09/22/nx-s1-5968588/openai-navier-stokes-problem-mathematicians-learn-little.....5

Tom B. Brown et al., *Language Models are Few-Shot Learners*, ARXIV (Jul. 22, 2020), https://arxiv.org/pdf/2005.14165.....13

Jean-Stanislas Denain, Anson Ho, & Jamie Sevilla, *How many digital workers could OpenAI deploy?*, EPOCH AI (Oct. 3, 2025), https://epoch.ai/gradient-updates/how-many-digital-workers-could-openai-deploy.....14

Jim Edwards, *Cloudflare CEO says Google is abusing its monopoly in search to feed its AI*, FORTUNE (Nov. 13, 2025, 4:32 AM), https://fortune.com/2025/11/13/cloudflare-ceo-google-abusing-monopoly-search-ai/.....17

Suchin Gururangan, et al., *Whose Language Counts as High Quality? Measuring Language Ideologies in Text Data Selection*, in EMNLP 2022, THE 2022 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, PROCEEDINGS OF THE CONFERENCE, 2562 (Dec. 2022), https://aclanthology.org/2022.emnlp-main.165.pdf 11

Tiffany Hsu & Stuart A. Thompson, *Disinformation Researchers Raise Alarms About A.I. Chatbots*, NEW YORK TIMES (Feb. 8, 2023, 3:21 PM), https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html.....7

Cade Metz, *OpenAI Says It Has Cracked One of Math's 'Millennium Problems'*, NEW YORK TIMES (Sep 8, 2026, 5:39 PM), https://www.nytimes.com/2026/09/08/science/openai-proof-millennium-problem.html 5

Cade Metz, *The Liftoff Scenario That Terrifies A.I. Doomsayers*, NEW YORK TIMES, (Sep 16, 2024), https://www.nytimes.com/2026/09/16/science/ai-recursive-self-improvement.html.....4

Rachel Metz, Brody Ford, & Madlin Mekelburg, *Microsoft Spent Over $100 Billion on OpenAI Partnership*, BLOOMBERG (May 13, 2026, 1:15 PM), https://www.bloomberg.com/news/articles/2026-05-13/microsoft-spent-over-100-billion-on-openai-partnership-to-date.....5

Ruchika Joshi & Michal Luria, *AI-Powered Deception: A Deeper Dimension of Dark Design Patterns in Conversational AI Tools and Platforms*, CENTER FOR DEMOCRACY & TECH (SEP. 15, 2025). https://cdt.org/insights/ai-powered-deception-a-deeper-dimension-of-dark-design-patterns-in-conversational-ai-tools-and-platforms/ 7

David Klepper & Ali Swenson, *AI-generated disinformation poses threat of misleading voters in 2024 election*, PBS (May 14, 2024, 7:52 PM), https://www.pbs.org/newshour/politics/ai-generated-disinformation-poses-threat-of-misleading-voters-in-2024-election.....7

Sri Muppidi, *OpenAI Considers New Financing at a $1.5 Trillion Valuation*, NEW YORK TIMES (Sep. 16, 2026, 10:57 AM), https://www.nytimes.com/2026/09/16/business/dealbook/openai-new-funding-round.html.....2

Ben Shimkus, *It took OpenAI's agents 130 billion tokens to crack a 90-year-old math problem*, BUSINESS INSIDER (Sep. 9, 2026, 11:55 AM), https://www.businessinsider.com/openai-math-problem-solved-tokens-cost-altman-2026-9.....6

Steve Stecklow & Poppy McPherson, *We set out to craft the perfect phishing scam. Major AI chatbots were happy to help.*, REUTERS (Sep 15, 2025, 10:30 AM), https://www.reuters.com/investigates/special-report/ai-chatbots-cyber/.....7

Loreben Tuquero, *AI agents are hacking systems without any input from humans. How did we get here?*, PBS (Sep. 2, 2026, 2:46 PM), https://www.pbs.org/newshour/science/ai-agents-are-hacking-systems-without-any-input-from-humans-how-did-we-get-here.....7

Juliana Castro Varón & Dylan Freedman, *Will A.I. Make Your Brain Lazy? Here's What the New Research Actually Shows.*, NEW YORK TIMES (Sep. 29, 2026) https://www.nytimes.com/interactive/2026/09/29/magazine/ai-chatbots-brain-development-study.html.....7

Other Authorities

AI Insights, CLOUDFLARE, https://radar.cloudflare.com/ai-insights#crawl-to-refer-ratio (last visited Sep. 11, 2026).....17

CommonPile at https://proceedings.neurips.cc/paper_files/paper/2025/hash/52acc050138d6f40dad6f12f91a4ce22-Abstract-Datasets_and_Benchmarks_Track.html.....13

Market capitalization of Microsoft (MSFT), https://companiesmarketcap.com/microsoft/marketcap/ (last visited Oct. 21, 2026).....2

Introducing ChatGPT search, OPENAI (Oct. 31, 2024), https://openai.com/index/introducing-chatgpt-search/.....17

SpaceX Super Intelligence (@SpaceXSI), X (Jul. 10, 2025, 12:01 AM) https://x.com/SpaceXAI/status/1943158495588815072 (at 12 minutes and 55 seconds into the video).....4

INTEREST OF AMICI CURIAE

Amici curiae are organizations and individuals dedicated to promoting the rule of law and the sound administration of justice. Amicus David Atkinson is the lead author of this brief and has written extensively on the legal implications of GenAI, including works addressing copyright and fair use issues. The Amici do not represent any specific commercial interest or party to this litigation. Rather, Amici's interest is ensuring that statutory frameworks are interpreted faithfully according to their plain language and legislative intent. Amici submits this brief to offer a framework to assist the judiciary in analyzing claims of fair use.

SUMMARY OF ARGUMENT

If one were designing legal arguments sure to swallow copyright law, one could hardly do better than the arguments made by Defendants in this case. Defendants are demanding an exemption from copyright law far beyond what any court has ever granted to human beings, and more expansive than any court has granted to any other company. The companies are arguing for generative artificial intelligence ("GenAI") exceptionalism, and if Defendants' use is fair use under any theory they or their amici have suggested, then nearly every use by nearly every human is also fair use because the arguments can universally apply to every individual.

Defendants want the court and the public to believe that written works are intended solely to be read, implying that learning from the works is, at most, an incidental aim for readers and that informing readers is not the intent and purpose of news publishers. In other words, Defendants want everyone to believe that their use of written works is special, privileged, and entirely unconstrained by copyright law, even though they use the works for the same purpose—learning from the work—as does anyone who picks up a newspaper or scrolls through a news article.

Defendants also assert that their special privilege rests on the claim that their products, including ChatGPT and Copilot, supposedly provide vast public benefits. Some people do benefit from chatbots, but it would be foolhardy to believe chatbot benefits exceed those provided by humans who do not rely on sweeping fair use arguments. This includes the discovery of every field of science and math, every notable genre of art, and every profound invention. No general-purpose AI system, let alone a chatbot, has done nearly as much to promote the progress of society. This is despite Defendants already training their models on orders of magnitude more content than any human is exposed to. If humans can accomplish so much without dismantling copyright law, then surely one tech company aiming for a roughly $1.5 trillion valuation[1] and another with a market cap of over $3.5 trillion[2] can do the same.

Moreover, simply being novel or impressive is not sufficient to be "transformative" in the copyright sense of the term. The fact that a model consumes text derived from various sources and that its mathematical underpinnings enable it to produce outputs that differ from the source material is little different, as a practical matter, from human beings reading a variety of different material and our brains producing novel written work, something that we have done for millennia.

The Defendants also claim that their models somehow only identify "the facts embodied in the articles,"[3] but this sophistry is targeted at those who may not understand how AI systems work. To provide a summary of a work, Defendants must copy the entire work, process it in its entirety, and only then produce an output. There is no magical software that can somehow skip all

---

[1] Sri Muppidi, *OpenAI Considers New Financing at a $1.5 Trillion Valuation*, NEW YORK TIMES (Sep. 16, 2026, 10:57 AM), https://www.nytimes.com/2026/09/16/business/dealbook/openai-new-funding-round.html.

[2] *Market capitalization of Microsoft (MSFT)*, https://companiesmarketcap.com/microsoft/marketcap/ (last visited Oct. 21, 2026).

[3] Defendants' Memorandum of Points and Authorities In Support of Motion for Summary Judgment at 13, *In Re: OpenAI, Inc., Copyright Infringement Litigation*, No. 1:25-md-03143-SHS-OTW (S.D.N.Y. Sep. 4, 2026) (ECF 1753).

the expression of a work and extract only facts and ideas. Those facts are embedded within expression, and defendants undeniably rely on that expression in their training data and with their retrieval augmented generation ("RAG") products to understand the context of the source, including which facts matter most and what they mean or imply. No court has held that a person or non-AI company can reproduce all the works of the world for free so long as they only provide summaries of the contents rather than distribute the original, or so long as they don't share publicly what's stored in their brains after reading the content, and there is no compelling reason to bend the law for the sake of AI companies.

At base, Defendants are asking for a breathtaking form of exceptionalism to enrich themselves while leaving the rest of society more impoverished by reducing the incentive for humans to create.

This brief will *first* address the purpose of copyright law and how Defendants' technology falls short of it compared to humans. *Second*, it will address the most common misguided arguments, including that because models (a) learn differently from humans;[4] (b) use only non-protected facts and ideas; (c) generate new material when generating outputs;[5] and (d) require humongous training datasets, their use is fair use. All of these arguments miss the mark. *Third*, it will explain why there is no implied license to take and use the copyrighted material at issue. The final section will describe the likely negative outcomes if the court supports GenAI exceptionalism.

[4] *See* Order Denying the Plaintiffs' Motion for Partial Summary Judgment and Granting Meta's Cross-Motion for Partial Summary Judgment at 17, *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417-VC (N.D. Cal. Jun. 25, 2025) (ECF 598).

[5] *See Id.*

ARGUMENT

I. Defendants' use of copyrighted works does not satisfy the purpose of copyright law and does not justify fair use.

"What copyright law cares about, above all else, is preserving the incentive for human beings to create artistic and scientific works."[6] Therefore, it's fair to ask whether granting Defendants sweeping fair use would promote such progress. But the inquiry must not end there. To ensure the court's reasoning is principled, it must also consider why granting such a sweeping fair use exemption to AI companies would promote science and the useful arts, whereas granting the same exemption to humans would not. After reviewing all plausible arguments, the only conclusion must be that GenAI exceptionalism is unjustified.

The Supreme Court has noted that fair use "identifies various factors that enable a court to apply an 'equitable rule of reason' analysis to particular claims of infringement."[7] That equitable analysis might favor a technology that dramatically accelerates progress. Defendants' products do not. Even GenAI CEOs acknowledge their models have yet to make significant breakthroughs.[8] Sam Altman claimed in mid-2025 that "2026 will likely see the arrival of systems that can figure out novel insights."[9] But in September 2026, the *New York Times* quoted Mark Chen, OpenAI's chief research officer, saying, "[t]he big value in human researchers is that they generate all these creative ideas all the time. . . . Our models are just not doing that right now."[10] That admission

---

[6] *Id*. at 1.

[7] *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417, 447–48 (1984).

[8] In July 2025, Elon Musk stated that "I would expect Grok to . . . literally discover new technologies that are actually useful no later than next year and maybe end of this year. . . . And it might discover new physics next year and within two years, I'd say almost certainly." SpaceX Super Intelligence (@SpaceXSI), X (Jul. 10, 2025, 12:01 AM) https://x.com/SpaceXAI/status/1943158495588815072 (at 12 minutes and 55 seconds into the video).

[9] Sam Altman, *The Gentle Singularity*, SAM ALTMAN BLOG (Jun. 10, 2025, 5:12 PM), https://blog.samaltman.com/the-gentle-singularity.

[10] Cade Metz, *The Liftoff Scenario That Terrifies A.I. Doomsayers*, NEW YORK TIMES, (Sep 16, 2024), https://www.nytimes.com/2026/09/16/science/ai-recursive-self-improvement.html.

comes even though Defendants have already acquired and used virtually all publicly available machine-readable data in the world, and Microsoft alone has spent over $100 billion on the OpenAI partnership.[11]

In contrast, even without a broad exemption from copyright law and relying on the same troves of information, humans have developed theories such as quantum mechanics, inventions such as computers, discoveries such as antibiotics, and new genres of art including rock and roll and impressionism. Defendants' chatbot products have not created any comparable breakthroughs despite training on billions of documents, videos, images, recordings, and code without authorization.

Even OpenAI's most compelling counterexample, solving hard math problems, is marred by controversy and suffers from at least five issues:

1. Mathematicians have alleged that OpenAI may have cribbed from their work after having already made significant headway on the Millennium Problem.
2. Mathematicians, including Terence Tao, recipient of the Fields Medal, have said that merely deriving answers could undermine progress in math because it obscures the learning that leads to breakthroughs.[12]

[11] Rachel Metz, Brody Ford, & Madlin Mekelburg, *Microsoft Spent Over $100 Billion on OpenAI Partnership*, BLOOMBERG (May 13, 2026, 1:15 PM), https://www.bloomberg.com/news/articles/2026-05-13/microsoft-spent-over-100-billion-on-openai-partnership-to-date.

[12] Geoff Blumfiel, *AI solved one of math's hardest problems. Humanity learned nothing (so far)*, NPR (Sep. 22, 2026, 7:08 AM), https://www.npr.org/2026/09/22/nx-s1-5968588/openai-navier-stokes-problem-mathematicians-learn-little ("But mathematicians say that, so far, humanity has learned very little from the solution the company's AI model purportedly found."); Cade Metz, *OpenAI Says It Has Cracked One of Math's 'Millennium Problems'*, NEW YORK TIMES (Sep 8, 2026, 5:39 PM), https://www.nytimes.com/2026/09/08/science/openai-proof-millennium-problem.html ("Dr. Tao is among the many mathematicians who have publicly warned that the latest A.I. systems could end up damaging the field of mathematics. If A.I. technologies can solve the most difficult problems with little input from human mathematicians, he says, they could weaken human understanding of the field. 'The effort needed to solve problems is often very instructive. It teaches you something. It's like going to the gym and having a goal to lift a weight a hundred times,' he explained. 'Now, A.I. can solve questions without really getting any value out of them. It's like having machines that can lift weights for you at the gym.'").

3. The method of solving the most famous problem was essentially brute force, with some 10,000 AI agents burning through several million dollars in compute in just a few days.[13]
4. The training material used for math is mostly open-source or in the public domain, so copyright is largely not implicated. Advanced math models therefore do not depend on a fair use exemption, and their achievements cannot support one.
5. The well-formulated problems and the math used to solve the problems were developed by humans. OpenAI's solutions were not akin to "discovering" the number zero, imaginary, or irrational numbers, or the development of algebra or calculus, for example.

In sum, defendants have yet to propose a novel problem that, if solved, would meaningfully improve society, let alone solve such a problem. It does not make sense to permit mass infringement on the hope that perhaps someday the breakthrough will occur. Even if it does, it will remain less impressive than the breadth and depth of human achievements.

If granting fair use for an entity is likely to lead to profound breakthroughs, it's far more likely to come from humans exposed to a relatively infinitesimal amount of copyrighted work. Yet Congress and the courts have recognized the wisdom of not granting a broad exemption from the law for humans because, though it may lead to near-term breakthroughs, it would undermine incentives to continue progress over time. Progress must be thought of in decades and centuries, not the sugar high of months or years. The speculative benefits of what AI "might," "will," or "could" do cannot justify present-tense mass infringement.

[13] Ben Shimkus, It took OpenAI's agents 130 billion tokens to crack a 90-year-old math problem, BUSINESS INSIDER (Sep. 9, 2026, 11:55 AM), https://www.businessinsider.com/openai-math-problem-solved-tokens-cost-altman-2026-9.

Defendants may have over a billion users,[14] but use is not benefit, and some benefit is not net benefit. Some people undoubtedly gain utility from the products, but the models have also been used for fraud,[15] deception,[16] misinformation,[17] disinformation,[18] and computer hacking,[19] have contributed to cognitive surrender,[20] and have served other uses that actively harm society. It is far from clear that the nation or the world is better off today than before ChatGPT debuted in 2022 and the system that became Copilot debuted in early 2023. Contrary to OpenAI's claim that "[the] public benefits are vast,"[21] there is not enough empirical evidence to rest a fair use defense on the technology's contribution to science and the useful arts, and the argument weakens further once its harms are weighed against its benefits.

II. If Defendants' use of Plaintiffs' works is fair use, then so must virtually all unauthorized and uncompensated uses by humans.

[14] *See* Defendants' Memorandum of Points and Authorities In Support of Motion for Summary Judgment at 7, *In Re: OpenAI, Inc., Copyright Infringement Litigation*, No. 1:25-md-03143-SHS-OTW (S.D.N.Y. Sep. 4, 2026) (ECF 1753).

[15] *See* Steve Stecklow & Poppy McPherson, *We set out to craft the perfect phishing scam. Major AI chatbots were happy to help.*, REUTERS (Sep 15, 2025, 10:30 AM), https://www.reuters.com/investigates/special-report/ai-chatbots-cyber/.

[16] *See* Ruchika Joshi & Michal Luria, *AI-Powered Deception: A Deeper Dimension of Dark Design Patterns in Conversational AI Tools and Platforms*, CENTER FOR DEMOCRACY & TECH (SEP. 15, 2025). https://cdt.org/insights/ai-powered-deception-a-deeper-dimension-of-dark-design-patterns-in-conversational-ai-tools-and-platforms/

[17] *See* Tiffany Hsu & Stuart A. Thompson, *Disinformation Researchers Raise Alarms About A.I. Chatbots*, NEW YORK TIMES (Feb. 8, 2023, 3:21 PM), https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html.

[18] *See* David Klepper & Ali Swenson, *AI-generated disinformation poses threat of misleading voters in 2024 election*, PBS (May 14, 2024, 7:52 PM), https://www.pbs.org/newshour/politics/ai-generated-disinformation-poses-threat-of-misleading-voters-in-2024-election.

[19] *See* Loreben Tuquero, *AI agents are hacking systems without any input from humans. How did we get here?*, PBS (Sep. 2, 2026, 2:46 PM), https://www.pbs.org/newshour/science/ai-agents-are-hacking-systems-without-any-input-from-humans-how-did-we-get-here.

[20] *See* Juliana Castro Varón & Dylan Freedman, *Will A.I. Make Your Brain Lazy? Here's What the New Research Actually Shows.*, NEW YORK TIMES (Sep. 29, 2026) https://www.nytimes.com/interactive/2026/09/29/magazine/ai-chatbots-brain-development-study.html.

[21] Defendants' Memorandum of Points and Authorities In Support of Motion for Summary Judgment at 1, *In Re: OpenAI, Inc., Copyright Infringement Litigation*, No. 1:25-md-03143-SHS-OTW (S.D.N.Y. Sep. 4, 2026) (ECF 1753).

There are strong doctrinal and technical reasons why Defendants' use of Plaintiffs' works is not fair use. But each fair use factor can also be tested against a simpler principle: AI companies should not receive privileges that courts would refuse to extend to humans absent strong justification. Applied to the purpose of the use, the amount taken, and the ongoing use of stored copies, that principle defeats Defendants' position.

There would be little dispute that it would not be lawful for students to make an unauthorized copy of a textbook for a course rather than pay for it, even if the students learn from the book, do not particularly enjoy reading it, the book has a minuscule effect on their thinking relative to all the other content the students are exposed to in their lives, the students do not quote the book at length verbatim, do not distribute unauthorized copies of the book, and the content of the book is necessary for them to learn in order to become marketable. And yet, the Defendants argue that their use of unauthorized copies for pre-training is fair use for those very reasons.

It's also unclear why LLMs should be allowed to use an unlimited number of complete books or articles to learn, but humans cannot pirate even a single book that they may need. If the argument is that more books and articles lead to better models, that holds for human brains as well. This exposes the weakness of Judge Chhabria's reasoning in finding that "feeding a whole book to an LLM does more to train it than would feeding it only half of that book. With this in mind, it was 'reasonably necessary' for Meta to 'make use of the entirety of the works.'"[22]

Similarly, no student is permitted to make unauthorized copies of the hundreds, thousands, or millions of works they think might be useful to them and keep them as potential references, even if they use only portions of the works and their subsequent writings do not merely regurgitate

[22] *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1049-50 (N.D. Cal. 2025) (quoting *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 98 (2d Cir. 2014).

the content they reference. But the defendants claim their retrieval augmented generation products are shielded by fair use for analogous actions. If we want to preserve copyright, we cannot accept the argument that nearly all uncompensated and unauthorized uses by AI companies and humans is fair use.

A. AI systems learning from patterns is not relevantly different from how humans learn.

At its broadest, Defendants' argument is that because their AI systems learn from the works they copy, the copying is fair use. No one, however, may go to a bookstore or library, make a complete copy of a book or newspaper, and avoid liability because they learned from it. Defendants should be held to the same standard.

More narrowly, defendants argue that models don't learn like humans, and they imply that this difference is relevant to a fair use analysis. As Judge Chhabria put it: "[A]n LLM's consumption of a book is different than a person's. An LLM ingests text to learn 'statistical patterns' of how words are used together in different contexts. It does so by taking a piece of text from its training data, removing a word from that text, predicting what that word will be, and updating its general understanding of language based on whether it was right or wrong—and then repeating this exercise billions or trillions of times with different text. This is not how a human reads a book."[23]

Models may learn by detecting patterns in text, but so do humans. Reading is not innate; it requires repeated exposure to text. People learn to read by associating words with other words, sentences with other sentences, and concepts with other concepts. When they do not know what a word means, they rely on context clues, and grade-school vocabulary lessons often require students

[23] *Kadrey*, 788 F. Supp. 3d at 1044.

to complete partial sentences, much as a model predicts a missing word. Each time, human brains adapt. Nobody can explain how, exactly, but neurochemical transmitters adjust as some connections between some neurons strengthen while others weaken. The neural networks that the AI systems at issue rely on were, in fact, loosely inspired by brain neurons. They are a simplified version of what happens in everyone's heads. There is no legal reason to find that a simplified version of a system must be granted greater leniency under fair use than the more complex system on which it is based.

Moreover, differences in the learning process are irrelevant to a fair use analysis. That an LLM embeds statistical patterns using math is of no consequence and, if anything, is less impressive, mysterious, and complex than a human brain transforming photons and sound waves into neurochemical signals and storing them as knowledge and skills in living cells that are constantly changing. The argument that learning from statistical modeling rather than biological adaptation favors fair use is simply an argument for GenAI exceptionalism.

B. As with humans, AI models cannot obtain facts without also consuming the expression of works.

Arguments that models are only interested in uncopyrightable facts and ideas similarly fail. In journalism, facts and ideas often cannot be neatly separated from the expression within which they're embedded. No Defendant's AI system can extract the facts and ideas from the vast majority of the content at issue without first copying the copyrighted work, including all of its expressive elements. Defendants elide the fact that they download and process the entire text before presenting a summary as output.

When AI systems demonstrate a certain style or tone in their outputs, it is only possible because they are trained on the expression in the training data, which includes Plaintiffs' works. A model cannot accurately imitate a particular journalist merely by extracting factual elements

from that journalist's articles, for example. Rather, the model incorporates elements of the journalists' expression.[24] It is not as if when humans imitate someone, we retain some essence of the person's expression in our brains, but when models imitate someone, they somehow do so while only relying on the underlying ideas of the person's writing.

The truth is that the Defendants have preferences for the type of data they train on. They need varied syntax, edited content, high signal-to-noise-ratio content, and other features that the articles at issue embody. Relying on less-refined data yields a less useful model, even when the underlying facts and ideas are identical.

If the model's sole focus were the underlying facts and ideas, it could just train on websites hosting hackneyed news summaries rather than source material from professional journalists and authors. The justification would also apply to virtually every math and science student. Few students read textbooks because they're enthralled by the expression describing the Krebs cycle or differential equations. Rather, the material is a means of accessing the ideas the authors hope to convey. Still, students are not permitted to acquire all their textbooks that happen to be "publicly available" from pirate libraries or elsewhere without authorization. Quite the contrary: they must pay a small ransom each semester for the privilege of learning.

This was also the case in *American Geophysical Union v. Texaco, Inc*., where the court held that researchers using a for-profit library could not avoid copyright infringement by systematically reproducing journal articles. Nobody in *Texaco* thought the scientists read the journals purely out of love for the expression in the journals and that learning was an incidental outcome of their enjoyment. They wanted the knowledge, insights, and ideas in the papers. Yet

[24] *See,* Suchin Gururangan, et al., *Whose Language Counts as High Quality? Measuring Language Ideologies in Text Data Selection*, in EMNLP 2022, THE 2022 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, PROCEEDINGS OF THE CONFERENCE, 2562 (Dec. 2022), https://aclanthology.org/2022.emnlp-main.165.pdf (explaining that GPT-3 filter does not upweight factuality but rather mimics news type articles).

fair use did not excuse how they pursued the facts and ideas. The court was correct in *Texaco,* and there is no reason to accept GenAI exceptionalism to evade a similar outcome here.

C. Humans, like AI systems, usually generate novel outputs and rarely say or write anything that matches what we've seen, heard, or read verbatim.

The argument that fair use must apply to defendants' AI because the training material enables novel outputs from models is also unhelpful to defendants. Humans also nearly exclusively create and respond to questions and requests with novel or uncopyrightable expressions. We may have read hundreds, perhaps thousands, of books and articles, but few of us can recite any significant chunk of any of them with much fidelity. Rather, we read or listen to the material and paraphrase what we consumed (if we comment on it at all). Also, as with models, no single work is guaranteed to be of any consequence to learning; rather, the cumulative effect is what is important. But unlike Defendants, humans generally pay — either directly or indirectly — for the material we learn from.

Again and again, Defendants focus on how their AI systems' outputs allegedly do not infringe. However, if outputs were all that mattered, humans would almost never infringe on any works we reproduced without authorization to learn from. It's a rare occurrence for an individual to produce an infringing work. Rarer still is when the infringing work is in response to an off-the-cuff question, as in the examples provided by defendants, "(*e.g.*, 'when does MoMA open today')".[25] If the basis of the argument must rest on outputs, the Defendants have again failed to demonstrate why the court should treat their product differently than it would treat a human.

[25] Defendants' Memorandum of Points and Authorities In Support of Motion for Summary Judgment at 12, *In Re: OpenAI, Inc., Copyright Infringement Litigation*, No. 1:25-md-03143-SHS-OTW (S.D.N.Y. Sep. 4, 2026) (ECF 1753).

D. The inefficient learning of AI systems should not excuse further infringement.

Defendants may argue that their products require millions or tens of millions of works from Plaintiffs to learn.[26] But claiming that such an extreme volume of material is necessary to create AI systems is also not a compelling defense. Copyright law offers no exception that permits weaker learners to acquire more works without authorization than quicker learners. The fact that a human reading ten news articles a week for 80 years may "only" read 42,000 articles should not make their use less shielded from copyright law than that of AI models, which apparently need access to millions of articles. The inefficiency of model learning cannot grant them sweeping fair use rights, nor can it mean that the less efficient they are at learning, the greater their rights must be. Society does not allow struggling students to take more copyrighted works without authorization because they struggle to learn a concept from just one or a few authorized sources. If defendants want to infringe less often, it is incumbent on them to develop technology that does not depend on mass infringement;[27] it must not be society's burden to permit mass infringement for the sake of for-profit companies.

III. Defendants' differences only reinforce that they should not be granted fair use.

To the extent there are meaningful differences between Defendants' products and humans, those differences only serve to reinforce that a finding of fair use would be inappropriate. For example, Defendants' AI systems can read several orders of magnitude more content than any human,[28] memorize far more of it than any human ever could, and generate outputs far faster than

[26] *See, e.g.*, *Kadrey*, 788 F. Supp. 3d at 1050 (when discussing the third fair use factor: "Everyone agrees that LLMs work better if they are trained on more high-quality material. . . . So feeding a whole book to an LLM does more to train it than would feeding it only half of that book." (internal citations omitted)).

[27] *See generally*, for example, CommonPile at https://proceedings.neurips.cc/paper_files/paper/2025/hash/52acc050138d6f40dad6f12f91a4ce22-Abstract-Datasets_and_Benchmarks_Track.html.

[28] Even OpenAI's older model, GPT-3, was trained on 45 terabytes of data. *See* Tom B. Brown et al., *Language Models are Few-Shot Learners*, ARXIV (Jul. 22, 2020), https://arxiv.org/pdf/2005.14165.

any human ever will.[29] Each of these is a difference in kind rather than degree, and they alter the character, reach, and market consequences of defendants' copying.

The inhuman abilities of Defendants' products mean they are far more likely to store a copy of a work within their systems, to regurgitate those works as outputs, and to drown the markets of plaintiffs. When Congress made the most recent substantive update to copyright law in 1976, they were not contemplating a system that could gulp down over 100 terabytes of data (for context, 100 terabytes can store at least several million text documents, like PDFs) and retain detailed information about a significant portion of it, including verbatim details of many of the works. Congress was also likely contemplating a human pace of expressive production. For example, an author might put out a few decent-quality novels a year on the high end. In contrast, Defendants' AI systems can pump out news summaries, essays, and book-length outputs in seconds or minutes, nonstop, year-round.

A technology's capacity to produce more material, serve more users, or operate more quickly does not establish that the uncompensated copying used to build and operate it should be privileged. Otherwise, the greater a Defendant's capacity to exploit copyrighted works and displace their markets, the stronger its fair-use claim would become. That rule would invert copyright's incentive structure: the uses most capable of appropriating value from authors would receive the broadest immunity from paying them.

IV. There is no implied license to reproduce Plaintiffs' works.

An implied license exists when there is no formal or written permission, but the actions, conduct, or circumstances of a property or intellectual property owner show that they intended to

[29] OpenAI's systems were estimated to process around 20 *trillion* tokens per day in October 2025. Jean-Stanislas Denain, Anson Ho, & Jamie Sevilla, *How many digital workers could OpenAI deploy?*, EPOCH AI (Oct. 3, 2025), https://epoch.ai/gradient-updates/how-many-digital-workers-could-openai-deploy.

allow someone else to use it. The Defendants helpfully quote from *Field v. Google*, which stated that "[c]onsent to use the copyrighted work need not be manifested verbally and may be inferred based on silence where the copyright holder knows of the use *and encourages it*." (emphasis added). The Second Circuit has only stated: "We have not yet ruled 'on the precise circumstances under which an implied non-exclusive license will be found.' *ABKCO Music, Inc. v. Sagan* (2d Cir. 2022), citing *Psihoyos v. Pearson Educ., Inc*., 855 F. Supp. 2d 103, 120 (S.D.N.Y. 2012) ((internal quotation marks omitted). Moreover, the Second Circuit noted that an implied license requires a "meeting of the minds between the parties to permit the particular usage at issue," *Psihoyos*, 855 F. Supp. 2d at 124 (internal quotation marks omitted)). There was no meeting of the minds here.

A. Publicly available is not the same as public domain.

Defendants argue that "[i]t is hard to imagine a fairer use than one a copyright holder can block at will."[30] They go on to say that News Plaintiffs' "knowing decision not to exclude their works from Copilot gives rise to an implied license."[31] But the onus shouldn't be on the websites to opt out. Publicly available is not the same as public domain; works aren't in the public domain by default, and they are not fair game to copy and use absent explicit notice from the copyright holder. If Defendants were correct, virtually all unauthorized copies would be fair use because most copyright holders would be unaware any given copy was made, and therefore unable to protest.

As noted in the Plaintiffs' brief, the Defendants acquired content through third parties such as Common Crawl and through anonymous scrapes, so blocking defendant bots via robots.txt

[30] Defendant Microsoft Corporation's Memorandum of Law In Support Of Motion For Summary Judgment In News Plaintiffs' Consolidated Cases at 3, *In Re: OpenAI, Inc., Copyright Infringement Litigation*, No. 1:25-md-03143-SHS-OTW (S.D.N.Y. Sep. 4, 2026) (ECF 1698).

[31] *Id.*

would be of limited use or entirely useless anyway.[32] A right to opt-out is not really a right if the Defendants can simply obtain the content from another source.

The argument for fair use based on public availability is also intellectually dishonest. Almost anything can be made publicly available, including ill-gotten content.[33] For example, someone could circumvent technological measures meant to limit a work's reproduction or distribution and then post it on the public internet, making it publicly available and, according to defendants, perfectly fine to use. In fact, pirate websites containing digital versions of at least millions of books, articles, and magazines are publicly available. Granting an implied license to publicly available content is little different from granting a license to any content the defendants can acquire.

B. Defendants' ravenous scraping is unlike traditional search engines and undermines the internet's economics.

Defendants also seem to misunderstand how the economics of the internet work. Websites traditionally permitted, and even encouraged, web indexers to crawl them because it made them more discoverable and increased web traffic. Increased web traffic, in turn, brought benefits to the websites. The assumptions that made websites encourage crawling for search, and therefore suggested an implied liacense, do not apply to crawling to train AI models.

The difference between the effect of AI scraping and traditional search engines is stark. We can see this in data from Cloudflare, through which about a fifth of all internet traffic goes, and which tracks crawl-to-refer ratios. Cloudflare's CEO claimed that "in the past, for every two

[32] News Plaintiffs' Combined Summary Judgment Brief at 10, *In Re: OpenAI, Inc., Copyright Infringement Litigation*, No. 1:25-md-03143-SHS-OTW (S.D.N.Y. Sep. 4, 2026) (ECF 1709).

[33] In a separate lawsuit, Meta admitted to this in an Answer. Defendant Meta Platforms, Inc.'s Answer to the Complaint at 12, *Carreyrou v. Anthropic PBC*, No. 5:25-cv-10897-PCP (N.D. Cal. Feb. 26, 2026) (ECF 109) ("Meta admits that it obtained data from publicly available sources such as Books3, LibGen, Z-Library, and Common Crawl …").

pages that Google crawled to inform its search engine, it would, on average, send one visitor to those sites. . .”[34] It was a mutually beneficial relationship. In contrast, as of September 11, 2026, OpenAI crawled 333 webpages for every referral it provided. Microsoft’s ratio was 38:1.[35] Generative AI search using RAG is utterly incomparable to traditional search engines.

With RAG, the chatbot fetches current articles at the time of the user's query, ingests them in full, and then generates an output that eliminates the need for users to click on the source material. This is, in fact, the intent. When introducing ChatGPT Search, OpenAI stated that “[g]etting useful answers on the web can take a lot of effort. It often requires multiple searches and digging through links to find quality sources and the right information for you. Now, chat can get you to a better answer. . .”[36]

The basis of the bargain was that search engines could earn money from ad revenue and websites would benefit from an increase in web traffic. That is no longer the case. A bot visiting a page to scrape its content is not the same as a human visiting a webpage. Human visitors may subscribe to a service on the page, generate ad revenue, make a (possibly impromptu) purchase, promote the website to friends and followers, and otherwise provide some benefit to the website owner, helping to keep the website sustainable. There are, in fact, several reasons why a person or entity may want their content to be public while *not* wanting GenAI bots to scrape them.[37]

The RAG bots by Defendants, in contrast, do none of these. Their bots don’t generate ad revenue; they don’t buy anything from the websites; they don’t share interesting webpages from their nonexistent social media pages. They parasitically extract and exploit. If the search bots were

---

[34] Jim Edwards, *Cloudflare CEO says Google is abusing its monopoly in search to feed its AI*, FORTUNE (Nov. 13, 2025, 4:32 AM), https://fortune.com/2025/11/13/cloudflare-ceo-google-abusing-monopoly-search-ai/.
[35] AI Insights, CLOUDFLARE, https://radar.cloudflare.com/ai-insights#crawl-to-refer-ratio (last visited Sep. 11, 2026).
[36] *Introducing ChatGPT search*, OPENAI (Oct. 31, 2024), https://openai.com/index/introducing-chatgpt-search/.
[37] *See* David Atkinson, *Unfair Learning: GenAI Exceptionalism and Copyright Law*, 25 Chi.-Kent J. Intell. Prop. 43 (2026).

such a boon for websites, defendants would have shared click-through rates, provided internal crawl-to-referral ratios, or provided other data showing how their actions clearly benefit websites in ways comparable to traditional search engine indexing. Indeed, if the benefits were obvious, Plaintiffs would have likely encouraged *more* scraping.

V. If the court affirms GenAI exceptionalism, it will undermine progress in science and the useful arts.

Defendants seek a privilege that would be unthinkable for even the most accomplished individual. Granting it would also harm society, directly contradicting the constitutional aim of promoting science and the useful arts.

If the court upholds the Defendants' view, similarly situated companies will likely enjoy the same exceptional treatment, including Anthropic, Meta, Google, and a slew of start-ups. Without having to pay for some of the most important content they train on while simultaneously providing a competing product that serves as an answer engine that can replace the training content, a foreseeable chain of events will likely unfold: plaintiffs will have to cut back on the number of journalists they employ and the number or quality of stories they publish, local news deserts will expand, and relatively expensive and time-consuming investigative journalism will decrease. A finding of fair use that does no more than humans to promote science and the useful arts, while undermining an essential institution of modern society, cannot be squared with copyright's purpose.

CONCLUSION

For these reasons, the court should refuse to grant Defendants exceptional treatment under the law. Declaring their use fair use would undermine incentives for creators to keep working

toward the progress of science and the useful arts, and it would establish that AI companies receive special treatment under the law.

DATED: October 30, 2026

Respectfully submitted,

/s/ *Kylie M. Huff*

Kylie M. Huff
Alexa Millinger – *S.D.N.Y. Admission Pending*
Hinckley, Allen & Snyder LLP
20 Church Street
Hartford, CT 06103
T: 860-725-6200
F: 860-278-3802
KHuff@hinckleyallen.com
AMillinger@hinckleyallen.com

Attorneys for proposed Amici Curiae AI Safety Asia, Brian J. A. Boyd, Buddhism & AI Initiative, Cara Project, Inc., Charles Camosy, David Atkinson, EdTech Law Center PLLC, Future of Life Institute, Irreplaceable, John C. Havens, Noēsis Collaborative, Progressive Democrats of America, RadicalxChange Foundation, The Tech Oversight Project, Walter Scheirer, and Young People's Alliance

CERTIFICATE OF COMPLIANCE WITH WORD-COUNT LIMITAION

Pursuant to Local Civil Rule 7.1(c), undersigned counsel hereby certifies that this Memorandum of Law complies with the word count limit specified in the Court's Order of August 7, 2026 (Dkt. No. 1663) limiting the proposed brief to 6000 words or less, as the brief contains 5989 words excluding material exempted by Local Civil Rule 7.1(c). Undersigned counsel relied on the word count of the word-processing program used to prepare this document.

DATED: October 30, 2026

Respectfully submitted,

/s/ *Kylie M. Huff*

Kylie M. Huff
Alexa Millinger (SDNY Admission Pending)
Hinckley, Allen & Snyder LLP
20 Church Street
Hartford, CT 06103
T: 860-725-6200
F: 860-278-3802
KHuff@hinckleyallen.com
AMillinger@hinckleyallen.com

Attorneys for proposed Amici Curiae AI Safety Asia, Brian J. A. Boyd, Buddhism & AI Initiative, Cara Project, Inc., Charles Camosy, David Atkinson, EdTech Law Center PLLC, Future of Life Institute, Irreplaceable, John C. Havens, Noēsis Collaborative, Progressive Democrats of America, RadicalxChange Foundation, The Tech Oversight Project, Walter Scheirer, and Young People's Alliance

CERTIFICATION

I hereby certify that on October 30, 2026, a copy of the foregoing was filed electronically and served by mail on anyone unable to accept electronic filing. Notice of this filing will be sent by email to all parties by operation of the Court's electronic filing system or by mail to any counsel or self-represented party unable to accept electronic filing, as indicated on the Notice of Electronic Filing. Parties may access this filing through the Court's CM/ECF System.

/s/ *Kylie M. Huff*
Kylie M. Huff