EXHIBIT A

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION This Document Relates To: 1:25-cv-04315-SHS-OTW 1:23-cv-11195-SHS-OTW 1:23-cv-08292-SHS-OTW	**25-MD-3143 (SHS) (OTW)**

BRIEF OF A.I. PROGRESS, INC. AS AMICUS CURIAE IN SUPPORT OF DEFENDANTS' MOTIONS FOR SUMMARY JUDGMENT

TABLE OF CONTENTS

Page

INTEREST OF AMICUS CURIAE 1

INTRODUCTION AND SUMMARY OF ARGUMENT 1

ARGUMENT 5

I. Fair Use Has Long Enabled Transformative Technologies That Advance Copyright's Purpose of Expanding Public Knowledge and Understanding. 5

II. Generative AI Training Is Fair Use and Serves Copyright's Core Purpose of Expanding Public Knowledge and Understanding. 7

A. Generative AI Training Is a Justified, "Quintessentially Transformative" Use. 7

B. Generative AI Training Does Not Cause Cognizable Market Harm. 12

1. Generative AI Training Does Not Provide a Substitute for Copyrighted Works. 12

2. Publicly Announced Business Agreements Do Not Establish a Licensing Market for AI Training. 15

3. Copyright Owners Cannot Manufacture Market Harm by Defining the Relevant Market as a License for the Challenged Use. 17

4. Copyright Does Not Protect Authors From Competition or "Market Dilution." 18

CONCLUSION 21

TABLE OF AUTHORITIES

Page(s)

Cases

Am. Geophysical Union v. Texaco Inc.,
60 F.3d 913 (2d Cir. 1994).....15

Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith,
598 U.S. 508 (2023)..... *passim*

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

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

Bartz v. Anthropic PBC,
787 F.Supp.3d 1007 (N.D. Cal. 2025) 11, 18

Bill Graham Archives v. Dorling Kindersley Ltd.,
448 F.3d 605 (2d Cir. 2006).....18

Campbell v. Acuff-Rose Music, Inc.,
510 U.S. 569 (1994).....8, 12

Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.,
499 U.S. 340 (1991).....5

Fox News Network, LLC v. TvEyes, Inc.,
883 F.3d 169 (2d Cir. 2018).....14, 15

Google LLC v. Oracle Am., Inc.,
593 U.S. 1 (2021).....12, 19

Hachette Book Grp., Inc. v. Internet Archive,
115 F.4th 163 (2d Cir. 2024) 14

Kadrey v. Meta Platforms Inc.,
788 F.Supp.3d 1026 (N.D. Cal. 2025) 11, 18, 19, 20

Kelly v. Arriba Soft Corp.,
336 F.3d 811 (9th Cir. 2003) 6

Perfect 10, Inc. v. Amazon.com, Inc.,
508 F.3d 1146 (9th Cir. 2007) 7

Romanova v. Amilus Inc.,
138 F.4th 104 (2d Cir. 2025) ..... *passim*

Sega Enters. Ltd. v. Accolade, Inc.,
977 F.2d 1510 (9th Cir. 1992) 9, 19

Sony Corp. of Am. v. Universal City Studios, Inc.,
464 U.S. 417 (1984) 5, 6

Twentieth Century Music Corp. v. Aiken,
422 U.S. 151 (1975) 5

Statutes and Legislative History

17 U.S.C. § 102(b) 9

17 U.S.C. § 106 5

H.R. Rep. No. 94-1476 (1976) 5

U.S. Const. art. I, § 8, cl. 8 5

Other Authorities

A. Feder Cooper et al., *Report of the 1st Workshop on Generative AI and Law*, ARXIV (Dec. 3, 2023), https://arxiv.org/pdf/2311.06477 3

Anthropic, *The Anthropic Economic Index* (Feb. 10, 2025),
https://www.anthropic.com/news/the-anthropic-economic-index 20

Anthropic, *Building AI for Cyber Defenders* (Sept. 29, 2025),
https://www.anthropic.com/research/building-ai-cyber-defenders 3

Daniil A. Boiko et al., *Autonomous Chemical Research with Large Language Models*, 624 Nature (2023), https://www.nature.com/articles/s41586-023-06792-0 14

Edward Lee, *The Ethics of Innovation: p2p Software Developers and Designing Substantial Noninfringing Uses Under the Sony Doctrine*, 62 J. Bus. Ethics (2005). "By 1999, 88.6 percent of all U.S. households owned a VCR." Kelu L. Sullivan, *Orphan Works at the Dawn of Digitization*, 18 Rich. J.L. & Tech. (2012) 6

Emergency Event Response System, https://eers.us/ (last visited Sept. 17, 2026) 2

Google Signs Deal with AP to Deliver Up-To-Date News Through Its Gemini AI Chatbot, The Associated Press (Jan. 15, 2025), https://www.ap.org/media-center/ap-in-the-news/2025/google-signs-deal-with-ap-to-deliver-up-to-date-news-through-its-gemini-ai-chatbot/ 16

Jeffrey S. Allen, *Monitoring AI Adoption in the US Economy*, FEDS Notes (Apr. 3, 2026), https://www.federalreserve.gov/econres/notes/feds-notes/monitoring-ai-adoption-in-the-u-s-economy-20260403.html 1

Jordan Hoffmann et al., *Training Compute-Optimal Large Language Models* (2022), https://arxiv.org/pdf/2203.15556.....10

Kevin J. Delaney & Jeffrey A. Trachtenberg, *Publishers Challenge Google's Book-Scanning Efforts*, Wall St. J. (Oct. 20, 2005, 12:01 AM ET).....7

Kritika Lamba, *Getty, Perplexity Sign Multi-Year Licensing Deal to Boost AI-powered Search Visuals*, Reuters (updated Oct. 31, 2025), https://www.reuters.com/business/getty-perplexity-sign-multi-year-licensing-deal-boost-ai-powered-search-visuals-2025-10-31/ 16

4 Melville B. Nimmer & David Nimmer, *Nimmer on Copyright* (2026).....17

Meta, *Bringing More Real-Time News and Content to Meta AI* (Dec. 5, 2025), https://about.fb.com/news/2025/12/bringing-more-real-time-news-and-content-to-meta-ai/ 16

Meta, *Fostering Innovation and Economic Opportunity: The Llama Impact Grant Recipients* (Apr. 29, 2025), https://about.fb.com/news/2025/04/llama-impact-grant-recipients/ 2

Noah Losch et al., *Fine-Tuning LLMs on Small Medical Datasets: Text Classification and Normalization Effectiveness on Cardiology Reports and Discharge Records* (Mar. 27, 2025), https://arxiv.org/html/2503.21349 13

Org. for Econ. Coop. & Dev., *AI Use by Individuals Surges Across the OECD as Adoption by Firms Continues to Expand* (Jan. 28, 2026), https://www.oecd.org/en/about/news/announcements/2026/01/ai-use-by-individuals-surges-across-the-oecd-as-adoption-by-firms-continues-to-expand.html.....2

Org. for Econ. Coop. & Dev., *Skills in the AI Age*, OECD Artificial Intelligence Paper No. 60 (2026), https://www.oecd.org/content/dam/oecd/en/publications/reports/2026/07/skills-in-the-ai-age_e8d8c1e6/972bd15e-en.pdf.....1

Ping Guo et al., *Teaching Large Language Models to Translate on Low-Resource Languages with Textbook Prompting*, in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (2024), https://aclanthology.org/2024.lrec-main.1362.pdf.....14

Shalini Gakhar & Raj Kumar Singh, *Leveraging Large Language Models (LLMs) for GeoAI-Enabled Digital Agro-Advisory*, 7 Frontiers in Remote Sensing (2026), https://www.frontiersin.org/journals/remote-sensing/articles/10.3389/frsen.2026.1839369/full 14

Shekoofeh Azizi & Bryan Perozzi, *How a Gemma Model Helped Discover a New Potential Cancer Therapy Pathway*, Google (Oct. 15, 2025), https://blog.google/innovation-and-ai/products/google-gemma-ai-cancer-therapy-discovery/.....2

Solo Tech, https://getsolo.tech/ (last visited Sept. 17, 2026).....2

Travis Clark et al., *The 28-Year Rise of Netflix, from DVD-by-Mail Service to Hollywood Juggernaut*, Business Insider (Dec. 5, 2025), https://www.businessinsider.com/netflix-history-streaming-growth-story-hollywood-disruption-subscribers-2022-6.....6

Tyler A. Chang et al., *Scalable Influence and Fact Tracing for Large Language Model Pretraining*, Int'l Conf. on Learning Representations (2025), https://openreview.net/forum?id=gLa96FlWwn.....11

Yue Jiang et al., *UrbanLLM: Autonomous Urban Activity Planning and Management with Large Language Models*, in *Findings of the Association for Computational Linguistics: EMNLP 2024* 1810 (2024), https://aclanthology.org/2024.findings-emnlp.98.pdf.....13

Zheng Dai & David K. Gifford, *Outputs of Generative Diffusion Models Are Often Unattributable*, 17 Nature Commc'ns 6974 (2026), https://www.nature.com/articles/s41467-026-75667-5.....11

Zhengliang Liu et al., *Radiology-Llama2: Best-in-Class Large Language Model for Radiology*, arXiv (Aug. 29, 2023), https://arxiv.org/abs/2309.06419.....2

Zhenyu Lei et al., *Harnessing Large Language Models for Disaster Management: A Survey*, in *Findings of the Association for Computational Linguistics: ACL 2025* (2025), https://aclanthology.org/2025.findings-acl.750.pdf 13

INTEREST OF AMICUS CURIAE

Amicus curiae A.I. Progress, Inc. is a coalition of leading AI developers committed to the advancement of AI tools that can deliver benefits across societies and economic sectors, from public safety to scientific discovery. A.I. Progress champions the essential role of existing U.S. copyright law in fostering innovation, protecting intellectual property, and ensuring America's competitiveness in AI. A.I. Progress has a strong interest in this case because the fair use principles applied here could have far-reaching consequences not only for AI innovation, but for the public benefits it makes possible.[1]

INTRODUCTION AND SUMMARY OF ARGUMENT

A ruling against fair use in this case would threaten not merely a single industry, but the broader economy due to the rapidly expanding ecosystem of generative AI tools already benefiting Americans across society. AI is no longer an emerging technology awaiting public adoption. It is now part of everyday life. As of November 2025, approximately 41% of U.S. workers reported work-related generative AI use, nearly half of Americans used it outside the workplace, and 78% of the U.S. workforce was employed by firms that had adopted AI, including approximately 54% employed by firms using large language models ("LLMs"). Jeffrey S. Allen, *Monitoring AI Adoption in the US Economy*, FEDS Notes (Apr. 3, 2026), https://www.federalreserve.gov/econres/notes/feds-notes/monitoring-ai-adoption-in-the-u-s-economy-20260403.html. According to the Organisation for Economic Co-operation and Development, business use of AI in its member countries increased from roughly 7% to 20%, while more than one-third of individuals used generative AI tools in 2025. Org. for Econ. Coop. & Dev., *Skills in the AI Age* 12, OECD Artificial Intelligence Papers No. 60 (2026),

[1] No party or its counsel authored this brief, in whole or in part, or funded the preparation or submission of this brief.

https://www.oecd.org/content/dam/oecd/en/publications/reports/2026/07/skills-in-the-ai-age_e8d8c1e6/972bd15e-en.pdf; Org. for Econ. Coop. & Dev., *AI Use by Individuals Surges Across the OECD as Adoption by Firms Continues to Expand* (Jan. 28, 2026), https://www.oecd.org/en/about/news/announcements/2026/01/ai-use-by-individuals-surges-across-the-oecd-as-adoption-by-firms-continues-to-expand.html.

Far from being a niche technology, generative AI is actively being deployed across the economy for a wide range of applications, from helping people navigate emergencies, to accelerating scientific and medical breakthroughs, to expanding access to knowledge and protecting critical infrastructure. Even a small sampling demonstrates LLMs' remarkable versatility. During natural disasters, the Emergency Event Response System ("E.E.R.S.") helps victims locate shelters, access government resources, and obtain recovery assistance; researchers use C2S-Scale to analyze cellular data and identify potential new avenues for cancer treatment; at the Mayo Clinic, Radiology-Llama2 helps healthcare providers interpret and communicate complex medical information; Solo Tech uses LLMs to bring educational, agricultural, and healthcare resources to underserved rural communities, even without reliable internet access; and LLM-powered cybersecurity tools help protect critical infrastructure by identifying vulnerabilities before they can be exploited. E.E.R.S., https://eers.us/ (last visited Sept. 17, 2026); Shekoofeh Azizi & Bryan Perozzi, *How a Gemma Model Helped Discover a New Potential Cancer Therapy Pathway*, Google (Oct. 15, 2025), https://blog.google/innovation-and-ai/products/google-gemma-ai-cancer-therapy-discovery/; Zhengliang Liu et al., *Radiology-Llama2: Best-in-Class Large Language Model for Radiology*, arXiv (Aug. 29, 2023), https://arxiv.org/abs/2309.06419; Solo Tech, https://getsolo.tech/ (last visited Sept. 17, 2026); Meta, *Fostering Innovation and Economic Opportunity: The Llama Impact Grant Recipients*

(Apr. 29, 2025), https://about.fb.com/news/2025/04/llama-impact-grant-recipients/; Anthropic, *Building AI for Cyber Defenders* (Sept. 29, 2025), https://www.anthropic.com/research/building-ai-cyber-defenders.

These are not speculations about what LLMs may someday deliver. They are existing real-world applications of a technology that is already yielding tangible public benefits. And they share a feature that is significant under settled fair-use principles: not one of them delivers its benefit by giving the public access to the works on which the models were trained. The disaster victim receives directions to shelter and recovery resources; the cancer researcher, new avenues for investigation; the clinician, help interpreting complex medical information; the rural community, access to educational, agricultural, and healthcare resources; and the cybersecurity professional, identification of potential vulnerabilities—not passages from a copyrighted book. The technological benefits and underlying training datasets are decoupled. The sheer range of these beneficial uses confirms that the purpose of AI training—a key consideration in the fair-use inquiry—is not to generate substitutes for authors' works. Rather, during training, an LLM analyzes vast amounts of data and copies of works to identify patterns, structures, and relationships in language, enabling entirely *new* capabilities that bear little resemblance to the expressive works on which they were trained. *See* A. Feder Cooper et al., *Report of the 1st Workshop on Generative AI and Law*, ARXIV (Dec. 3, 2023), https://arxiv.org/pdf/2311.06477.

Copyright law has repeatedly faced similar moments, with transformative technologies challenged as threats to copyright owners and markets. Yet courts concluded that those technologies served new and socially beneficial purposes under fair use. The result was not harm to copyright's objectives, but dramatic expansions in access to information, knowledge, and new creative works.

The same principles apply here. Generative AI training is a transformative use because it does not use copyrighted works to communicate their expression. Courts applying the Supreme Court's fair-use framework have already held that training LLMs serves a different purpose from the underlying works and therefore the first fair-use factor strongly favors fair use.

Nor does AI training cause cognizable market harm under the Copyright Act. The fourth fair-use factor asks whether a challenged use supplies a substitute for the original work. AI training does not distribute copyrighted works, permit users to access them, or provide substitute copies of works on which AI models have been trained. Nor can Plaintiffs manufacture market harm by characterizing every commercial agreement involving AI as evidence of a market for AI *training*. A review of public announcements describing agreements between AI companies and content owners reveals that many concern fundamentally different commercial uses, including product integrations, API access, content distribution, and real-time information services, rather than the licensing of copyrighted works for AI training.

Finally, Plaintiffs' theory rests on the notion that generative AI may enable the creation of additional, original expressive works that compete in the marketplace with existing works. But copyright has never granted authors a right to suppress lawful competition. To the contrary, encouraging the creation of more expression by more creators is one of the principal ways copyright advances its constitutional purpose.

The Court should reject Plaintiffs' efforts to expand copyright beyond its lawful scope. Fair use has long protected transformative technologies that advance knowledge and serve the public interest. As it should here.

ARGUMENT

I. Fair Use Has Long Enabled Transformative Technologies That Advance Copyright's Purpose of Expanding Public Knowledge and Understanding.

"The primary objective of copyright is not to reward the labor of authors, but 'to promote the Progress of Science and useful Arts.'" *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 349 (1991) (quoting U.S. Const. art. I, § 8, cl. 8). As the Second Circuit recently explained, "[t]he ultimate goal of copyright is to expand public knowledge and understanding," and copyright seeks to achieve that goal by giving creators limited exclusive rights in their works, which encourages them to create and share works. *Romanova v. Amilus Inc.*, 138 F.4th 104, 109 (2d Cir. 2025) (quoting *Authors Guild v. Google, Inc.*, 804 F.3d 202, 212 (2d Cir. 2015) ("*Google Books*")). Copyright therefore balances the creation and availability of expressive works for public benefit. *Id.* (quoting *Twentieth Century Music Corp. v. Aiken*, 422 U.S. 151, 156 (1975)).

Copyright owners have never enjoyed "complete control over all possible uses of [their] work." *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 432 (1984). Instead, copyright confers a limited set of enumerated rights, each expressly subject to important statutory limitations. *See* 17 U.S.C. § 106 (providing that copyright owners' exclusive rights are "[s]ubject to sections 107 through 122" of the Copyright Act). Among those limitations, Congress recognized fair use as "one of the most important and well-established limitations on the exclusive right of copyright owners" and emphasized its flexibility to accommodate "rapid technological change." H.R. Rep. No. 94-1476, at 65–66 (1976). That flexibility has repeatedly enabled transformative technologies to emerge and flourish, notwithstanding predictions that they would harm authors, disrupt established markets, or undermine copyright itself.

Perhaps the most famous example is the VCR. In *Sony*, the Supreme Court concluded that private, noncommercial "time-shifting" constituted fair use. 464 U.S. at 455. It did so despite motion picture studios claiming that the nascent technology allowing consumers to record television broadcasts for later viewing threatened to ruin the entertainment industry.

That decision proved prescient. Far from harming copyright markets, the VCR fueled demand for programming. Consumers gained the ability to watch programs on their own schedules and the emerging technology "spawned an entirely unforeseen market, with sales of tapes for home viewing later comprising the largest source of revenue for the movie industry." Edward Lee, *The Ethics of Innovation: p2p Software Developers and Designing Substantial Noninfringing Uses Under the Sony Doctrine*, 62 J. Bus. Ethics 147, 148 (2005). "By 1999, 88.6 percent of all U.S. households owned a VCR." Kelu L. Sullivan, *Orphan Works at the Dawn of Digitization*, 18 Rich. J.L. & Tech. 6, ¶ 2 n.6 (2012) (citation omitted). Video rental stores evolved into DVD-by-mail services and, ultimately, streaming platforms, expanding public access to creative works while creating substantial new revenue streams for copyright owners. *See* Travis Clark et al., *The 28-Year Rise of Netflix, from DVD-by-Mail Service to Hollywood Juggernaut*, Business Insider (Dec. 5, 2025), https://www.businessinsider.com/netflix-history-streaming-growth-story-hollywood-disruption-subscribers-2022-6.

The same pattern emerged with internet search engines. In *Kelly v. Arriba Soft Corp.*, a search engine copied copyrighted photographs and displayed thumbnail images to help users locate information online. 336 F.3d 811, 818 (9th Cir. 2003). The Ninth Circuit held the use was transformative because the thumbnails served a different function than the original photographs, acting as navigational tools rather than substitutes for the images themselves. *Id.* at 818–20. The court reaffirmed those principles in *Perfect 10, Inc. v. Amazon.com, Inc.*, recognizing the highly

transformative role search engines play in helping users locate information across the internet. 508 F.3d 1146, 1165–67 (9th Cir. 2007). Those decisions helped pave the way for search engines that are now essential to education, commerce, research, and civic participation.

The same principle enabled the creation of Google Books. Google digitized millions of books, creating a searchable index that allows users to locate information across the world's published literature. The plaintiffs challenged the project as infringing, but the Second Circuit held that creating the searchable index was highly transformative and constituted fair use. *Google Books*, 804 F.3d at 216–17, 225. That decision dramatically expanded public knowledge by enabling access to information once buried in distant or forgotten collections. As one commentator observed, a contrary ruling would have made it "essentially impossible to maintain a search engine." Kevin J. Delaney & Jeffrey A. Trachtenberg, *Publishers Challenge Google's Book-Scanning Efforts*, Wall St. J. (Oct. 20, 2005, 12:01 AM ET), https://www.wsj.com/articles/SB112973759679773260.

These examples demonstrate a recurring truth: technologies once criticized as threats to copyright markets often become powerful engines of public benefit when fair use permits their development. That lesson is especially important here. A ruling that prevents or substantially impedes the training of generative AI models risks foreclosing a rapidly growing class of LLM-based technologies already helping people navigate emergencies, accelerate scientific and medical breakthroughs, expand access to knowledge, and protect critical infrastructure.

II. Generative AI Training Is Fair Use and Serves Copyright's Core Purpose of Expanding Public Knowledge and Understanding.

A. Generative AI Training Is a Justified, "Quintessentially Transformative" Use.

There can be no serious dispute that generative AI training is transformative under the first fair-use factor. The Supreme Court has emphasized that this factor turns on whether the

secondary use serves a different purpose than the original work. The question is whether the new use merely "supersede[s] the objects" of the original or instead "adds something new, with a further purpose or different character." *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 528 (2023) (quoting *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994)). *Warhol* explained that transformative uses are "justified" because, "in a broad sense," they "further[] the goal of copyright, namely, to promote progress of science and the arts, without diminishing the incentive to create." *Id.* at 510–11. And, "[i]n a narrower sense, a use may be justified because copying is reasonably necessary to achieve the user's new purpose." *Id.*; *accord Romanova*, 138 F.4th at 114–116 (identifying justifications for copying that have been held to be fair uses). While commerciality of the secondary use may be relevant, it is not dispositive. *Warhol*, 598 U.S. at 531.

Warhol reaffirmed, rather than retreated from, decisions holding that copying is transformative when undertaken to extract information rather than exploit expression. The Court expressly relied on *Google Books*, which held that Google Books was transformative because it enabled users to access information about the books rather than substitute for them. *Google Books*, 804 F.3d at 217. The searchable database facilitated new forms of research, including the analysis of "word frequencies, syntactic patterns, and thematic markers" and the study of "nomenclature, linguistic usage, and literary style." *Id.* at 209. Likewise, in *Authors Guild, Inc. v. HathiTrust*, the Second Circuit held that copying millions of books to create a searchable database was "quintessentially transformative." 755 F.3d 87, 97 (2d Cir. 2014). A search result, the court explained, differs fundamentally "in purpose, character, expression, meaning, and message" from the underlying book. *Id.* Together, these decisions establish that copying to

derive information, identify patterns, and enable new forms of research serves a distinct purpose that strongly favors fair use.

Generative AI training falls squarely within that tradition. Copyright protects expression, not the facts, ideas, concepts, and information embodied in a work. 17 U.S.C. § 102(b). Nor does copyright extend to the functional elements embodied in a work. *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1522–28 (9th Cir. 1992) (finding fair use where computer code was copied to discern the functional patterns, structures, and relationships necessary to understand the software's operation). Accordingly, copyright owners have no exclusive right to control the extraction of information about language, patterns, or relationships reflected in their works. As the Second Circuit explained, copyright does not confer "an exclusive right to furnish" information about a work itself. *Google Books*, 804 F.3d at 225–26.

And the Second Circuit has specifically recognized that such informational and functional uses justifies copying. That principle applies where the copying "provided information to the public about the copied work, or enabled the furnishing of valuable information on any subject of public interest or rendered a valuable service to the public, in most cases limited to circumstances in which the benefit was provided without allowing public access to the copy." *Romanova*, 138 F.4th at 115. Generative AI training fits comfortably within these recognized justifications: it uses copyrighted works to derive information about patterns and relationships across vast bodies of material and, in turn, enables a new technological capability that can furnish valuable information and services to the public.

The real-world applications discussed above demonstrate the point: LLMs are already helping disaster victims find resources, researchers explore potential cancer treatments, healthcare providers interpret complex information, and underserved communities access

essential services. These applications exemplify precisely the kind of valuable information and public services that *Romanova* recognizes as justifying copying.

The training that enables these tools follows precisely that structure. Research has shown that LLMs perform better when trained on far more data than previously assumed, and that the amount of training data should increase as models become larger. *See* Jordan Hoffmann et al., *Training Compute-Optimal Large Language Models* 1–2 (2022), https://arxiv.org/pdf/2203.15556. A broad corpus is thus copied so that a system can analyze the works and develop new capabilities; what the public receives is the resulting capability, not access to the copied datasets used in training. That aligns AI training with *Romanova's* recognition that copying may be justified when it enables valuable information or services without providing public access to the copies. Indeed, the separation is even more pronounced for AI training than in the uses in *Google Books* and *HathiTrust*, where users could receive snippets or search results drawn from the underlying works. Training itself, by contrast, provides the public no portion of any work at all.

Rather than using works for their original expressive purpose, LLM training involves analyzing vast collections of works to learn patterns, relationships, and structures in language, images, and other forms of expression. Models are not designed to store works for later retrieval or republication. Instead, they learn statistical relationships that enable them to respond to prompts and generate *new* content. The copying thus serves precisely the kinds of informational and functional purposes that the Second Circuit has recognized as legitimate justifications for copying. *Romanova*, 138 F.4th at 115.

Nor do outputs derive from any single work. Modern models are trained on millions or billions of works across datasets, with any individual work's influence diffused across the

model's parameters. The relationship between any particular input and any particular output is therefore not one-to-one. *See* Tyler A. Chang et al., *Scalable Influence and Fact Tracing for Large Language Model Pretraining*, Int'l Conf. on Learning Representations 1, 4 (2025), https://openreview.net/forum?id=gLa96FlWwn. Researchers have found that the influence of any individual work becomes increasingly diffused as models are trained on larger datasets, making it difficult and often impossible to trace a particular output to a specific work, creator, or artist. *See* Zheng Dai & David K. Gifford, *Outputs of Generative Diffusion Models Are Often Unattributable*, 17 Nature Commc'ns 6974 (2026), https://www.nature.com/articles/s41467-026-75667-5.

Courts applying the Supreme Court's fair-use framework have already held that LLM training is not merely transformative, but "quintessentially" or "highly" transformative. In *Bartz v. Anthropic PBC*, the court held that training LLMs on copyrighted works was "quintessentially transformative" because it uses those works to derive patterns, structures, and relationships in order to create a new technological capability, rather than to reproduce or supplant the works themselves. 787 F.Supp.3d 1007, 1022 (N.D. Cal. 2025). Likewise, in *Kadrey v. Meta Platforms Inc.*, the court recognized that LLMs learn "statistical patterns" of language rather than communicate the expressive content of books, and ultimately concluded that LLM training serves a "further purpose" and possesses a "different character" from the books on which it trains, making it "highly transformative." 788 F.Supp.3d 1026, 1044 (N.D. Cal. 2025).

Those conclusions follow directly from settled fair-use principles. A use that extracts information about language, patterns, and relationships to create a new technology does not "supersede[] the objects" of the original works. *Warhol*, 598 U.S. at 528. Nor does it substitute for those works in the marketplace. And it falls comfortably within the Second Circuit's

recognition that copying may be justified where it enables the furnishing of valuable information or renders a valuable service to the public without providing the public access to the copies themselves. *Romanova*, 138 F.4th at 115. Like the search and research tools permitted in *Google Books* and *HathiTrust*, AI training uses copyrighted works for a fundamentally different purpose that advances copyright's ultimate goal of promoting learning, creativity, and progress.

B. Generative AI Training Does Not Cause Cognizable Market Harm.

1. Generative AI Training Does Not Provide a Substitute for Copyrighted Works.

The fourth fair-use factor asks whether the secondary use "usurps the market of the original work," *HathiTrust*, 755 F.3d at 99, or "brings to the marketplace a competing substitute for the original," *Google Books*, 804 F.3d at 223.

Copyright does not, however, protect copyright owners from every economic effect that may accompany transformative innovation. As the Supreme Court explained, "a potential loss of revenue is not the whole story," as a "lethal parody, like a scathing theatre review," may "kil[l] demand for the original." *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 35 (2021) (citing *Campbell*, 510 U.S. at 591–92) (alteration in original). Such harm, "even if directly translated into foregone dollars, is not 'cognizable under the Copyright Act.'" *Id.* (citing *Campbell*, 510 U.S. at 592).

Courts have repeatedly found no cognizable market harm where a transformative use does not provide substitute access to protected expression. In *Google Books*, for example, the Second Circuit concluded that the service posed no meaningful threat to the market for books because users could access only "discontinuous, tiny fragments" of text, not complete works or substitute copies. 804 F.3d at 223–24.

The same principle controls here. Generative AI training is not designed to disseminate copyrighted works, allow users to read them, or provide substitute access to expressive content in datasets. Its purpose is to derive information about language, structure, and relationships from a vast corpus of material to be able to create *new* content. Like the databases in *Google Books* and *HathiTrust*, AI training does not compete with the market for the underlying works because it does not make those works available to the public as substitutes. Nor does the mere potential for infringing outputs by itself render the separate, earlier process of AI training infringing. *Warhol* directs the fair-use inquiry to the particular use challenged as infringing, rather than a separate use occurring earlier or later. 598 U.S. at 533–34. The possibility that a downstream user might prompt a model to produce an infringing output cannot determine whether the antecedent copying undertaken to train the model is itself fair use. The Court must evaluate training for what training does and the purpose it serves.

Many documented real-world LLM deployments, such as those discussed above, do not produce outputs that compete with expressive works at all. Instead, they support disaster response and infrastructure planning, as well as produce clinical classifications, crop assessments, translations for low-resource languages, scientific analyses, and other functional or informational outputs. *See, e.g.*, Zhenyu Lei et al., *Harnessing Large Language Models for Disaster Management: A Survey*, in *Findings of the Association for Computational Linguistics: ACL 2025* 14,528 (2025), https://aclanthology.org/2025.findings-acl.750.pdf; Yue Jiang et al., *UrbanLLM: Autonomous Urban Activity Planning and Management with Large Language Models*, in *Findings of the Association for Computational Linguistics: EMNLP 2024* 1810 (2024), https://aclanthology.org/2024.findings-emnlp.98.pdf; Noah Losch et al., *Fine-Tuning LLMs on Small Medical Datasets: Text Classification and Normalization Effectiveness on*

Cardiology Reports and Discharge Records (Mar. 27, 2025), https://arxiv.org/html/2503.21349; Shalini Gakhar & Raj Kumar Singh, *Leveraging Large Language Models (LLMs) for GeoAI-Enabled Digital Agro-Advisory*, 7 Frontiers in Remote Sensing (2026), https://www.frontiersin.org/journals/remote-sensing/articles/10.3389/frsen.2026.1839369/full; Ping Guo et al., *Teaching Large Language Models to Translate on Low-Resource Languages with Textbook Prompting*, in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* 15,685 (2024), https://aclanthology.org/2024.lrec-main.1362.pdf; Daniil A. Boiko et al., *Autonomous Chemical Research with Large Language Models*, 624 Nature 570 (2023), https://www.nature.com/articles/s41586-023-06792-0. The hypothetical use of an LLM to generate a novel that competes with an author's work thus represents only a narrow slice of potential LLM applications, not their characteristic form.

Plaintiffs invoke *Internet Archive* and *TVEyes* for the proposition that the fourth factor considers both actual market harm and the potential harm that would result from widespread similar conduct. ECF No. 1881 ("Book Pls. Partial Mot. Summ. J."), at 23 (citing *Hachette Book Grp., Inc. v. Internet Archive*, 115 F.4th 163, 189 (2d Cir. 2024)); *Fox News Network, LLC v. TvEyes, Inc.*, 883 F.3d 169, 179 (2d Cir. 2018). But even under that framework, AI training does not cause cognizable market harm. Unlike the non-transformative uses in *Internet Archive* and *TVEyes*, which provided the public with complete or lengthy copies of the underlying copyrighted works without any new purpose, AI training does not distribute works or provide substitute access to them. Instead, it identifies patterns, structures, and relationships. As a result, even widespread AI training would not place competing copies of copyrighted works in the marketplace—thus, a very different scenario than those in *Internet Archive* and *TVEyes*. *See*

Internet Archive, 115 F.4th 163, 188 (2d Cir. 2024) (holding that defendant distributed copies of the plaintiffs' books "to the public in full"); *TVEyes*, 883 F.3d at 179 (finding that defendant "likely provide[d] . . . users with all of the Fox programming that they seek and the entirety of the message conveyed by Fox to authorized viewers of the original").

These decisions reflect a simple principle: absent substitute access to copyrighted works or competing copies of the same works, alleged economic injury is too remote from copyright's core concern with market substitution to carry significant weight under the fourth factor.

2. Publicly Announced Business Agreements Do Not Establish a Licensing Market for AI Training.

Plaintiffs point to publicly announced business agreements between AI companies and content owners as purported evidence of an existing or emerging market for AI training. Book Pls. Partial Mot. Summ. J. at 37–38. But that argument conflates agreements involving AI with agreements licensing content for AI training. A closer look at public announcements of these agreements describe product integrations, API access, content distribution, and real-time information services—not standalone purchases of AI-training rights.

Plaintiffs' reliance on the sheer number of these agreements does not bridge that gap. They invoke "more than 120 licenses between content providers and AI companies," *id.* at 37, but the number of agreements says nothing about whether they establish a market for AI training. What matters is what those agreements actually license. Training is distinct from other downstream uses of AI systems, and agreements licensing those distinct uses do not establish a market for the use of copyrighted works to train AI models. *Texaco* underscores the distinction. There, the court credited potential licensing revenues because a "viable market" existed for licensing the particular right implicated by the challenged use. *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 930 (2d Cir. 1994). Here, by contrast, commercial agreements covering

materially different downstream uses do not establish a cognizable market for generative AI training.

For example, Google's partnership with the Associated Press combines access to real-time content through Reddit's Data API and AP's news reporting with Google's AI tools and infrastructure to enhance the quality, timeliness, and functionality of consumer-facing AI products. Matt O'Brien, *Google Signs Deal with AP to Deliver Up-To-Date News Through Its Gemini AI Chatbot*, The Associated Press (Jan. 15, 2025), https://www.ap.org/media-center/ap-in-the-news/2025/google-signs-deal-with-ap-to-deliver-up-to-date-news-through-its-gemini-ai-chatbot/.

Similarly, Perplexity's arrangement with Getty Images integrates licensed photographs into search and discovery tools through APIs, attribution features, and related-content displays. Kritika Lamba, *Getty, Perplexity Sign Multi-Year Licensing Deal to Boost AI-powered Search Visuals*, Reuters (updated Oct. 31, 2025), https://www.reuters.com/business/getty-perplexity-sign-multi-year-licensing-deal-boost-ai-powered-search-visuals-2025-10-31/. The same pattern appears in Meta's agreements with organizations such as CNN, Fox News, and USA TODAY. Those deals integrate content into Meta AI's user-facing products through features such as news summaries, links, and real-time information. Meta, *Bringing More Real-Time News and Content to Meta AI* (Dec. 5, 2025), https://about.fb.com/news/2025/12/bringing-more-real-time-news-and-content-to-meta-ai/.

These agreements share a common feature: they are multifaceted commercial partnerships that bundle content access with consumer-facing product features, distribution mechanisms, and technology services. Much of their value derives from improving user experiences, expanding content distribution, generating traffic, or supporting downstream

products and services. That reality undercuts Plaintiffs' licensing-market theory. If anything, industry agreements show that sophisticated parties negotiating AI-related licenses routinely bargain over a wide range of commercial rights and services extending well beyond model training itself.

More fundamentally, evidence of a functioning market for one use does not establish a licensing market for a materially different use. The agreements discussed above reflect familiar copyright markets in which content is displayed, distributed, accessed, or incorporated into consumer-facing products. Plaintiffs seek to use evidence of those markets to establish a separate market for the antecedent and fundamentally different use of works to train AI models. But evidence of commercial demand for licensed display and product uses does not establish a cognizable licensing market for the distinct and transformative use of copyrighted works in AI training.

3. Copyright Owners Cannot Manufacture Market Harm by Defining the Relevant Market as a License for the Challenged Use.

Nor can Plaintiffs establish market harm simply by asserting a right to license the challenged use itself. As Professor Nimmer has explained, there is an inherent "danger of circularity" in defining the relevant market as the market for licensing the very use alleged to be fair, because "it is always a given" that a copyright owner can claim the loss of such a market. 4 Melville B. Nimmer & David Nimmer, *Nimmer on Copyright* § 13F.08[B] (2026).

The Second Circuit has recognized the same limitation. "[N]ot every effect on potential licensing revenues enters the analysis under the fourth factor"; courts consider only "traditional, reasonable, or likely to be developed markets." *Texaco*, 60 F.3d at 929–30.

For that reason, courts have consistently rejected efforts to monopolize licensing markets for transformative uses. In *Bill Graham Archives v. Dorling Kindersley Ltd.*, the Second Circuit

warned that if the absence of a licensing fee automatically constituted market harm, the fourth factor would "always favor the copyright holder." 448 F.3d 605, 614 (2d Cir. 2006). Similarly, *HathiTrust* held that because full-text search was transformative and non-substitutive, it was "irrelevant" whether libraries might have purchased licenses. 755 F.3d at 99–100. More recently, both *Kadrey* and *Bartz* rejected the argument that AI training harms an emerging licensing market because copyright law does *not* grant authors exclusive control over transformative uses of that nature. *Kadrey*, 788 F.Supp.3d at 1052; *Bartz*, 787 F.Supp.3d at 1031–32.

History underscores the wisdom of that rule. Search engines, image-search tools, digital research databases, and Google Books all could have been characterized as impairing potential licensing opportunities. Courts nevertheless rejected those arguments because the uses were transformative and did not serve as substitutes for the underlying works.

Generative AI training fits within that framework. The challenged act of training does not distribute the training works, display them to users, or provide users substitute access to them. Rather, AI training uses copyrighted works to extract information about language, patterns, and relationships in order to develop new content and technological capabilities. Where a use is transformative and does not reveal substantial protected expression, alleged losses arising from foregone licensing opportunities do not constitute cognizable market harm.

4. Copyright Does Not Protect Authors From Competition or "Market Dilution."

No Supreme Court or Second Circuit decision has held that copyright protects authors from competition. As explained in *Bartz*, the creation of new expressive works that may compete with existing works "is not the kind of competitive or creative displacement that concerns the Copyright Act," which "seeks to advance original works of authorship, not to protect authors against competition." 787 F.Supp.3d at 1031–32. Treating increased competition as cognizable

harm would transform copyright from a system that promotes creativity into one that suppresses it.

Every novel competes with other novels, every newspaper article with other sources of information, and every film for consumers' limited time and attention. Yet copyright has never treated such competition as actionable harm. To the contrary, copyright is designed to encourage the creation of more works and more competitors. The Ninth Circuit recognized this principle in *Sega*, explaining that copyright law seeks to foster creative expression for the public benefit. 977 F.2d 1510 at 1523. Likewise, the Supreme Court has instructed courts to consider whether a challenged use promotes the creation of new expression and thereby advances copyright's goals. *Google*, 593 U.S. at 35.

The implications of Plaintiffs' theory would be extraordinary. Any technology that lowers the cost of producing expression could be accused of "diluting" demand for existing works. Desktop publishing software, digital cameras, social media, search engines, and self-publishing platforms all dramatically increased the volume of available expression. Yet copyright law has never treated such an increase in creative output as actionable harm.

Although the court in *Kadrey*, sua sponte, raised market dilution as a potential theory of market harm, it nonetheless recognized the serious defects in that theory. There, despite finding no evidence that the model's outputs harmed the market for plaintiffs' books, the court speculated that AI systems might someday generate large numbers of competing works. *Kadrey*, 788 F.Supp.3d at 1035–36, 1052. At the same time, however, the court acknowledged that "an LLM trained only on public domain works could still be capable of quickly generating large numbers of books that could compete for sales with copyrighted books." *Id.* at 1053. That observation exposes the fundamental defect in the dilution theory. If the same alleged harm

would occur even when training is conducted solely on public-domain materials, then the harm cannot stem from the use of copyrighted works. It necessarily stems from the production of new expressive works that compete with existing ones for consumers' time, attention, and purchasing decisions. In other words, the theory does not identify harm resulting from the exploitation of copyrighted expression; it identifies harm resulting from increased competition. Copyright law has never recognized the latter as a cognizable injury.

The observed pattern of LLM use confirms as much. According to a published report, arts, design, entertainment, sports, and media tasks accounted for only about one in ten Claude.ai queries, while computer and mathematical tasks accounted for more than a third of all queries, by far the largest occupational category. *See* Anthropic, *The Anthropic Economic Index* (Feb. 10, 2025), https://www.anthropic.com/news/the-anthropic-economic-index. In practice, as noted above, LLMs are being deployed across medicine, scientific research, disaster response, agriculture, and cybersecurity planning to analyze information, classify conditions, assess risks, and provide functional guidance—uses far removed from generating books, articles, or other competing expressive works. Plaintiffs' dilution theory thus elevates one application of LLM technology into its characteristic use, when documented evidence shows otherwise. And as the *Kadrey* court's own observation confirms, even for that narrow application, the asserted harm flows from the creation of new competing works, not from the use of copyrighted works for training.

Nor would such a rule be confined to AI. It would permit copyright owners to claim injury whenever any new technology increases the supply of lawful expression and intensifies competition. Copyright safeguards against the unauthorized copying of protected expression, not the emergence of new competitors offering their own lawful expression.

CONCLUSION

The Court should reject Plaintiffs' attempt to extend copyright beyond its established limits. AI training uses copyrighted works for a fundamentally different purpose, does not provide the public with substitute access to those works, and does not cause the market harm cognizable under the Copyright Act. Instead, it enables a broad array of transformative technologies that are already delivering public benefits. Copyright exists to promote creativity and progress, not to suppress lawful innovation or protect authors from competition. Settled fair-use principles therefore compel the conclusion that generative AI training is fair use.

Dated: October 30, 2026

Respectfully submitted,

/s/ Anna B. Chauvet

FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, LLP
Anna B. Chauvet (pro hac vice)
anna.chauvet@finnegan.com
901 New York Avenue NW
Washington, DC 20001
Telephone: (202) 408-4000
Facsimile: (202) 408-4400

Matthew C. Berntsen (SDNY Bar No. MB8752)
matthew.berntsen@finnegan.com
Two Seaport Lane
Boston, MA 02210
Telephone: (617) 646-1600
Facsimile: (617) 646-1666

Attorneys for A.I. Progress, Inc.

CERTIFICATE OF COMPLIANCE

I certify that this brief contains 5,741 words, excluding the caption and title, table of contents, table of authorities, signature block, certificate of compliance, and certificate of service, as calculated by Microsoft Word, and complies with the Court's 6,000-word limit for proposed amicus briefs.

/s/ Anna B. Chauvet
Anna B. Chauvet (pro hac vice)

CERTIFICATE OF SERVICE

I, the undersigned, hereby certify that on October 30, 2026, a copy of the foregoing document was filed with the clerk of the Court using the CM/ECF system which will send notification to all counsel of record.

/s/ Anna B. Chauvet
Anna B. Chauvet (pro hac vice)