Sabita J. Soneji (SBN 224262)
TYCKO & ZAVAREEI LLP
1970 Broadway – Suite 1070
Oakland, CA 94612
Tel: (510) 254-6808
ssoneji@tzlegal.com

Attorney for Amici Curiae

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

RICHARD KADREY, et al.,

Individual and Representative Plaintiffs,

v.

META PLATFORMS, INC.,

Defendant.

Case No. No. 3:23-cv-03417-VC

AMICUS BRIEF OF COPYRIGHT LAW PROFESSORS

Table of Contents

Table of Authorities

Cases

Statutes

Other Authorities

STATEMENT OF INTEREST OF AMICI CURIAE

Amici, listed in the Appendix, are professors who teach and have written about copyright law and related subjects. Our sole interest in this case is in the development of copyright law in a way that serves the public interest.[1]

SUMMARY OF ARGUMENT

Meta's claim that its unauthorized copying of plaintiffs' works to train its large language models is fair use is a breathtaking request for greater legal privileges than courts have ever granted human authors. It should be rejected. The use of copyrighted works to train generative models is not "transformative," because using works for that purpose is not relevantly different from using them to educate human authors, which is a principal original purpose of all of plaintiffs' works. That training use is also not "transformative" because its purpose is to enable the creation of works that compete with the copied works in the same markets – a purpose that, when pursued by a for-profit company like Meta, also makes the use undeniably "commercial." Use of works to train large language models is also not a "nonexpressive use," because it incorporates the expressive choices of the authors of those works into the models. While "intermediate copying" to enable the discovery of functional software interfaces has justifiably been held to be a fair use, the copying at issue here has no such purpose. And while the measure of actual damages for copyright infringement importantly does not include the value of skills learned from infringing copies, limiting damages is crucially different from granting immunity from all liability. Under all of those circumstances, the case for fair use is weak.

[1] Counsel for the parties did not author this brief in whole or in part. No person other than Amici Curiae or their counsel contributed money to fund preparation or submission of this brief. The Amici Curiae are submitting this brief pursuant to the permission granted by Judge Chhabria in his order of April 1, 2025, Docket No. 513, Kadrey v. Meta Platforms, Inc., No. 3:23-cv-03417-VC.

ARGUMENT

I. META'S COPYING OF THE ENTIRETY OF PLAINTIFFS' WORKS TO IMPART CERTAIN ABILITIES OR SKILLS TO ITS LARGE LANGUAGE MODELS IS NOT TRANSFORMATIVE

Ever since the Supreme Court's decision in *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994), an important factor in fair use inquiry has been whether the copier's use is "transformative." *Id.* at 579. A use can be "transformative" either because it modifies a particular work in order to parody, comment on, or otherwise add new meaning to that particular work, *see id.*; or because it uses that work for a purpose different from the purpose for which it was created, and in a different market than the market for which the work was created. *See Authors Guild v. Google, Inc.*, 804 F.3d 202, 214-218 (2d Cir. 2015); *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 526, (2023); Rebecca Tushnet, *Content, Purpose, or Both?*, 90 Wash. L. Rev. 869, 869–70 (2015) (distinguishing between content-transformativeness" and "purpose-transformativeness") Meta's use is not "transformative" in any of the meanings of that term.

A. Meta's Copying Makes No Transformative Use of Plaintiffs' Works Because It Does Not Parody, Comment On, or Add New Meaning to Any Particular Work

The copying of reasonable portions of a copyright work can be a fair use if the result is an altered work that "adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message." *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994). However, Meta has copied plaintiffs' works in their entireties, and its purpose in creating its large language model is not to create a parody of any particular copyrighted work in the training set, nor to comment on or otherwise add new meaning to any particular work in that set. It would be the rarest of occasions on which an output of Llama contained enough of a

particular work in the training set to amount to a prima facie infringement of that work under 17 U.S.C. § 106, and yet would be saved by the fair use defense under 17 U.S.C. § 107 because it parodied, commented upon, or otherwise added new meaning to that training work. Thus, Meta's training use makes no "transformative use" of plaintiffs' works in that sense of the term.

B. Meta's Copying Does Not Have a Transformative Purpose Because a Central Purpose of Plaintiffs' Works Has Always Been to Increase Knowledge and Skill

The copying of a copyrighted work, even in its entirety, can also be a fair use if the copying "serves a new and different function" than that for which the work was originally created. *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 96 (2d Cir. 2014); *Authors Guild v. Google, Inc.*, 804 F.3d 202, 217 (2d Cir. 2015). Meta asserts that it is using the plaintiffs' works for a different purpose than that for which they were created. "The works that AI models are trained on . . . were created for expressive purposes," while "models use training data . . . to develop an entirely new and innovative service that . . . produces valuable new content." Meta Platforms, Inc., *Comments of Meta on U.S. Copyright Office's Notice of Inquiry on Artificial Intelligence & Copyright* [Docket No. 2023–06], at 19 (Oct. 30, 2023). Meta's fellow generative AI developer, OpenAI, elaborates that while works like plaintiffs' were created for their "standalone entertainment value," the generative AI developers are using those works for the different purpose of "helping computer programs learn the patterns inherent in human-generated media." OpenAI, *Comments of OpenAI on U.S. Copyright Office's Notice of Inquiry and Request for Comment* [Docket No. 2023-06], at 11–12 (Oct. 30, 2023). Those assertions are false, because they inaccurately describe both the original purpose of plaintiffs' works, and Meta's purpose in engaging in the copying.

Plaintiffs created their works, not just to "entertain" unchanging human beings, but to inform, educate, and enlighten us. Regardless of the subjective intent of a work's author in

creating a work, we human beings are always learning from works, and are always learning patterns that are spread across multiple works. Thus, the imputed purpose of those works must include the development of knowledge and skills.

Our human use of copyrighted works to gain knowledge or to acquire skills has never been sufficient for a finding of fair use. If it were, then courts would routinely find almost all unauthorized violations of the exclusive rights of copyright holders by humans to be fair use. The reason is that such use tends to increase knowledge or build skill in the person who views the movie, hears the song, or reads the book in much the same way as consuming more videos, songs, and books improves Llama's "knowledge" and its ability to produce substitutes for some human literary creativity. *See* David Atkinson, Unfair Learning: GenAI Exceptionalism and Copyright Law, https://ssrn.com/abstract=4975857 (2025). As the Supreme Court recognized in *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417 (1984), it is difficult to distinguish between some human use that produces knowledge and other human use that doesn't, because all human use tends to increase the knowledge or skill of the user. "A teacher who copies to prepare lecture notes is clearly productive. But so is a teacher who copies for the sake of broadening his personal understanding of his specialty." *Id.* at, 455 n. 40. Thus, the assertion that plaintiff's works were created only to entertain is false. The purpose of plaintiffs' works, and of all works authored by human beings, includes the acquisition of knowledge and skill.

If Meta's description of the original purpose of plaintiffs' works is much too narrow, its description of its own purpose is too broad and optimistic. Generating text, images, and other works at the request of human beings may be an "entirely new service" in the sense that it has never before been done by computers, but human beings have been providing that service for millennia. Indeed, human beings are much more tested and proven creators of new knowledge

than large language models. Over the period that copyright has protected authors, and even in the time since large language models have existed, humans have done immeasurably more than large language models to create new fields of study, solve complex novel problems, develop innovative solutions and insightful theories, and create new genres of music and art.[2] If the goal is to increase knowledge and promote the progress of science and useful arts, there is far stronger evidence to support a stance of giving humans free rein to use all copyrighted works than there is to allow a for-profit corporation building a large language model to do the same. Thus, when Meta and its supporters advocate for a virtually unrestricted right to reproduce and consume any copyrighted work in the name of knowledge production, they are making an argument that *anyone* should be allowed to consume *any* work without authorization or compensation to copyright holders.

Courts have never interpreted fair use so broadly, and for good reason. Such free-wheeling disregard for copyright protections would certainly undermine and destroy the incentives to continue progress in science and art over time. If human creators cannot share information without the assurance that others will not be able to benefit from the information unjustly–often to the detriment of the original creator–then they lose an incentive to create and share whatever they do create. Even if human creators have non-economic motivations for creating, the lack of copyright protection would deny them the opportunity to earn a living from creating, and many of them wouldn't be able to create as much, because they would have to earn

[2] See, e.g., Thomas Wolf, "The Einstein AI model," https://thomwolf.io/blog/scientific-ai.html:

> [R]eal scientific breakthroughs will come not from answering known questions, but from asking challenging new questions and questioning common conceptions and previous ideas. . . . In my opinion this is one of the reasons LLMs, while they already have all of humanity's knowledge in memory, haven't generated any new knowledge by connecting previously unrelated facts. They're mostly doing "manifold filling" at the moment - filling in the interpolation gaps between what humans already know, somehow treating knowledge as an intangible fabric of reality.

Wolf is the co-founder and Chief Science Officer of Hugging Face, Inc., "a company . . . that develops computation tools for building applications using machine learning." Wikipedia, "Hugging Face" (last accessed April 7, 2025)

a living by doing something else. The professional authors who are the named plaintiffs in this lawsuit, and thousands of others in the proposed plaintiff class, may write from love of writing, but without copyright, they could not have produced all of the books that Meta has copied without authorization. Thus, Meta's purpose in copying plaintiffs' works is not transformatively different from the purpose for which they were produced, and that purpose, which includes imparting knowledge and skill, cannot itself justify a blanket exemption from copyright liability.

Since not only Meta, but OpenAI, Google, Anthropic, xAI, and many others are creating generative AI models through unauthorized use of works protected by copyright, a decision in Meta's favor here will have the cascading effect of undermining the rights of copyright holders in every domain rapidly, with ease, and at a scale unimaginable just a few years ago.

C. Meta's Copying Does Not Have a Transformative Purpose Because Its Ultimate Goal Is to Enable the Creation of Works that Compete with Plaintiffs' Works

In its most recent fair use decision, the Supreme Court took pains to explain that the "transformativeness" inquiry is not just about "new meaning or purpose" in the abstract, but is also about whether the defendant's use will end up competing in a market with the plaintiff's work. In *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 535 (2023) ("*Warhol"*) the Supreme Court held that the Andy Warhol Foundation did not make a "transformative use" of the Andy Warhol silkscreen that copied from Lynn Goldsmith's photograph of the musician and composer Prince, because "[b]oth are portraits of Prince used in magazines to illustrate stories about Prince." *Id.* at 535. Even though the Warhol silkscreen and the Goldsmith photograph were not "perfect substitutes," the Warhol Foundation's licensing of the silkscreen "shared the objectives" of Goldsmith's photograph. *Id.* More recently, in *Thomson Reuters Enterprise Center. GMBH v. Ross Intelligence. Inc.*, No. 1:20-CV-613-SB, 2025 WL 458520 (D. Del. Feb. 11, 2025) ("*Ross Intelligence*"), the District of Delaware, citing *Warhol*,

held that Ross Intelligence's use of Westlaw headnotes to train an AI legal research tool was not transformative because their product was going to be competing with Westlaw. *Id.* at *7.

Here, Meta's ultimate purpose in copying plaintiff's works "shares the objective" of those works. Meta wants Llama to produce output that competes with its training works. It is not surprising that an online market for freelance writing experienced a marked decline in demand immediately after the November 30, 2022 public release of OpenAI's ChatGPT service. *See* Jin Liu, Xingchen Xu, Xi Nan, Yongjun Li, and Yong Tan, *"Generate" the Future of Work through AI: Empirical Evidence from Online Labor Markets*, https://arxiv.org/abs/2308.05201 (last revised June 6, 2024). Suddenly, ChatGPT, another large language model, was competing with the works that freelance writers had created or could create. Nor is it surprising that "stock" is the single most common word (other than prepositions and articles) in a set of 400 million image-text pairs used to train Stable Diffusion and other text-to-image models. *See* Robert Brauneis, *Copyright and the Training of Human Authors and Generative Machines*, 48 Colum. J. L. & Arts 1, 38, 38 n. 124 (2025). Text-to-image tools are being trained on the stock images that their outputs are most likely to replace.

To be sure, not every generative AI output competes with every work in the training set. In the end, though, the point is that all human authors have had to pay something, directly or indirectly, for most the works from which they have learned, because that is how the copyright system justifiably works. Meta now wants to compete against human authors, past and future, without having to pay for the works from which Llama learns, while the human authors continue to have to pay. That puts human authors at an unfair disadvantage entirely separate from whatever technical advances may make Llama faster and cheaper. Fair use should not be the means by which Meta achieves that unfair advantage.

II. META'S PURPOSE IN TRAINING LLAMA IS UNDENIABLY COMMERCIAL

The first enumerated fair use factor in § 107 of the Copyright Act is "the purpose and character of the use, including whether such use *is of a commercial nature* or is for nonprofit educational purposes," 17 U.S.C. § 107(1) (emphasis added), and thus it is not surprising that from its first fair use case to its most recent, the Supreme Court has held that the fact that a use is "commercial" weighs against fair use. *See, e.g., Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417, 448-450 (1984), *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 583–84 (1994); *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 531 (2023). Meta is a for-profit company, and its purpose in creating Llama is to make a profit. While there are many ways that Meta could create revenue from Llama, we know about at least one of them: Meta has revenue-sharing agreements with companies that host Llama services. *See* Kyle Wiggers, *Meta has revenue sharing agreements with Llama AI model hosts, filing reveals*, Techcrunch (March 21, 2025), https://techcrunch.com/2025/03/21/meta-has-revenue-sharing-agreements-with-llama-ai-model-hosts-filing-reveals/. To be sure, the Supreme Court has also held that when a use is "transformative," the fact that it is commercial matters less. *See Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994). In this case, however, as explained immediately above, Meta's ultimate purpose is not transformative, but is instead to train a tool to be able to generate output that competes with plaintiffs' copied works. Thus, Meta's use is commercial to the core.

III. META IS NOT MAKING "NONEXPRESSIVE USE" OF PLAINTIFFS' WORKS

Meta and its supporters argue that the models only train on the "nonexpressive" elements of the copyrighted works in the training set. This presents two problems for them to overcome. First, it is unclear why they believe that the way models "learn" from training data and produce

expressive outputs is for legal purposes meaningfully different from how a human does the same. Second, it incorrectly presumes that the models are only focused on nonexpressive elements.

As for the first, Meta's argument that it should be allowed to train on unauthorized works because it is just learning from the works' nonexpressive elements only makes sense if the process is legally distinct from how a human learns when exposed to works. Meta implies that humans learn from the expressive elements of, say, sentences, when reading a book, but when Llama consumes a book, it *only* learns from the non-expressive elements. That is, it suggests that humans are not learning to recognize patterns, understand context, or make associations between words, sentences, and paragraphs. But of course, as humans, we do all of that.

Words in isolation within a sea of text do not have meaning. If you open any book and read a single word, it will reveal nothing. It is only in context that the words have any significance. This context, in turn, is learned from repeated exposure. Reading is a skill, not an innate ability. To read a book, the brain must constantly make assumptions and predictions based on what the person has already read. If this is starting to sound like how models learn by deriving associations between tokens, it should.

In an effort to suggest enormous distance between the copyrighted training works and its model, Meta quickly resorts to abstract technical language in describing how its models learn from those works. Words are fragmented into "trillions of tokens"; "patterns and information" are "extracted" from those tokens; and they are then converted into "billions of numerical weights" in a complex model. Defendant Meta Platforms, Inc.'s (1) Notice of Motion and Motion for Partial Summary Judgment; and (2) Opposition to Plaintiff's Motion for Partial Summary Judgment at 13 (March 28, 2025). Yet that effort to build remoteness from the copyrighted training works fails. One might as well attempt to defend against a claim of human

copyright infringement by describing the infringer's experience and learning in similar abstract technical detail. The infringement defendant didn't "see" anything. Rather, many billions of photons hit the book's surface; some of those billions reached a lens, which focused them onto a retina, which converted them into electronic signals, which then resulted in electronic and chemical changes in some portion of over 100 billion neurons with over 100 trillion connections, some of those changes being transitory, and others more permanent. The technical description of human processing and learning is even more mysterious because not even expert specialists in neuroscience know how the brain works at the neuronal level.

The complexity and mystery of that description do not change the fact that the person was reading the book and was gaining linguistic knowledge and skills that they could later use to create their own literary works. In exactly the same way, the technical description of the construction of Llama does not change the fact that Meta's entire purpose is to have the model produce ordinary text that people can read, and that can substitute for text that human beings could have created. To hear Meta's argument, one would think that when we as humans read a book, we only see the expressive elements (whatever that may mean to Meta) and we only store expressive elements in our neurons, so when we summarize a book or create a derivative work we are relying on the expressive elements of the book we read, while Llama is only using the non-expressive elements. That is an attempted illusion, generated by shifting the level at which the processes are described.

Second, if "expression" means the creative choices made by authors of copyrighted works, then Meta's use of those works is for the purpose of exploiting the expressive elements of those works. *See* Robert Brauneis, *Copyright and the Training of Human Authors and Generative Machines*, 48 Colum. J. L. & Arts 1, 22-26, 30-35 (2025) (distinguishing between

different meanings of the term "expression," and noting that large language models and human authors both use the creative choices in the works from which they learn). Expression is the specific means through which meaning is conveyed, by means of selecting and arranging specific expressive elements. In other words, expression is not what you say, *it's how you say it.* It is which words you choose to use, in which order and with which juxtapositions. That is the only thing copyright unambiguously protects in a literary work. *See Harper & Row, Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539, 547 (1985) ("The copyright [in a work] is limited to those aspects of the work—termed 'expression'—that display the stamp of the author's originality."). And that is precisely what LLM training appropriates, incorporates into its models, and makes lucrative use of. What Meta wants isn't a model that "knows English," but rather a model that contains patterns taken from the way human authors actually express themselves. Indeed, Meta wants something even more specific than this, which is why it does not just train its models on human-authored works already in the public domain, which it could do without copyright concern. Meta wants a model containing the patterns used by human authors *in the here and now*. It wants a model capable of deciding, in response to a prompt, "What expression would the actual contemporary human authors who wrote the works in my training set use to respond to this?" That is why Meta chose to copy plaintiffs' works without authorization.

The case law Meta claims supports its position is unhelpful to it. Two of the cases -- *Authors Guild, Inc. v. Google, Inc.,* 804 F.3d 202 (2d Cir. 2015), and *Authors Guild, Inc. v. HathiTrust,* 755 F.3d 87 (2d Cir. 2014) -- are about Google Book Search, a product that enables users to find books that contain search terms or strings the users enter, and thus tells users something *about* the books. Unlike Llama, Google Book Search identifies particular books in which users might be interested, with only limited display of "snippets" of those books; often

provides links to facilitate purchase of the books by the user, and allows copyright owners to opt out of the display of any text of the books. *See Authors Guild v. Google,* 804 F.3d at 208-210. Similarly, in *A.V. v. iParadigms, LLC,* 562 F.3d 630 (4th Cir. 2009), the product -- a plagiarism checker -- simply determined how much of the text that was to be checked matched the text of some copyrighted work, thus revealing something *about* that copyrighted work.

In none of those cases was the technology focused on exploiting the expressive elements of the works for their expressiveness. Put differently, the expressiveness of the works was beside the point. The technology at issue was content agnostic. None of the products would function any differently if the book or paper scanned or uploaded to them consisted of random groupings of letters rather than actual words and sentences. Meta and other LLM developers, on the other hand, filter out such works because they rely on the expressiveness of the works.

IV. META IS NOT ENGAGING IN ANY ALLOWABLE "INTERMEDIATE COPYING" OF PLAINTIFFS' WORKS

Meta and its supporters also argue that its use of plaintiffs' works is fair because it is only "intermediate copying" that does not appear in any final output of Meta's models. Ironically, the case that Meta and its supporters most frequently cite for the principle that "intermediate copying" is always a noninfringing use, *Sega Enterprises Ltd. v. Accolade, Inc.,* 977 F.2d 1510 (9th Cir. 1992). *("Sega"),* explicitly and thoroughly repudiates that principle. In *Sega,* Accolade argued that "intermediate copying does not infringe the exclusive rights granted to copyright owners in section 106 of the Copyright Act unless the end product of the copying is substantially similar to the copyrighted work." *Id.* at 1517. The Ninth Circuit flatly rejected that argument. "Neither the language of the Act nor the law of this circuit supports" Accolade's arguments that intermediate copying is never infringing. *Id.* at 1518. "[T]he Copyright Act does not distinguish between unauthorized copies of a copyrighted work on the basis of what stage of the alleged

infringer's work the unauthorized copies represent. . The intermediate copying done by Accolade therefore falls squarely within the category of acts that are prohibited by the statute." *Id.*

Rather, both *Sega* and the subsequent Ninth Circuit decision in *Sony Computer Entertainment, Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000) ("*Connectix")*, stand for an important but much narrower principle of fair use. In both cases, the purpose of the copying was only to identify the functional elements in operating system software, which enabled human creators to then use their own creativity to develop games that could run on those operating systems. As the District of Delaware recently held in *Ross Intelligence*, 2025 WL 458520, at *8, "intermediate copying" outside of that context, such as the copies Ross Intelligence made to train an AI legal research tool, does not deserve the same fair use treatment.

Here, Meta is not training on books to learn functional properties so Meta can then provide all of the creative expression that Llama will use to produce its outputs. In fact, Meta provides *none* of the expressiveness the model depends on to be useful. Instead, Meta harvests all of that expression from works authored by others. In the case of the works authored by the plaintiffs in this case, it harvests that expression from works still under copyright and without authorization from the copyright owners. That falls squarely outside of the fair use principle developed in *Sega* and *Connectix*, and within the "intermediate copying" that *Sega* firmly establishes as infringement and *Ross Intelligence* confirms is undeserving of fair use treatment.[3]

[3] Closely related to the debunked "intermediate copying" escape hatch is the argument that training use is "internal copying" that for that reason alone avoids all infringement liability. *See* [Proposed] Amicus Brief of Intellectual Property Law Professors at 5 (filed March 31, 2025) (referring to "intermediate uses" as "internal copying"). Trademark law observers will recall that tech companies that sold keyword advertising claimed that their use of trademarks as keywords could not possibly ever infringe those trademarks, because their use of those trademarks was "internal." *See, e.g.,* Memorandum of Law in Support of Defendant Google Inc.'s Motion to Dismiss at 5-10, *Rescuecom Corp. v. Google Inc*., 5:04-cv-01055-NAM-GHL (N.D.N.Y., filed November 8, 2004). Yet the Second and Ninth Circuits both eventually rejected that attempt to short-circuit the inquiry into likelihood of confusion. *See Rescuecom Corp. v. Google Inc.*, 562 F.3d 123, 130 (2d Cir. 2009) (squarely rejecting the position that "an alleged infringer's use of a trademark in an internal software program insulates the alleged infringer from a charge of infringement, no matter how likely the use is to cause confusion in the marketplace"); *Network Automation, Inc. v.*

V. WHILE SOME OF THE VALUE THAT META DERIVES FROM ITS UNAUTHORIZED COPYING IS NOT WITHIN THE COPYRIGHT MEASURE OF DAMAGES, IT CANNOT CONVERT A LIMITATION ON DAMAGES INTO COMPLETE IMMUNITY FROM INFRINGEMENT LIABILITY

One of Meta's arguments boils down to the assertion that because some of the value that it is gaining from infringement is not within copyright law's measure of damages, there is no infringement – or, put more simply, because it doesn't owe everything, it owes nothing. The error in that logic is self-evident. A limitation on damages serves a very important role, but its purpose is not to abolish the tort altogether. If someone makes and reads an unauthorized copy of a textbook on bookkeeping and learns how to practice bookkeeping from that text, the actual damages in an infringement action do not include the earnings that the infringer makes from subsequently working as a bookkeeper. *See, e.g.,* Mark A. Lemley, *The Fruit of the Poisonous Tree in IP Law*, 103 Iowa L. Rev. 245, 260-262 (2017). That limitation is in place in part to separate the proper sphere of copyright law from that of patent law. As the Supreme Court held in *Baker v. Selden*, 101 U.S. 99 (1880), "the copyright of [a] treatise [does not] give the exclusive right to the art or manufacture described therein." *Id.* at 102. Yet that does not mean that making an unauthorized copy of a book for the purpose of learning the skill that it teaches results in no infringement and no damages at all. Indeed, in the very same case, the Supreme Court acknowledged that "a work on the subject of book-keeping . . . may be the subject of a copyright . . . and, considered . . . as the work of an author, conveying information on the subject of book-keeping, and containing detailed explanations of the art, it may be a very valuable acquisition to the practical knowledge of the community." *Id.* While the actual damages are

Advanced Systems Concepts, Inc., 638 F.3d 1137 (9th Cir. 2011) (also finding sufficient use in a keyword advertising case, and engaging in a likelihood of confusion inquiry). Similarly, in the present case, the attempt to deploy "internal copying" as a magic shield against copyright infringement liability should be rejected. Copying that takes place "internally," but with the purpose and effect of extracting expression for reuse in generative AI output, should be infringing copying.

limited to the profits that the copyright owner would have made from selling or licensing the book, or to the profits that the unauthorized copier did make from selling or licensing the infringing copy, *see* 17 U.S.C § 504 (b) (establishing the measures of actual damages for copyright infringement), the copier is still liable for copyright infringement. Meta's attempt to alchemically transmute a limitation on damages into a grant of immunity fails.

VI. CONCLUSION

Meta's use of plaintiffs' works to train its large language model is a commercial use that takes the expression in those works, that is not transformative, and that does not qualify as the kind of "intermediate copying" that courts have found to be a fair use. Meta's purpose is to create a model that will produce outputs that compete with the works on which it was trained, and with their authors. While the copyright actual measure of damages does not extend to all of the value Meta has created by copying, that should not immunize it from all liability. Taken together, these factors weigh conclusively against a finding of fair use.

Dated: April 11, 2025

By: *Sabita J. Soneji*
Sabita J. Soneji (SBN 224262)
TYCKO & ZAVAREEI LLP
1970 Broadway – Suite 1070
Oakland, CA 94612
Tel: (510) 254-6808
ssoneji@tzlegal.com

Attorney for Amici Curiae

APPENDIX[4]

David Atkinson
Assistant Professor of Instruction
The University of Texas at Austin

Jonathan M. Barnett
Torrey H. Webb Professor of Law
University of Southern California Gould School of Law

Robert Brauneis
Michael J. McKeon Professor of Intellectual Property Law
The George Washington University Law School

Jon M. Garon
Professor of Law
Nova Southeastern University Shepard Broad College of Law

Philippa Loengard
Executive Director of the Kernochan Center for Law, Media and the Arts, and Lecturer in Law
Columbia Law School

Adam Mossoff
Professor of Law
George Mason University Antonin Scalia Law School

Christopher M. Newman
Associate Professor of Law
George Mason University Antonin Scalia Law School

Trevor G. Reed
Charter Professor of Law
Arizona State University Sandra Day O'Connor College of Law

Zvi Rosen
Assistant Professor of Law
Southern Illinois University Simmons Law School

Bhamati Viswanathan
Assistant Professor of Law
New England Law

[4] Institutional affiliations are listed for identification purposes only.