Elizabeth Brannen (SBN 226234)
ebrannen@stris.com
Lauren Martin (SBN 294367)
lmartin@stris.com
STRIS & MAHER LLP
17785 Center Court Dr N, Ste 600
Cerritos, CA 90703
T: (213) 995-6800
F: (213) 261-0299

Counsel for Plaintiffs

Kyle Roche (*pro hac vice* forthcoming)
kroche@fnf.law
FREEDMAN NORMAND FRIEDLAND LLP
155 E. 44th Street, Ste 915
New York, NY 10017
T: (646) 494-2900

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

TOSCA RENO, DAVID BARKER, JEANETTE BICKNELL, MICHAEL BIEHL, MATTHEW DAVID BROZIK, ED BURYN, MATTHEW CALLAHAN, DANIEL CHESTER, GEORGE DEMARTINO, MICAELA DI LEONARDO, CHAD EASTHAM, DAVE EDLUND, MARK EDWARDS, GREGORY FACCONE, RUSSELL FIELDING, JOHN FOUBERT, ILAN FUCHS, GEORGE GALE, GIANFRANCO GIANFRATE, ERIKA GINSBERG-KLEMMT, WAYNE GLADSTONE, LYUBOMIR GRUYITCH, JOHN GUZLOWSKI, NICHOLAS HARTLEP, HENDRIK HARTOG, ELIZABETH HINTON, LISA JAHN-CLOUGH, JOSH KARP, KENNETH KNOBLAUCH, PATTI LATHER, NATHANIEL MARSHALL, CHRISTIE MELLOR, ALFRED MIERZEJEWSKI, BRETT MILLIER, MARK MONTROSE, JANET MULLANY, BRENDAN O'LEARY, ROBERT OSTERGREN, TOM POLAND, DIXA RAMIREZ, MICHAEL ROCQUE, YUVAL RON, HOWARD RUFFNER, SVEN SAALER, SCOTT SCHAFFER, AMY SCHAPIRO, SUZANNE SCHLOSBERG, ALEX SMITH, SAMANTH SUBRAMANIAN, WENFANG TANG, ERIC

Civil Case No.:

COMPLAINT

DEMAND FOR JURY TRIAL

VAN DE VELDE, PHILLIP WEICKER, MICHAEL WOLF-BRANIGIN, CHARLES YOUNG, VERSHAWN YOUNG, and NEAL ZASLAW,

Plaintiffs,

v.

META PLATFORMS, INC.,

Defendant.

1. Plaintiffs Tosca Reno, David Barker, Jeanette Bicknell, Michael Biehl, Matthew David Brozik, Ed Buryn, Matthew Callahan, Daniel Chester, George DeMartino, Micaela Di Leonardo, Chad Eastham, Dave Edlund, Mark Edwards, Gregory Faccone, Russell Fielding, John Foubert, Ilan Fuchs, George Gale, Gianfranco Gianfrate, Erika Ginsberg-Klemmt, Wayne Gladstone, Lyubomir Gruyitch, John Guzlowski, Nicholas Hartlep, Hendrik Hartog, Elizabeth Hinton, Lisa Jahn-Clough, Josh Karp, Kenneth Knoblauch, Patti Lather, Nathaniel Marshall, Christie Mellor, Alfred Mierzejewski, Brett Millier, Mark Montrose, Janet Mullany, Brendan O'Leary, Robert Ostergren, Tom Poland, Dixa Ramirez, Michael Rocque, Yuval Ron, Howard Ruffner, Sven Saaler, Scott Schaffer, Amy Schapiro, Suzanne Schlosberg, Alex Smith, Samanth Subramanian, Wenfang Tang, Eric Van de Velde, Phillip Weicker, Michael Wolf-Branigin, Charles Young, Vershawn Young, and Neal Zaslaw (collectively "Plaintiffs") bring this action against Meta Platforms, Inc. ("Meta" or "Defendant"), and allege as follows:

I. INTRODUCTION

2. This case concerns Defendant Meta's exploitation of Plaintiffs' copyrighted works to build its "Llama" and "Muse" families of large language models ("LLMs"). Defendant is in open and active competition with its tech company rivals to win what many have deemed the "generative AI arms race." Participants in this race are under immense pressure to build more, better, and faster, with an eye towards ensuring that *its* generative AI models are widely adopted. Broad adoption of a company's generative AI models will yield great profits, the thinking goes; on the converse, failure to move quickly means missing the boat on this trillion-dollar industry. To remain competitive in this race, Meta turned to notorious online "shadow libraries" and similar pirated datasets to obtain vast quantities of copyrighted materials—books, articles, training materials, and the like—which it desperately wanted to train and optimize its LLM models. Meta could have—*and should have*—paid copyright owners, including Plaintiffs, for licenses to use their copyrighted works in connection with training its Llama and Muse models. Instead, Meta fed from, and then back into, the dark market for digitally available copyrighted materials. Ultimately, Meta downloaded, reproduced, distributed, and defaced these copyrighted works, including Plaintiffs' Works, to get ahead in the

generative AI arms race and improve its bottom line. Plaintiffs bring this action to hold Meta accountable for its brazen acts of copyright infringement.[1]

3. LLMs are a form of "generative artificial intelligence" or "generative AI." These models are designed to process and emit natural language. Over the last several years, many of the world's biggest and wealthiest technology companies have entered the market to develop, distribute, and commercialize generative AI models, believing these and other forms of artificial intelligence have massive potential for growth in revenue and profits. Because artificial intelligence, including generative AI, is viewed by industry leaders as the next foundational layer of the digital economy, the pressure to build more, better, and faster has led to the aforementioned AI "arms race."

4. Against this backdrop, LLM developers needed data upon which to train their models. AI researchers quickly discovered that published content—*e.g.*, books, articles, and training materials—is really the gold standard when it comes to LLM training material. Unlike Internet content and many other forms of text, published material typically is structured, long-form, and highly polished. These materials embody the expressive output of their creators, are thoughtfully planned, and take weeks, months, and years to develop and publish. In simple terms, if you want to teach a machine how to "speak" or write like a human—capable of telling stories, using analogies, and making jokes—feed it material that mimics how we express ourselves in our most articulate and complete forms.

5. For its Llama training datasets, Meta needed gold standard training data. To get its hands on published content, Meta considered and explored the possibility of licensing copyrighted materials for use in its LLM training. The company's head of generative AI even discussed spending up to $100 million on licensing such materials. Ultimately, however, Meta abandoned this idea in favor of a faster and cheaper approach: downloading copyrighted materials from Internet-based "shadow libraries" and other datasets that contained the contents of these shadow libraries. These shadow libraries, each of which contained tens or hundreds of thousands of unauthorized digital

[1] Unless otherwise indicated, references to Meta's "Llama" and "Muse" refer to all versions of Llama and Muse, respectively, in any stage of their development or deployment.

copies of copyrighted works, had for years been the subject of criminal prosecutions, civil lawsuits, and widespread warnings within the technology industry. As such, by the time Meta turned to these sources to obtain training material for its Llama models, it knew or should have known that these "libraries" were the beneficiaries and propagators of copyright infringement.

6. To download copyrighted works from these shadow libraries, Meta often torrented them. "Torrenting" works by breaking a file into many small pieces and distributing those pieces across a network of participating computers. A user downloads portions of a file from numerous other computers that already possess the file and software reassembles those pieces into a completed whole on the user's machine.

7. But torrenting protocols not only facilitate downloading, they also facilitate and encourage uploading. And through Meta's torrenting of copyrighted material from illicit shadow libraries, Meta became a distributor of unauthorized copies of protected works. Stated somewhat differently, Meta used torrenting protocols that facilitated its reuploading of copyrighted materials into peer-to-peer file-sharing networks. In Copyright Act terms, Meta distributed copyrighted materials without consent or authorization. And through this distribution, Meta also contributed to further acts of infringement, allowing other users on these peer-to-peer networks to unlawfully reproduce and distribute copyrighted works.

8. In addition to ripping pirated copies, reproducing, and distributing without authorization Plaintiffs' copyrighted materials, Meta also stripped these materials of copyright management information, enabling, facilitating, and concealing the infringement of these works. Meta did this to optimize its models' performance.

9. Meta's infringement did not stop at shadow libraries. Beginning in or around mid-2025—after it had been sued repeatedly for training Llama on pirated books—Meta launched a book-scanning operation, referred to herein as "Project Vellum," through which it acquired physical copies of books en masse, scanned them into digital form, ran the scanned images through optical character recognition software to extract their text, and fed the resulting digital copies into the training of its next-generation "Muse" family of LLMs, including Muse Spark (released in April 2026) and Muse Glimmer (released in August 2026).

10. Meta did not seek, let alone obtain, a license to reproduce any of these books. Buying a single copy of a book conveys ownership of that one copy; it does not convey the copyright owner's exclusive right to reproduce the work. Yet that is exactly the right Meta arrogated to itself to build a commercial training library on the cheap. Meta's unauthorized reproduction of Plaintiffs' Works through Project Vellum is an independent act of infringement, separate and apart from its piracy of those same works from shadow libraries.

11. Adding insult to all of this injury, Meta's use of Plaintiffs' copyrighted materials facilitated Meta's creation of AI models capable of generating content that directly competes with and will compete directly with Plaintiffs' content. "Learning from" the creativity and expression embodied in thousands upon thousands of copyrighted works, including Plaintiffs' Works, Llama and Muse have the capability to flood the market with free and paid content that vies with Plaintiffs' for consumer attention.

12. The result of Meta's use of copyrighted materials to train Llama? Meta expects its Llama models to generate somewhere in the realm of $460 billion to $1.4 trillion in revenue. The Copyright Act doesn't allow Meta to get a free ride on the backs of Plaintiffs and the other creators whose copyrighted works it exploited to build its trillion-dollar generative AI models. Plaintiffs bring this action to hold Meta accountable for the infringement that enabled its rise in the generative-AI marketplace, and to enforce the fundamental principle that creative expression cannot be taken, copied, or exploited without permission or compensation.

13. To redress Meta's repeated, unlawful, and *en masse* infringement of their works, Plaintiffs seek (1) damages, (2) permanent injunctive relief barring Meta's ongoing infringement, and (3) any additional remedies the law provides.

14. Plaintiffs elect not to bring this case as a class action because the Copyright Act entitles them to recover individualized statutory damages, determined by a jury, for Meta's infringement and related conduct. Plaintiffs desire to retain full control of their case and avoid having their rights diluted by being swept into sprawling class-action settlements structured to resolve claims for pennies on the dollar. Recent history has shown that certain class actions and proposed settlement(s) seem to serve the tech conglomerate-infringers, not creators. LLM

companies should not be able to so easily extinguish thousands upon thousands of high-value claims at bargain-basement rates, eliding what should be the true cost of their massive willful infringement.

15. That is not how Plaintiffs plan to proceed. Under established Supreme Court precedent, "the amount of statutory damages is a question for the jury."[2] The Copyright Act thus vests authors with the right to have a jury evaluate the willfulness of infringement and assign a damages amount tailored to Meta's conduct.

16. In sum, the Copyright Act's statutory-damages and attorneys' fee regime empowers individual authors and publishers to hold infringers accountable without the need for class action treatment. That is what Plaintiffs have chosen to do.

II. PARTIES

A. Plaintiffs

17. Plaintiff Tosca Reno is an author who resides in Canada. Tosca Reno is the author of *THE START HERE DIET Three Simple Steps That Helped Me Transition from Fat to Slim...For Life* and *Your Best Body Now*. One or more of Reno's works are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Reno's works. Reno's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

18. Plaintiff David Barker is an author who resides in Maryland. David Barker is the author of *One Nation, Two Realities: Dueling Facts in American Democracy*; *Representing Red and Blue: How the Culture Wars Change the Way Citizens Speak and Politicians Listen*; *Rushed to Judgment: Talk Radio, Persuasion, and American Political Behavior*; and *The Politics of Truth in Polarized America*. One or more of Barker's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is

[2] *Feltner v. Columbia Pictures Television, Inc.*, 523 U.S. 340, 353 (1998).

accordingly a reasonable inference that Meta illegally downloaded Barker's works. Barker's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

19. Plaintiff Jeanette Bicknell is an author who resides in Canada. Jeanette Bicknell is the author of *Philosophy of Song and Singing: An Introduction*. Bicknell's work is contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Bicknell's work. Bicknell's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

20. Plaintiff Michael Biehl is an author who resides in Florida. Michael Biehl is the author of *Doctored Evidence: A Suspense Novel*; *Lawyered to Death*; *Nursing a Grudge*; and *Seven Mile Bridge*. One or more of Biehl's works are contained in pirated online libraries, including Z-Library and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Biehl's works. Biehl's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

21. Plaintiff Matthew David Brozik is an author who resides in New York. Matthew David Brozik is the author of *Danger ...with a Hard G*; *Odder Space*; *Spelled Wrong*; *Taking Ivy Seriously*; *The Admirable Vice Provost and Other Stories*; *The Government Manual for New Pirates*; *The Government Manual for New Superheroes*; *The Government Manual for New Wizards*; *The Vowels of the Earth*; *Twin Freaks*; *Whimsy & Soda*; and *Xavier, the Creative Little Gator*. One or more of Brozik's works are contained in pirated online libraries, including Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Brozik's works. Brozik's works were also published in print and were commercially available for

purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

22. Plaintiff Ed Buryn is an author who resides in California. Ed Buryn is the author of *The William Blake Tarot of the Creative Imagination*; *The William Blake Tarot of the Creative Imagination, 3rd Edition*; *Vagabonding in the U. S. A.: A Guide for Independent Travelers and Foreign Visitors*; and *Vagabonding in the U. S. A.: A Guide to Independent Travel*. One or more of Buryn's works are contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Buryn's works. Buryn's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

23. Plaintiff Matthew Callahan is an author who resides in Washington. Matthew Callahan is the author of *Shadowborne: Book One*. Callahan's work is contained in pirated online libraries, including Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Callahan's work. Callahan's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

24. Plaintiff Daniel Chester is an author who resides in Delaware. Daniel Chester is the author of *Optimal Automated Process Fault Analysis*. Chester's work is contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Chester's work. Chester's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

25. Plaintiff George DeMartino is an author who resides in Colorado. George DeMartino is the author of *Modern Macroeconomic Theories of Cycles and Crisis: A Methodological Critique*;

The Economist's Oath: On the Need for and Content of Professional Economic Ethics; *The Oxford Handbook of Professional Economic Ethics*; and *The Tragic Science: How Economists Cause Harm (Even as They Aspire to Do Good)*. One or more of DeMartino's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded DeMartino's works. DeMartino's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

26. Plaintiff Micaela Di Leonardo is an author who resides in Illinois. Micaela Di Leonardo is the author of *Black Radio/Black Resistance: The Life and Times of the Tom Joyner Morning Show*; the foreword to *Exotics at Home: Anthropologies, Others, American Modernity*; *Gender at the Crossroads of Knowledge: Feminist Anthropology in the Postmodern Era*; *New Landscapes of Inequality: Neoliberalism and the Erosion of Democracy in America*; *The Gender/sexuality Reader: Culture, History, Political Economy*; the foreword to *The People of the Abyss*; *The Varieties of Ethnic Experience*; and *The Varieties of Ethnic Experience: Kinship, Class, and Gender Among California Italian-Americans*. One or more of Di Leonardo's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Di Leonardo's works. Di Leonardo's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

27. Plaintiff Chad Eastham is an author who resides in Tennessee. Chad Eastham is the author of *Guys Are Waffles, Girls Are Spaghetti*; *Guys Like Girls Who*; *The Truth About Breaking Up, Making Up and Moving On*; *The Truth About Guys*; and *Truth About Dating, Love, and Just Being Friends*. One or more of Eastham's works are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is

accordingly a reasonable inference that Meta illegally downloaded Eastham's works. Eastham's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

28. Plaintiff Dave Edlund is an author who resides in Oregon. Dave Edlund is the author of *Crossing Savage*; *Deadly Savage*; *Hunting Savage: A Peter Savage Novel*; *Methanol Fuel Cell Systems*; and *Relentless Savage*. One or more of Edlund's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Edlund's works. Edlund's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

29. Plaintiff Mark Edwards is an author who resides in Michigan. Mark Edwards is the author of *Faith and Foreign Affairs in the American Century*. Edwards's work is contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Edwards's work. Edwards's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

30. Plaintiff Gregory Faccone is an author who resides in Florida. Gregory Faccone is the author of *Unwelcome Star*. Faccone's work is contained in pirated online libraries, including Z-Library and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Faccone's work. Faccone's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

31. Plaintiff Russell Fielding is an author who resides in South Carolina. Russell Fielding is the author of *The Wake of the Whale: Hunter Societies in the Caribbean and North Atlantic*. Fielding's work is contained in pirated online libraries, including LibGen and Z-Library. Meta has

directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Fielding's work. Fielding's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

32. Plaintiff John Foubert is an author who resides in Tennessee. John Foubert is the author of *Lessons Learned*; *Lessons Learned: How to Avoid the Biggest Mistakes Made by College Resident Assistants*; *Lessons Learned: How to Avoid the Biggest Mistakes Made by College Resident Assistants, 2nd Edition*; *Protecting Your Children from Internet Pornography, Understanding the Science, Risks, and Ways to Protect Your Kids*; *The Men's and Women's Programs: Ending Rape Through Peer Education*; *The Men's Program*; and *The Men's Program: How to Successfully Lower Men's Likelihood of Raping*. One or more of Foubert's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Foubert's works. Foubert's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

33. Plaintiff Ilan Fuchs is an author who resides in New York. Ilan Fuchs is the author of *Jewish Women's Torah Study*. Fuchs's work is contained in pirated online libraries, including LibGen. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Fuchs's work. Fuchs's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

34. Plaintiff George Gale is an author who resides in New York. George Gale is the author of *Dying on the Vine: How Phylloxera Transformed Wine* and *Theory of Science: An Introduction to the History, Logic, and Philosophy of Science*. One or more of Gale's works are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, and Z-Library.

Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Gale's works. Gale's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

35. Plaintiff Gianfranco Gianfrate is an author who resides in Massachusetts. Gianfranco Gianfrate is the author of *Corporate Valuation*. Gianfrate's work is contained in pirated online libraries, including Z-Library and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Gianfrate's work. Gianfrate's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

36. Plaintiff Erika Ginsberg-Klemmt is an author who resides in Florida. Erika Ginsberg-Klemmt is the author of *The Complete Anchoring Handbook: Stay Put on Any Bottom in Any Weather*. Ginsberg-Klemmt's work is contained in pirated online libraries, including LibGen. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Ginsberg-Klemmt's work. Ginsberg-Klemmt's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

37. Plaintiff Wayne Gladstone is an author who resides in New York. Wayne Gladstone is the author of *Agents of the Internet Apocalypse*; *Notes from the Internet Apocalypse*; and *Reports on the Internet Apocalypse*. One or more of Gladstone's works are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Gladstone's works. Gladstone's works were also published in print and were commercially available

for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

38. Plaintiff Lyubomir Gruyitch is an author who resides in France. Lyubomir Gruyitch is the author of *Advances in the Linear Dynamic Systems Theory*; *Einstein's Realtivity Theory: Correct, Paradoxical and Wrong*; *Linear Continuous-Time Systems*; *Nonlinear Systems Tracking*; *Stability Domains*; *Time and Time Fields: Modeling, Relativity, and Systems Control*; *Time: Fields, Relativity and Systems*; *TIME. TRUE RELATIVITY Physical and Mathematical Bases*; and *Tracking Control of Linear Systems*. One or more of Gruyitch's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Gruyitch's works. Gruyitch's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

39. Plaintiff John Guzlowski is an author who resides in Virginia. John Guzlowski is the author of *Echoes of Tattered Tongues: Memory Unfolded* and *Suitcase Charlie*. One or more of Guzlowski's works are contained in pirated online libraries, including Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Guzlowski's works. Guzlowski's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

40. Plaintiff Nicholas Hartlep is an author who resides in Kentucky. Nicholas Hartlep is the author of *A Segmented Assimilation Theoretical Study of the 2002 Asian American Student Population*; *Asian/Americans, Education, and Crime: The Model Minority as Victim and Perpetrator*; *Pay to Play: Race and the Perils of the College Sports Industrial Complex*; *The Neoliberal Agenda and the Student Debt Crisis in U.S. Higher Education*; and *What Makes a Star Teacher: 7 Dispositions That Support Student Learning*. One or more of Hartlep's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has

directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Hartlep's works. Hartlep's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

41. Plaintiff Hendrik Hartog is an author who resides in New Jersey. Hendrik Hartog is the author of *American Public Life and the Historical Imagination*; *Law in Everyday Life*; *Law in the American Revolution and the Revolution in the Law: A Collection of Review Essays on American Legal History*; *Man and Wife in America: A History*; *Public Property and Private Power: The Corporation of the City of New York in American Law, 1730-1870*; *Someday All This Will Be Yours: A History of Inheritance and Old Age*; *The Properties of the Corporation: New York City and Its Law, 1730*; and *The Trouble with Minna: A Case of Slavery and Emancipation in the Antebellum North*. One or more of Hartog's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Hartog's works. Hartog's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

42. Plaintiff Elizabeth Hinton is an author who resides in Connecticut. Elizabeth Hinton is the author of *America on Fire: The Untold History of Police Violence and Black Rebellion Since the 1960s* and *From the War on Poverty to the War on Crime: The Making of Mass Incarceration in America*. One or more of Hinton's works are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Hinton's works. Hinton's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

43. Plaintiff Lisa Jahn-Clough is an author who resides in Maine. Lisa Jahn-Clough is the author of *123 Yippie*; *A Big Bed for Jed*; *Alicia's Best Friends*; *Country Girl, City Girl*; *Felicity & Cordelia: A Tale of Two Bunnies*; *Me, Penelope*; *Missing Molly*; *My Friend and I*; *My Happy Birthday Book*; *Nothing but Blue*; *On the Hill*; *Simon and Molly Plus Hester*; and *We're Painting*. One or more of Jahn-Clough's works are contained in pirated online libraries, including Z-Library and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Jahn-Clough's works. Jahn-Clough's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

44. Plaintiff Josh Karp is an author who resides in Illinois. Josh Karp is the author of *A Futile and Stupid Gesture: How Doug Kenney and National Lampoon Changed Comedy Forever*; *Orson Wells's Last Movie*; and *Straight Down the Middle: Shivas Irons, Bagger Vance, and How I Learned to Stop Worrying and Love My Golf Swing*. One or more of Karp's works are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Karp's works. Karp's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

45. Plaintiff Kenneth Knoblauch is an author who resides in France. Kenneth Knoblauch is the author of *Modeling Psychophysical Data in R*. Knoblauch's work is contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Knoblauch's work. Knoblauch's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

46. Plaintiff Patti Lather is an author who resides in Ohio. Patti Lather is the author of *Engaging Science: Policy from the Side of the Messy*; *Getting Lost: Feminist Efforts Toward a Double(d) Science*; and *Getting Smart: Feminist Research and Pedagogy With/in the Postmodern*. One or more of Lather's works are contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Lather's works. Lather's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

47. Plaintiff Nathaniel Marshall is an author who resides in Wisconsin. Nathaniel Marshall is the author of *FINNA Poems*; *Out South*; and *Wild Hundreds*. One or more of Marshall's works are contained in pirated online libraries, including Z-Library and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Marshall's works. Marshall's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

48. Plaintiff Christie Mellor is an author who resides in Washington. Christie Mellor is the author of *Coco Brighton Invites You to a Theatrical Extravaganza*; *Fun Without Dick and Jane: Your Guide to a Delightfully Empty Nest*; *The Three-Martini Playdate: A Practical Guide to Happy Parenting*; *Were You Raised by Wolves*; and *You Look Fine, Really*. One or more of Mellor's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Mellor's works. Mellor's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

49. Plaintiff Alfred Mierzejewski is an author who resides in Texas. Alfred Mierzejewski is the author of *A History of the German Public Pension System: Continuity Amid Change*; *Ludwig Erhard: A Biography*; *The Collapse of the German War Economy, 1944-1945: Allied Air Power and the German National Railway*; *The Most Valuable Asset of the Reich*; *The Most Valuable Asset of the Reich: A History of the German National Railway: Vol. 1, 1920-1932*; and *Wheels Must Roll for Victory: Allied Air Power and the German War*. One or more of Mierzejewski's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Mierzejewski's works. Mierzejewski's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

50. Plaintiff Brett Millier is an author who resides in Vermont. Brett Millier is the author of *Elizabeth Bishop: Life and the Memory of It*; *Elusive Mastery: The Drafts of Elizabeth Bishop's "One Art"*; *Flawed Light: American Women Poets and Alcohol*; *Modesty and Morality: George Herbert, Gerard Manley Hopkins, and Elizabeth Bishop*; and *Selected Poems: 1950-2012*. One or more of Millier's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Millier's works. Millier's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

51. Plaintiff Mark Montrose is an author who resides in California. Mark Montrose is the author of *EMC and the Printed Circuit Board: Design, Theory, and Layout Made Simple*; *Printed Circuit Board Design Techniques for EMC Compliance*; *Printed Circuit Board Design Techniques for EMC Compliance, 2nd Edition*; and *Testing for EMC Compliance: Approaches and Techniques*. One or more of Montrose's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally

contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Montrose's works. Montrose's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

52. Plaintiff Janet Mullany is an author who resides in Maryland. Janet Mullany is the author of *A Certain Latitude*; *Dedication*; *Forbidden Shores*; *Hidden Paradise*; *Jane and the Damned*; *Jane Austen: Blood Persuasion*; *Little to Hex Her*; *Tell Me More*; *The Malorie Phoenix*; and *The Rules of Gentility*. One or more of Mullany's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Mullany's works. Mullany's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

53. Plaintiff Brendan O'Leary is an author who resides in Pennsylvania. Brendan O'Leary is the author of *Divided Nations and European Integration*; *How to Get Out of Iraq with Integrity*; *Power Sharing in Deeply Divided Places*; *Terror, Insurgency, and Re State: Ending Protracted Conflicts*; *The Future of Kurdistan in Iraq*; and *Theories of the State: The Politics of Liberal Democracy*. One or more of O'Leary's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded O'Leary's works. O'Leary's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

54. Plaintiff Robert Ostergren is an author who resides in Wisconsin. Robert Ostergren is the author of *A Community Transplanted: The Trans-Atlantic Experience of a Swedish Immigrant Settlement in the Upper Middle West, 1835-1915*; *Building Nazi Germany: Place, Space, Architecture, and Ideology*; *Europeans, Second Edition: A Geography of People, Culture, and*

Environment[3]; *The Europeans: A Geography of People, Culture and Environment*; and *Wisconsin Land and Life*. One or more of Ostergren's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Ostergren's works. Ostergren's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

55. Plaintiff Tom Poland is an author who resides in South Carolina. Tom Poland is the author of *Carolina Bays: Wild, Mysterious, and Majestic Landforms* and *Reflections of South Carolina*. One or more of Poland's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Poland's works. Poland's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

56. Plaintiff Dixa Ramirez is an author who resides in Rhode Island. Dixa Ramirez is the author of *Colonial Phantoms: Belonging and Refusal in the Dominican Americas, from the 19th Century to the Present* and *Dominicanas Presentes: Gender, Migration, and History's Legacy in Dominican Literature*. One or more of Ramirez's works are contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Ramirez's works. Ramirez's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

[3] This title reflects an error included in the U.S. Copyright Office registration data; the correct title is *The Europeans, Second Edition: A Geography of People, Culture, and Environment*.

57. Plaintiff Michael Rocque is an author who resides in Maine. Michael Rocque is the author of *Crime Prevention: Programs, Policies, and Practices*; *Fitting the Facts of Crime: An Invitation to Biopsychosocial Criminology*; *The Criminal Brain, Second Edition*; and *Understanding the Relationship Between Maturation and Desistance from Crime: A Life-Course Developmental Approach*. One or more of Rocque's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Rocque's works. Rocque's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

58. Plaintiff Yuval Ron is an author who resides in California. Yuval Ron is the author of *Divine Attunement: Music as a Path to Wisdom* and *Flirt with the Divine: Essays and Tales on Music and Spirituality*. One or more of Ron's works are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Ron's works. Ron's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

59. Plaintiff Howard Ruffner is an author who resides in California. Howard Ruffner is the author of *Moments of Truth: A Photographer's Experience of Kent State 1970*. Ruffner's work is contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Ruffner's work. Ruffner's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

60. Plaintiff Sven Saaler is an author who resides in Japan. Sven Saaler is the author of *Pan-Asianism: A Documentary History / Volume 1: 1850-1920*; *Pan-Asianism: A Documentary History / Volume 2: 1920-Present*; and *Routledge Handbook of Modern Japanese History*. One or

more of Saaler's works are contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Saaler's works. Saaler's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

61. Plaintiff Scott Schaffer is an author who resides in Canada. Scott Schaffer is the author of *Resisting Ethics*. Schaffer's work is contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Schaffer's work. Schaffer's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

62. Plaintiff Amy Schapiro is an author who resides in Maryland. Amy Schapiro is the author of *Fenwick: A Feisty Favorite* and *Millicent Fenwick: Her Way*. One or more of Schapiro's works are contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Schapiro's works. Schapiro's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

63. Plaintiff Suzanne Schlosberg is an author who resides in Oregon. Suzanne Schlosberg is the author of *Fitness for Dummies*; *Fitness for Travelers: The Ultimate Workout Guide for the Road*; *It's No Accident: Breakthrough Solutions to Your Child's Wetting, Constipation, UTIs, and Other Potty Problems*; *Kathy Smith's Fitness Makeover: A 10-Week Guide to Exercise and Nutrition That Will Change Your Life*; *More Than a Workout Log*; *Quit Smoking for Life*; *The Active Woman's Pregnancy Log: A Day-by-Day Diary and Guide to a Fit and Healthy Pregnancy*; *The Curse of the Singles Table: A True Story of 1001 Nights Without Sex*; *The Essential Breastfeeding Log: A Feedings Tracker and Baby-Care Organizer for Nursing Moms*; *The Essential Fertility Log*;

The Fat-Free Truth: 239 Real Answers to the Fitness and Weight-Loss Questions You Wonder About Most; *The Good Neighbor Cook Book*; *The Ultimate Diet Log: A Unique Food and Exercise Diary That Fits Any Weight-Loss Plan*; *The Ultimate Workout Log, Fourth Edition*; *The Ultimate Workout Log: An Exercise Diary and Fitness Guide*; *The Ultimate Workout Log: An Exercise Diary for Everyone*; *Weight Training for Dummies*; and *Weight Training for Dummies*®. One or more of Schlosberg's works are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Schlosberg's works. Schlosberg's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

64. Plaintiff Alex Smith is an author who resides in the United Kingdom. Alex Smith is the author of *Confessions of a Hypnotist: Everything You Ever Wanted to Know About Hypnosis but Were Afraid to Ask*. Smith's work is contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Smith's work. Smith's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

65. Plaintiff Samanth Subramanian is an author who resides in the United Kingdom. Samanth Subramanian is the author of *Following Fish* and *The Divided Island*. One or more of Subramanian's works are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Subramanian's works. Subramanian's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

66. Plaintiff Wenfang Tang is an author who resides in Hong Kong, China. Wenfang Tang is the author of *Party Intellectuals' Demands for Reform in Contemporary China*; *Populist Authoritarianism: Chinese Political Culture and Regime Sustainability*; *Proud and Angry: Political Culture in Post-British Hong Kong*; *Public Opinion and Political Change in China*; and *Social Change in Contemporary China: C. K. Yang and the Concept of Institutional Diffusion*. One or more of Tang's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Tang's works. Tang's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

67. Plaintiff Eric Van de Velde is an author who resides in California. Eric Van de Velde is the author of *Concurrent Scientific Computing*. Van de Velde's work is contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Van de Velde's work. Van de Velde's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

68. Plaintiff Phillip Weicker is an author who resides in California. Phillip Weicker is the author of *A Systems Approach to Lithium-Ion Battery Management*. Weicker's work is contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Weicker's work. Weicker's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

69. Plaintiff Michael Wolf-Branigin is an author who resides in Washington, D.C.. Michael Wolf-Branigin is the author of *Measuring Community Inclusion of People with Developmental Disabilities in a Spatial Context* and *Using Complexity Theory for Research and*

Program Evaluation. One or more of Wolf-Branigin's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Wolf-Branigin's works. Wolf-Branigin's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

70. Plaintiff Charles Young is an author who resides in Arkansas. Charles Young is the author of *Name, Rank, and Serial Number: Exploiting Korean War Pows at Home and Abroad*. Young's work is contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Young's work. Young's work was also published in print and was commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

71. Plaintiff Vershawn Young is an author who resides in Canada. Vershawn Young is the author of *Code-Meshing as World English: Pedagogy, Policy, Performance*; *From Bourgeois to Boojie: Black Middle-Class Performances*; *Neo-Passing: Performing Identity After Jim Crow*; *Other People's English: Code-Meshing, Code-Switching, and African American Literacy*; and *Your Average Nigga: Performing Race, Literacy and Masculinity*. One or more of Young's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Young's works. Young's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

72. Plaintiff Neal Zaslaw is an author who resides in New York. Neal Zaslaw is the author of *Mozart's Piano Concertos: Text, Context, Interpretation*; *The Birth of the Orchestra: History of an Institution, 1650-1815*; *The Classical Era—from the 1740s to the End of the 18th*

Century; *The Compleat Mozart: A Guide to the Musical Works of Wolfgang Amadeus Mozart*; and *The Mozart Repertory: A Guide for Musicians, Programmers, and Researchers*. One or more of Zaslaw's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Zaslaw's works. Zaslaw's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

73. A non-exhaustive list of registered copyrights owned by Plaintiffs is included as Exhibit A (herein, the "Plaintiffs' Works").[4]

B. Defendant

74. Defendant Meta Platforms, Inc. is a Delaware corporation with its principal place of business in Menlo Park, California. Meta develops and distributes the Llama series of LLMs, including Llama 1, Llama 2, Llama 3, and Llama 4, which were trained using datasets sourced in part from shadow libraries and other datasets containing pirated books. Meta also develops and distributes the Muse series of LLMs, including Muse Spark and Muse Glimmer, which were trained using, among other things, unauthorized digital copies of physical books that Meta scanned as part of Project Vellum.

III. JURISDICTION AND VENUE

75. This action arises under the Copyright Act of 1976, 17 U.S.C. § 101 *et seq*. This Court has subject-matter jurisdiction under 28 U.S.C. §§ 1331 and 1338(a) because Plaintiffs assert claims exclusively under federal copyright law.

76. This Court has personal jurisdiction over the Defendant. The Defendant has purposefully availed itself of the privilege of conducting business in this District and the State of California. The Defendant committed acts of copyright infringement in this District, directed

[4] Even where other individuals or entities are listed as copyright claimants on the relevant copyright registrations, Plaintiffs are legal or beneficial owners of all copyrights listed in Exhibit A.

conduct toward this District, or knowingly caused harm that was suffered in this District. The Defendant maintains substantial, continuous, and systematic contacts with this District.

77. Venue is proper in this District under 28 U.S.C. § 1400(a) because the Defendant or its agents resides or may be found in this District as a result of the infringing acts alleged herein. Venue is also proper under 28 U.S.C. § 1391(b)(2) because a substantial part of the events giving rise to Plaintiffs' claims—including the acquisition of pirated copies of Plaintiffs' Works, the acquisition, scanning, and digitization of physical copies of Plaintiffs' Works, the reproduction and ingestion of those copies into Defendant's training pipelines, the training and fine-tuning of the relevant LLMs, and the commercialization of the resulting models—occurred in this District.

IV. FACTUAL ALLEGATIONS

A. Plaintiffs and their Protected Works

78. Plaintiffs' Works are published books. Each was conceived, researched, written, and revised by its author over weeks, months, or years, and each was published and offered for sale to the reading public. Whatever their genre or subject matter, the works embody their authors' original creative choices—of structure, organization, selection, emphasis, tone, and expression—and constitute precisely the kind of structured, long-form, polished prose that, as alleged below, Meta and other LLM developers prize as training data. Collectively, the works reflect substantial original authorship intended to inform, educate, entertain, and engage readers.

79. Plaintiffs have either legal or beneficial ownership of the rights to reproduce, adapt, and distribute Plaintiffs' Works.

80. According to publicly available metadata, Plaintiffs' Works are contained in pirated online libraries and datasets including Books3 (and thus *The Pile*), Library Genesis or "LibGen," Z-Library, and Anna's Archive. As alleged below, Meta has directly or indirectly downloaded works contained in *The Pile*, LibGen, Z-Library, and Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded and reproduced Plaintiffs' Works.

81. Plaintiffs' Works were also published in print. On information and belief, physical copies of Plaintiffs' Works were commercially available for purchase—new or used—during the period in which Meta conducted Project Vellum. As alleged below, Meta acquired physical books

at scale, across every genre of published work, for the purpose of scanning and digitizing them as training data, and there is accordingly a reasonable inference that Meta acquired, scanned, and reproduced Plaintiffs' Works as part of that operation.

B. LLMs and the Generative AI "Arms Race"

82. "Generative artificial intelligence" or "generative AI" refers to systems and models that create outputs—such as text or images—that simulate human expression, often in response to user prompts.

83. Large language models are a form of generative AI, which are designed to process—or "understand"—and generate natural language. At a high level, LLMs operate by studying and "learning" the statistical relationship between words and other text (such as punctuation marks); upon further refinement by their developers, the models then process complex mathematical sequences to "predict" sequences of text based on the statistical relationships it has learned and been trained to recognize.

84. Development of LLMs requires, among other things, "training" the model on data—*i.e.*, the model's inputs. Training "requires identifying a formal measure or 'objective' for how well the model performs, and then repeatedly adjusting the model's parameters based on that objective as the model is exposed to training data."[5] While not all LLM developers use the same terminology, developers commonly reference a "pre-training" process—"in which a massive amount of computing power and data is spent to teach the model the broad foundations of language, grammar, and reasoning"—and a "post-training" or "fine-tuning" process "where the pre-trained model is further trained on a (relative to pre-training) smaller amount of carefully curated data of specific tasks."[6] For purposes of this Complaint, "training" refers to all stages of the LLM training process.

85. LLMs typically are trained, at least initially, by feeding them massive amounts of text data from which they can "learn" statistical relationships between and among text. In the process

[5] U.S. Copyright Office, *Copyright and Artificial Intelligence, Part 3: Generative AI Training Pre-Publication*, at 17 (May 2025), https://perma.cc/EY5U-EFUY.

[6] *Id.*

of compiling training data and using it during the training process, LLM developers as a matter of course typically create multiple copies of "raw" datasets. So, for instance, if an LLM developer downloads a copy of a digital book from some Internet-based source (*e.g.*, a shadow library), the developer typically will create a new "cleaned" copy of the book (*e.g.*, stripped of certain undesirable data), deduplicate that book against other potential copies, compile that book with other data to create a new compilation (or compilations), and store that book in multiple locations. In short, with respect to training data, the LLM development pipeline typically involves numerous reproductions of any given piece of the dataset.

86. A model's inputs—the training datasets—are directly tied to the model's performance. As one AI researcher has now famously said, a "model['s] behavior is not determined by architecture, hyperparameters, or optimizer choices. It's determined by your dataset, nothing else."[7]

87. In determining the corpus of training data for an LLM, a model's developer typically considers "the quantity of data, its quality, and the ultimate purpose(s) of the model."[8]

88. In terms of data quantity, the LLMs that have been developed and are being developed by leading technology companies are trained on enormous datasets, typically on the scale of terabytes. This is because AI researchers have found that "increasing the quantity of training data typically increases a model's 'performance.'"[9]

89. In terms of data quality, "[r]ecent research from major developers suggests that quality may even be a more important consideration than quantity."[10] "Garbage in, garbage out," the saying goes. AI researchers quickly discovered that published content—books in particular—makes for some of the best training material for LLMs. Unlike much content on the Internet and

[7] James Betker, *The "it" in AI Models is the Dataset*, (June 10, 2023), https://perma.cc/ZCH9-H53S.

[8] U.S. Copyright Office, *Copyright and Artificial Intelligence, Part 3: Generative AI Training Pre-Publication*, at 9.

[9] *Id.* at 10.

[10] *Id.* at 11.

many other forms of text, published material typically is structured, long-form, and highly polished. Published materials provide formal, extended prose that teaches models narrative structure, complex syntax, and coherent storytelling.

90. That books make for particularly valuable training material for LLMs is evidenced by the fact that several—if not most—of the major LLM developers either have paid for or expressly contemplated paying (large sums) for licenses to use copyrighted works to train their LLMs. Indeed, in an *amicus* brief filed in *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-3417 (N.D. Cal.), the Association of American Publishers cited *dozens* of "AI licensing deals for textual works," including deals involving Amazon, Microsoft, OpenAI, and Perplexity.[11] And, as the Court found in that litigation, Meta considered spending up to $100 million to license copyrighted materials in connection with LLM training.[12]

91. As alleged below, Meta's employees reached the same conclusion as other AI researchers—that published materials, and books in particular, were an essential ingredient for developing a high-performing LLM.

92. As discussed in the introduction, technology companies have treated generative AI as the next foundational layer of the digital economy. Industry leaders publicly describe an "AI arms race," in which they have redirected their corporate strategies to seize control of what they believe will become a new infrastructure layer for commerce, communication, and knowledge work.[13] For these companies, staying ahead of competitors is "code red."[14] Among other things, being too slow out of the blocks could mean ending up in last place; if by the time you publicly released *your* LLM

[11] *Kadrey v. Meta Platforms, Inc.*, ECF No. 535 at 12, Case No. 3:23-cv-3417-VC (N.D. Cal.) (Apr. 11, 2025).

[12] *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1040 (N.D. Cal. 2025).

[13] *See* Dr. Peter Asaro, *What is an 'Artificial Intelligence Arms Race' Anyway?*, 15 I/S: J.L. & Pol'y for Info. Soc'y 45 (2019).

[14] *See* Sharon Goldman, *Sam Altman declares 'Code Red' as Google's Gemini surges—three years after ChatGPT caused Google CEO Sundar Pichai to do the same*, FORTUNE (Dec. 2, 2025, 11:43 AM), https://perma.cc/J9MS-UQD2.

model, the public writ large and private industry had already adopted *other companies'* models, there would be a real risk that your model would be left on the shelf.

93. Like other participants in this race, Meta risked falling behind early industry leaders like OpenAI. Indeed, market analysts consistently reported that Meta had, in fact, fallen behind, with many pointing to Meta's intense focus on investment in the "Metaverse."[15]

94. Faced with the risks outlined above, Meta needed to move fast to develop and roll out Llama, its capstone LLM. And to do that, Meta needed to get its hands on high-quality training data.

C. Shadow Libraries, *The Pile*, and Torrenting

95. For years now, there have been illicit, online marketplaces for digital copies of books. These "shadow libraries," as they are commonly known, systematically acquire, store, index, and disseminate full-fidelity digital copies of copyrighted books. These online repositories maintain searchable catalogs, metadata, mirrors, bulk-download mechanisms, and—critically—complete downloadable archives designed to allow third parties to replicate the entire collection.[16] And they do all of this without authorization from authors or publishers.[17] In short, shadow libraries facilitate the reproduction and distribution of unauthorized copies of copyrighted works at industrial scale.

96. These libraries are widely known within both piracy communities and the technology sector as illegal sources of copyrighted books. Many have been the subject of criminal prosecutions, civil injunctions, domain seizures, and formal designation as "notorious piracy markets" by United

---

[15] Mike Proulx, *Meta: So Long, Metaverse; Hello, Superintelligence*, FORRESTER (July 30, 2025), https://perma.cc/BZ57-CR2D; Karen Hao, et al., *Mark Zuckerberg Was Early in AI. Now Meta Is Trying to Catch Up.*, WALL STREET JOURNAL (June 17, 2023, 12:00 AM), https://perma.cc/G548-B5RD.

[16] *See* Letter from Mary E. Rasenberger, CEO, Authors Guild, & Umair Kazi, Dir. of Pol'y & Advocacy, Authors Guild, to Daniel Lee, Assistant U.S. Trade Representative for Innovation & Intellectual Prop., Office of the U.S. Trade Representative (Oct. 7, 2022), https://perma.cc/XM4R-NDN3.

[17] *See* Riddhi Setty, *Rampant 'Shadow Libraries' Drive Calls for Anti-Piracy Action*, BLOOMBERG LAW (Oct. 19, 2022, 9:03 AM), https://perma.cc/F5VH-3BA6; Claire Woodcock, '*Shadow Libraries' Are Moving Their Pirated Books to the Dark Web After Fed Crackdowns*, VICE (Nov. 30, 2022, 11:38 AM), https://perma.cc/K9FA-VLPW.

States trade authorities. As centralized shadow libraries increasingly faced enforcement actions, including the seizure of domains, third parties responded by creating full mirrored copies of those repositories for decentralized redistribution.[18]

97. In 2018, an OpenAI employee downloaded pirated copies of books from Library Genesis, or "LibGen"—a shadow library repeatedly enjoined by federal courts[19]—and used those books to create two internal datasets OpenAI called "LibGen1" and "LibGen2," which OpenAI publicly referred to as "Books1" and "Books2."[20] OpenAI used those pirated corpora to train GPT-3, which it released in June 2020 to widespread commercial acclaim. OpenAI's use of a books corpora from a pirate library established the model for how to quickly and cheaply obtain published materials for use as LLM training data.

98. Within weeks of GPT-3's release, an open research collective, EleutherAI, formed with the express goal of replicating GPT-3's capabilities and democratizing access to large-scale language modeling. To do so, EleutherAI assembled and publicly released *The Pile*, a large (800+ gigabyte) general-purpose training dataset expressly designed to be downloaded, used, and incorporated into LLMs by academic researchers, startups, and commercial AI developers.[21]

99. Because OpenAI did not disclose the precise makeup of its training datasets, members of EleutherAI constructed a book corpus of their own: "Books3," consisting of

[18] Woodcock, *'Shadow Libraries' Are Moving Their Pirated Books to the Dark Web After Fed Crackdowns*.

[19] *See Cengage Learning, Inc. et al. v. Does 1-50 d/b/a Library Genesis*, ECF No. 36, Case No. 23-cv-8136 (S.D.N.Y. Sept. 24, 2024).

[20] *In re OpenAI, Inc., Copyright Infringement Litig.*, Case No. 1:2025-md-03143 (S.D.N.Y. 2025), ECF No. 846 at 9; Ashley Belanger, *OpenAI desperate to avoid explaining why it deleted pirated book datasets*, ARSTECHNICA (Dec. 1, 2025), https://perma.cc/9M7K-8DA9.

[21] Leo Gao et al., *The Pile: An 800GB Dataset of Diverse Text for Language Modeling*, arXiv, 1 (2020), https://perma.cc/NHV6-R8YE.

approximately 196,640 books.[22] Books3 comprises about 12 percent (just over 100 gigabytes) of *The Pile*.[23]

100. EleutherAI and the compiler of Books3, Shawn Presser, have confirmed the genesis of Books3 is the shadow library Bibliotik. Presser has publicly stated that Books3 represents "all of bibliotik,"[24] and an EleutherAI paper likewise confirms that Books3 "is a dataset of books derived from a copy of the contents of the Bibliotik private tracker."[25]

101. Bibliotik is a private, invitation-only torrent tracker that has long functioned as a centralized source of pirated e-books. The repository, which hosts and distributes hundreds of thousands of copyrighted books, is only accessible with the use of torrenting protocols, which are explained below.[26]

102. Bibliotik has been widely recognized in piracy communities, academic literature, and AI-research documentation as a shadow library devoted to copyrighted books. Unsurprisingly, then, *The Pile*'s datasheet acknowledges that "Books3 is almost entirely comprised of copyrighted works."[27] Bibliotik's illicit nature is no secret—it has been openly discussed for years prior to major tech companies' use of datasets derived directly from it.

103. In an interview with *The Atlantic*, Presser confirmed that the illuminating purpose behind Books3 was to ensure broad-based access to the tools necessary to create LLMs:

> [Presser] created Books3 in the hope that it would allow any developer to create generative-AI tools. "It would be better if it wasn't necessary to have something like

[22] Stella Biderman et al., *Datasheet for the Pile*, arXiv, 8 (2020), https://perma.cc/7KL2-LTLF.

[23] *Id.*

[24] Shawn Presser, X (Oct. 25, 2020 1:32 AM), https://perma.cc/7WRD-NHRX.

[25] Biderman et al., *Datasheet for the Pile*, arXiv, 8.

[26] Ruheni Mathenge, *The 12 Best Private Torrent Sites Still Working in 2026*, PRIVACYSAVVY (last accessed March 9, 2026), https://perma.cc/4V8M-3ALY.

[27] Biderman et al., *Datasheet for the Pile*, arXiv, 8.

Books3," he said. "But the alternative is that, without Books3, only OpenAI can do what they're doing."[28]

104. EleutherAI's compilation and distribution of *The Pile* and Books3 provided "off-the-shelf" access to a corpus of infringing works that any AI developer could download and immediately incorporate into a large-scale training pipeline. As a result, Books3's presence within *The Pile* facilitated wide downstream distribution and adoption of a corpus derived from a pirate book library.[29]

105. Bibliotik is just one of many Internet-based shadow libraries and, as alleged below, Meta relied on more than just *The Pile* in connection with its development of Llama.

106. As noted above, OpenAI's developers trained its LLMs on a books corpus derived from LibGen, one of the largest and longest-running shadow libraries in the world. LibGen hosts millions of pirated books, academic texts, and scholarly articles.[30] It operates as a centralized repository offering direct downloads of full-fidelity e-book files. It also distributes its entire database through bulk archives and torrent files, enabling third parties to download and locally host complete copies of its collection.[31] LibGen has been repeatedly enjoined by federal courts for copyright infringement and has been designated a "notorious market" by the United States Trade Representative.[32] Despite enforcement actions, LibGen has remained accessible through shifting

---

[28] Alex Reisner, *Revealed: The Authors Whose Pirated Books Are Powering Generative AI,* THE ATLANTIC (Aug. 19, 2023), https://perma.cc/P25L-S9FC.

[29] *See* Stella Biderman, *The Pile: An 800GB Dataset of Diverse Text for Language Modeling*, ELEUTHERAI (Dec. 31, 2020), https://perma.cc/JGT9-LLTK.

[30] Office of the U.S. Trade Representative, REVIEW OF NOTORIOUS MARKETS FOR COUNTERFEITING AND PIRACY, 27 (2024), https://perma.cc/N4WT-AHJ9 ("Libgen … hosts a large number of digital copies of books, manuals, journals, and other works, many of which are unauthorized copies of copyright protected content.").

[31] *See* Letter from Mary E. Rasenberger, CEO, Authors Guild, & Umair Kazi, Dir. of Pol'y & Advocacy, Authors Guild, to Daniel Lee, Assistant U.S. Trade Representative for Innovation & Intellectual Prop., Office of the U.S. Trade Representative at n.5.

[32] *See* Office of the U.S. Trade Representative, 2019 REVIEW OF NOTORIOUS MARKETS FOR COUNTERFEITING AND PIRACY, 27, https://perma.cc/22VN-9VD7.

domains, mirrors, and downloadable archives. Its persistence is a function of deliberate decentralization designed to evade shutdown.[33]

107. Z-Library (also known as "B-ok") is another well-known shadow library that emerged as an expanded and user-friendly derivative of LibGen. It incorporated large portions of LibGen's catalog while adding additional titles, metadata, and interface features.[34] Z-Library offered premium features—including faster downloads and higher volume limits—in exchange for payment, operating in effect as a commercial piracy service.[35] In 2022, Z-Library's domains were seized by law-enforcement authorities, and its operators were arrested and later indicted for criminal copyright infringement.[36] These actions confirmed what had long been publicly known: Z-Library was an illegal piracy operation. The seizure of Z-Library did not eliminate access to its content. Instead, third parties responded by creating full mirrors of its collection to ensure continued distribution.[37]

108. Another major player in the shadow library ecosystem is Anna's Archive. Anna's Archive is the most comprehensive and active shadow library currently in operation.[38]

109. Anna's Archive began in 2022 as "Pirate Library Mirror," initially hosting a mirrored copy of Z-Library. It later rebranded as "Anna's Archive" and expanded to aggregate and host the

---

[33] *See* Andrew Albanese, *Textbook Publishers Sue Notorious 'Shadow Library' Libgen,* PUBLISHERS WEEKLY (Sep. 14, 2023), https://perma.cc/3NPY-UJCX.

[34] Jordana Rosenfeld, *Z-Library*, ENCYCLOPEDIA BRITANNICA (last accessed March 9, 2026), https://perma.cc/4H26-GDCC.

[35] *See* Masood Farivar, *Two Russian Nationals Charged With Operating E-Book Piracy Site*, VOA (Nov. 16, 2022), https://perma.cc/6MPD-QNKB.

[36] Press Release, U.S. Dep't of Justice, U.S. Att'y's Off., E. Dist. of N.Y., *Two Russian Nationals Charged with Running Massive E-Book Piracy Website* (Nov. 16, 2022), https://perma.cc/CR4L-JLA3.

[37] *See* Woodcock, *'Shadow Libraries' Are Moving Their Pirated Books to The Dark Web*.

[38] *See* Soumyajyoti Mukherjee, *Who is Anna's Archive? All we know about pirate activist group behind 300 TB Spotify music library heist*, SOAPCENTRAL.COM (Dec. 24, 2025), https://perma.cc/DC5S-89Z8.

complete collections of LibGen, Z-Library, and other pirated sources.[39] Anna's Archive functions as a meta-library: it indexes, mirrors, and redistributes multiple shadow libraries simultaneously, offering users unified access to millions of pirated books.[40]

110. Anna's Archive offers paid tiers that provide high-speed or priority access to its pirated collections.[41] Through its downloadable archives and torrent-based distribution, Anna's Archive enables users to acquire and store local copies of millions of copyrighted books in bulk.[42]

111. Although individual domain names may change, Anna's Archive and its underlying datasets remain accessible through mirrors, torrents, and distributed storage systems.

112. According to Anna's Archive, "virtually all major companies building LLMs contacted us to train on our data. . . We have given high-speed access to about 30 companies."[43] Anna's Archive blog stated as recently as February 18, 2026 that if an *LLM was reading its blog* "you have likely been trained in part on our data."[44]

113. Many of these shadow libraries and datasets can be downloaded using "torrent," a file-sharing method used in peer-to-peer networks. Torrenting works by breaking a file into many small pieces and distributing those pieces across a network of participating computers. A user who torrents a shadow-library repository does not receive a single copy from a single source; rather, the user downloads portions of the library from numerous other computers that already possess the

---

[39] *See* Ernesto Van de Sar, *"Anna's Archive" Opens the Door to Z-Library and Other Pirate Libraries*, TORRENTFREAK (Nov. 19, 2022), https://perma.cc/U88R-WTR4.

[40] *Id.*

[41] *See If you're an LLM, please read this,* ANNA'S ARCHIVE (February 18, 2026), https://perma.cc/989X-GS3Q.

[42] *See* M. Luisa Simpson, *2024 Special 301 Out-of-Cycle Review of Notorious Markets: Request for Comments* ASSOCIATION OF AMERICAN PUBLISHERS (OCTOBER 2, 2024), https://perma.cc/P2E9-MYBY.

[43] *See Copyright reform is necessary for national security,* ANNA'S ARCHIVE (Jan. 31, 2025), https://perma.cc/3RVZ-6G5B.

[44] *See If you're an LLM, please read this,* ANNA'S ARCHIVE (Feb. 18, 2026), https://perma.cc/MJ7Z-3ZCL.

copyrighted books. Torrent software then reassembles those pieces into a complete library on the user's machine. Torrenting protocols, including BitTorrent, are often configured by default to *reupload* pieces of the copyrighted files to others on the network both during download ("leeching") and after download is complete ("seeding"). This means that each participant in the torrent both copies *and redistributes* the copyrighted works. By obtaining copyrighted materials through this leech-and-seed process, a user may make multiple unauthorized reproductions of, and engage in numerous distributions of, the copyrighted materials.

D. Meta's Deliberate Use of Pirated Datasets to Train the Llama Family of Models

114. As noted above, Meta develops and distributes the Llama series of LLMs, including Llama 1, Llama 2, Llama 3, and Llama 4. Meta has "invested billions of dollars to develop its generative AI offerings,"[45] and it in turn expects Llama to generate somewhere in the realm of $460 billion to $1.4 trillion in revenue.[46] None of the money that Meta invested into the development of its Llama models was used to pay for Plaintiffs' copyrighted works.[47]

115. Meta released the first of its Llama models, Llama 1, in February 2023.

116. In connection with the development of Llama, Meta viewed book-corpora as among its most valuable sources of training data. Llama's design goal was to emit particularly creative and expressive language, leveraging Meta's consumer platforms to "connect" with users through text.[48] To accomplish that, Meta needed to train on large quantities of high-quality books and other published materials.

[45] *Cambronne Inc. v. Meta Platforms, Inc.*, No. 3:26-cv-03725-VC (N.D. Cal.), ECF No. 109 ¶ 40 (Meta's Answer).

[46] *See Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1040 (N.D. Cal. 2025).

[47] *Cambronne*, ECF No. 109 at ¶¶ 2, 9, 42, 89 (admitting in *Cambronne* that "it did not obtain a license or pay for the use of [copyrighted] works").

[48] Jon Russell, *Mark Zuckerberg Announces New Team at Meta Working on A.I. Products for Instagram, WhatsApp*, CNBC (Feb. 27, 2023), https://perma.cc/JKE9-N4T7 ("Zuckerberg said that the team would build 'creative and expressive' tools to be used inside Meta's products.").

117. Meta employees repeatedly acknowledged the importance of books as training data. Specifically, they acknowledged it was "really important for [Meta] to get books data ASAP," and the "best resources [Meta] [could] think of are definitely books."[49]

118. Prior to the release of Llama 1, Meta downloaded copyrighted works from LibGen. As recognized in *Kadrey v. Meta Platforms, Inc.*, Meta first downloaded LibGen in 2022 "to investigate whether there was value in training Llama on the works it contained. If the answer was yes, the plan was to then set up licensing agreements for those or similar works."[50]

119. At the time Meta downloaded LibGen, it knew or should have known that LibGen was a repository of unauthorized copies of copyrighted works.

120. Ultimately, Meta decided not to use its pirated LibGen content in connection with the training of Llama 1. It did, however, rely on another dataset of pirated works—Books3—in training Llama 1.

121. In connection with Llama 1's release, Meta disclosed the composition of the model's "pre-training data."[51] Among the datasets used to train Llama 1 were "two book corpora," including "the Books3 section of ThePile."[52] In their research paper, Meta's developers described Books3 as "a publicly available dataset for training large language models," which they distinguished from the other book corpus it used, which they described as being in the "public domain."[53] This distinction was deliberate and notable, as Meta *knew* that Books3 was *not* a corpus of books in the "public domain."

122. Among the other datasets used by Meta to train Llama 1 were "five CommonCrawl dumps, ranging from 2017 to 2020," comprising 67 percent of the overall data mix, and "the publicly

---

[49] *Kadrey*, 788 F. Supp. 3d at 1040.

[50] *Id.*

[51] Hugo Touvron et al., *LLaMA: Open and Efficient Foundation Language Models*, arXiv, 2 (2023), https://perma.cc/LB97-ZK2C.

[52] *Id.*

[53] *Id.*

available C4 dataset," comprising 15 percent of the data mix.[54] "CommonCrawl is a publicly-available web archive that provides 'web extracted text' by removing markup and other non-text content from the scraped HTML files. This process produces around 20TB of scraped data per month."[55] The C4 dataset itself is derived from CommonCrawl.[56] In short, a massive amount (and the majority) of data used by Meta to train Llama was "scraped" from the Internet.

123. As alluded to above, Meta at one time had a "plan . . . to then set up licensing agreements" to obtain the copyrighted materials it wanted for use in training its AI models.[57] But this idea was short-lived. "[I]n spring 2023, after failing to acquire licenses and following escalation to CEO Mark Zuckerberg, Meta decided to just use the works acquired from LibGen as training data."[58]

124. Done with the idea of paying for licenses to use copyrighted materials, Meta pivoted to shadow libraries and web crawlers to obtain the training data it wanted. Meta turned to shadow libraries knowing that these were just that—repositories of unauthorized copies of copyrighted works.

125. "In early 2024, Meta also downloaded Anna's Archive, a compilation of shadow libraries including LibGen, Z-Library, and others."[59]

[54] Touvron et al., *LLaMA: Open and Efficient Foundation Language Models*, arXiv, 2.

[55] Colin Raffel et al., *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*, Journal of Machine Learning Research 21, 7 (2020), https://perma.cc/A4VM-V8XR.

[56] *Id.*

[57] *Kadrey*, 788 F. Supp. 3d at 1041; *see also Cambronne*, ECF No. 109 at ¶ 88 (Meta admitting that "it contacted certain publishers" and "internally discussed licensing certain types of data to train its Llama models").

[58] *Kadrey, 788 F. Supp. 3d* at 1041.

[59] *Id*.

126. Meta likewise downloaded copyrighted materials from Z-Library, Sci-Hub, and other shadow libraries.[60] Indeed, by April 2024, Meta had downloaded over 25 TB of data from Z-Library and 10 TB of data from LibGen.[61]

127. To obtain the massive quantities of data from these shadow libraries, Meta used torrenting protocols.[62] In using these protocols to obtain data from these shadow libraries, Meta reuploaded (and distributed) to peer-to-peer networks at least some of the copyrighted materials it had downloaded, whether through "leeching" or by "seeding."[63]

128. By uploading and distributing copyrighted materials on peer-to-peer sharing networks, Meta made those materials available for download to other third parties whose downloading and redistribution of copyrighted materials would constitute further infringement of the copyrighted materials.

129. There was no substantial or commercially significant non-infringing use of the copyrighted materials that Meta uploaded and distributed. Similarly, there was no substantial or commercially significant non-infringing use of Meta's uploading and redistribution of these materials.

130. In torrenting from shadow libraries, Meta knew or should have known that it was participating in a *peer-to-peer network that traded in unauthorized copies of copyrighted material*. Stated differently, Meta knew or should have known that it was facilitating further copyright infringement by making available to others on the peer-to-peer network unauthorized copies of copyrighted materials.

---

[60] *Cambronne*, ECF No. 109 (Meta Answer), ¶ 83.

[61] *Kadrey*, ECF No. 472 (Pls.' Partial Mot. for Summ. J.), at 11.

[62] *Cambronne*, ECF No. 109 at ¶ 83 (admitting that "it used BitTorrent to download certain portions of these publicly available datasets").

[63] *See Kadrey*, 788 F. Supp. 3d at 1041 ("To download these large datasets more quickly and without unnecessarily slowing down its networks, Meta torrented them. . . . There is no dispute that Meta torrented LibGen and Anna's Archive, but the parties dispute whether and to what extent Meta uploaded (via leeching or seeding) the data it torrented. A Meta engineer involved in the torrenting wrote a script to prevent seeding, but apparently not leeching.").

131. In connection with the development of Llama, certain Meta employees questioned, implicitly or explicitly, the legal or ethical implications of relying on copyrighted content, content from shadow libraries, and torrenting protocols. On information and belief, certain employees implicitly or explicitly raised concerns with training Llama on "pirated" content. On information and belief, certain employees implicitly or explicitly raised concerns with obtaining copyrighted materials from shadow libraries. And on information and belief, certain employees implicitly or explicitly raised concerns with torrenting from shadow libraries.

132. Notwithstanding what it knew about these peer-to-peer networks and concerns raised by employees, Meta torrented large quantities of data from these shadow libraries, including but not limited to terabytes of data from each of Anna's Archive and LibGen.

133. Meta continued to download from these shadow libraries even after it had been sued for copyright infringement based on its use of copyrighted works to train Llama.

134. Meta also relied on web crawlers to obtain massive amounts of Internet data for use in training Llama. On information and belief, the web crawler data that Meta obtained in connection with its Llama training contained copyrighted materials.

135. Upon downloading this material, Meta manipulated the downloaded data. Among other things, Meta processed the data, and in doing so, removed and altered certain text and information, including titles, author information, information about the copyright owner, and copyright notices. When it removed this information, Meta did so without the authority of the copyright owners.

136. As Meta has admitted elsewhere, Meta "created a script to remove repetitive text from the training data for its Llama models."[64] As admitted elsewhere, Meta's "program was designed to remove lines that contain the word 'copyright.'"[65]

137. As alleged above, the Plaintiffs' Works appear in various shadow libraries and/or pirated datasets, including Books3, LibGen, Anna's Archive, and Z-Library. As such, there is a

[64] *Kadrey*, ECF No. 485, ¶ 88 (Meta's Answer to Third Am. Compl.).

[65] *Kadrey*, ECF No. 485, ¶ 89.

reasonable inference that Meta downloaded, manipulated, reproduced, and distributed the Plaintiffs' Works.

138. As admitted elsewhere, Meta used at least Books3 and data it obtained from LibGen and Z-Library to train the Llama family of models.[66] On information and belief, it also used data it obtained from Anna's Archive and other shadow libraries to train the Llama family of models. After Llama 1, Meta stopped disclosing detailed information about the datasets that its models were trained on. Meta did this, in part, to conceal the fact that it had relied on and was relying on pirated materials to train Llama.

139. Meta has since confirmed, in interrogatory responses served in other litigation, the specific pirated datasets on which each Llama model was trained. According to Meta, the Books3 dataset (which Meta refers to internally as "b3g") was used to train Llama 1, Llama 2, Llama 3, and Llama 4; the LibGen "Scitech" and LibGen "Fiction" datasets were used to train Llama 3 and Llama 4; and the Z-Library dataset, which Meta obtained by torrenting Anna's Archive, was used to train Llama 4.[67] Meta has further admitted that, beginning in or about Spring 2023, a Meta engineer downloaded the Fiction and Scitech branches of LibGen using a direct-download script, and that Meta's downloads of both branches "are believed to have been substantially complete."[68] In other words, if a book was in LibGen, Meta downloaded it.

140. On information and belief, Meta, like Anthropic, downloaded, maintained, and retained pirated copies of works that were not specifically tied to training its LLM models.[69]

[66] *Cambronne*, ECF No. 109, ¶ 2 (Meta's Answer, admitting that "it obtained publicly available data from sources such as LibGen and Z-Library, portions of which were utilized to research, develop, and train its Llama models").

[67] Def. Meta Platforms, Inc.'s Second Supp. Resps. & Objs. to Pl.'s First Set of Interrogs. at 17–18, *Chicken Soup for the Soul, LLC v. Meta Platforms, Inc.*, No. 3:26-cv-02333-VC (N.D. Cal. Sept. 4, 2026), filed as ECF No. 140-6 (Sept. 12, 2026) (hereinafter "Meta Interrog. Resps.").

[68] *Id.* at 12–14.

[69] *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1014 (N.D. Cal. 2025) (Anthropic's amassing of " a central library").

141. Meta has incorporated Llama into several of its consumer-facing products, including Facebook, Instagram, and WhatsApp. And although Meta has offered free downloads of various versions of Llama, it considers or has considered Llama and the technology underpinning it to be an important part of its commercial success. In other words, Meta hopes and expects that its development of Llama will contribute to its current and future financial success.

E. Project Vellum: Meta's Mass Scanning of Physical Books to Train Its Muse Models

142. Meta's appetite for books did not abate after it was sued for training Llama on pirated copies. Nor did Meta turn to the licensing market. Instead, beginning in or around mid-2025, Meta began acquiring physical copies of books for use in training its next generation of LLMs.[70] Meta internally referred to this book-acquisition and scanning operation as "Project Vellum," and it is referred to as such herein.[71]

143. Project Vellum was an industrial-scale operation. Meta acquired physical books in bulk, catalogued them by ISBN, title, author, publisher, and year of publication, and routed them through a scanning pipeline. By Meta's own account, the acquired physical copies were "first scanned into digital form and then run through OCR"—optical character recognition—"to extract their textual data from the scanned images," and the extracted text was "then de-duplicated, processed and cleaned" for use as model training data.[72]

144. Meta used the digital copies it created through Project Vellum to train its "Muse" family of LLMs. In April 2026, Meta introduced the initial version of its Muse Spark model, which Meta made available to the public through certain Meta products and services and, in later versions,

[70] Meta Interrog. Resps. at 23.

[71] *Entrepreneur Media, LLC v. Meta Platforms, Inc.*, Case No. 25-cv-09579-VC (N.D. Cal.), ECF No. 113 at 3 (Meta's Administrative Motion to Quash Apex Deposition); Annelise Levy, *Meta Fails to Block Zuckerberg Deposition in AI, Copyright Suit*, BLOOMBERG LAW (Aug. 11, 2026), https://perma.cc/FMW7-CMGC.

[72] Meta Interrog. Resps. at 23; *see also id.*, Ex. D (identifying, by ISBN, title, author, publisher, and year of publication, works used to train Meta's Muse Spark models).

to developers through an application programming interface ("API"). In August 2026, Meta released its Muse Glimmer model to the open-source community for use by researchers, developers, and other third parties.[73] Meta has admitted that books it acquired and scanned as part of Project Vellum were used to train the Muse Spark family of models.[74]

145. At no point did Meta seek or obtain authorization from Plaintiffs—or from authors and publishers generally—to reproduce their works through Project Vellum. Meta did not license the right to scan, digitize, copy, or use these works as training data. Instead, Meta simply bought a copy of each book and treated that purchase as though it had acquired the copyright owner's exclusive rights.

146. It had not. The purchase of a physical copy of a book conveys ownership of that particular copy and nothing more. Ownership of a material object in which a work is embodied does not convey any rights in the copyrighted work itself. 17 U.S.C. § 202. And while the first-sale doctrine permits the owner of a lawfully made copy to sell or otherwise dispose of that copy, 17 U.S.C. § 109(a), it confers no right to reproduce the work. The exclusive right to reproduce a copyrighted work in copies belongs to the copyright owner alone. 17 U.S.C. § 106(1).

147. Scanning a book is reproduction. Each time Meta scanned a physical copy of a book, it created a new digital copy of the entire work. Each time Meta ran the scanned images through OCR, it created another copy—this time as machine-readable text. Each time Meta processed, "cleaned," de-duplicated, and tokenized that text, it created still more copies. And each time Meta used that text in a training run—including in the "number of internal models" that Meta trains "for experimental and evaluation purposes" before each public release, and in the "purpose-built variants" of its LLMs that Meta uses for discrete internal purposes—it created additional copies of the work.[75] Meta has admitted that "[a] number of transitory and persistent copies of each of these

[73] Meta Interrog. Resps. at 23, 25.

[74] *Id.* at 18 & Ex. D.

[75] *Id.* at 18.

works would have been created" in its pipeline and that it "has no feasible way to reliably ascertain the number of additional persistent or transitory copies" it made.[76]

148. Meta launched Project Vellum with full knowledge that the books it was scanning were protected by copyright. Unlike the pirated files Meta downloaded from shadow libraries—from which Meta stripped copyright notices and other copyright management information—Meta has admitted that "copyright-related information from these purchased books was not filtered out before being used for model training."[77] In other words, Meta scanned each book's copyright page, ingested the copyright notice, and trained on the work anyway.

149. Meta also launched Project Vellum with full knowledge that a market exists for licensing books for use in AI training. By mid-2025, Meta had itself explored licensing deals with publishers, and numerous other AI developers had entered into such deals. Meta chose to purchase single copies of books rather than pay for training licenses because it was cheaper. The retail price of a book compensates the author for one reader's copy; it does not compensate the author for the right to reproduce the work wholesale and to build a commercial product on its expressive content. Project Vellum was designed to bypass the licensing market, and it did.

150. On information and belief, Meta acquired, scanned, digitized, reproduced, and used physical copies of Plaintiffs' Works as part of Project Vellum, and used the resulting digital copies to train one or more of its Muse models.

151. Meta's reproduction and use of Plaintiffs' Works through Project Vellum constitutes an independent basis of liability, separate and apart from Meta's acquisition, reproduction, and distribution of pirated copies of Plaintiffs' Works from shadow libraries. Even if Meta had never downloaded a single pirated copy of any of Plaintiffs' Works, its unauthorized scanning, digitization, reproduction, and exploitation of physical copies of those works would independently violate Plaintiffs' exclusive rights under 17 U.S.C. § 106.

[76] *Id.* at 24.

[77] *Id.* at 23; *see also id.* at 19–20 (Meta "did not employ" its copyright-text filtering step "for scanned and digitized books it used as training data for its Muse models").

F. Meta's Conduct Harms the Market for Plaintiffs' Copyrighted Materials

152. Meta's conduct has a detrimental effect on the potential market for and value of Plaintiffs' Works, including by, among other things, developing products that create and are capable of creating content which serves as a direct substitute for Plaintiffs' Works, developing products that create content and are capable of creating content which serves as indirect substitutes for Plaintiffs' Works, and undermining Plaintiffs' ability to participate in and profit from the market for licensing their works for the purpose of training LLMs.

153. *First*, Meta's decision to download and use unauthorized copies of Plaintiffs' Works from shadow libraries—and to scan and use physical copies of Plaintiffs' Works without a license—deprived Plaintiffs of revenue in the form of licensing fees that they would have otherwise earned. As alleged herein, there is an existing market for licensing copyrighted materials such as Plaintiffs', including for use in the development of LLMs. Meta bypassed that market, and in doing so, deprived Plaintiffs of licensing revenue they would have earned. Meta now allows users in certain markets to opt out of having their data used to train its AI models, but it deprived Plaintiffs and many others of that choice. Meta's misconduct undermined Plaintiffs and many others, eliminating the bargaining power they should have had, and otherwise would have had, with respect to licensing terms for the use Meta made of their works.

154. *Second*, Llama and Muse, trained on Plaintiffs' copyrighted materials (and the protected works of others), are capable of generating outputs that compete directly with, and risk serving as replacements for, Plaintiffs' Works. As the U.S. Copyright Office has warned, "the speed and scale at which AI systems generate content pose a serious risk of diluting markets for works of the same kind as in their training data."[78]

155. *Third*, even if Llama and Muse models are restricted from outputting extended portions of verbatim text from copyrighted works, they are nevertheless capable of producing nearly

[78] U.S. Copyright Office, *Copyright and Artificial Intelligence, Part 3: Generative AI Training Pre-Publication*, at 65.

indistinguishable "versions" of copyrighted works such that a consumer would use the AI-generated version of the material rather than pay for a copy of the actual copyrighted work.

V. CLAIMS FOR RELIEF

COUNT I

Direct Copyright Infringement (17 U.S.C. § 501)—Pirated Copies Obtained from Shadow Libraries

156. Plaintiffs incorporate the allegations above.

157. Plaintiffs are the legal or beneficial owners of the copyrighted works listed in Exhibit A (referred to herein as the Plaintiffs' Works).

158. Defendant, without authorization from Plaintiffs, copied, downloaded, reproduced, ingested, parsed, embedded, and used pirated copies of Plaintiffs' Works in the development, training, fine-tuning, and deployment of its Llama family of models. These acts violated Plaintiffs' exclusive rights under 17 U.S.C. § 106.

159. Defendant's infringement occurred repeatedly throughout the lifecycle of its AI-model development. As alleged above, Defendant:

- acquired Plaintiffs' Works from shadow-library repositories and datasets containing pirated works from shadow libraries;
- distributed Plaintiffs' Works through the use of torrenting software, programs, or protocols;
- reproduced additional copies during ingestion, preprocessing, storage, deduplication, formatting, and/or tokenization; and
- while training its models, made even more copies of the text—because every training pass (each epoch and each step of gradient descent) automatically requires creating and working with fresh versions of that text.

160. Defendant's reproductions and distributions of Plaintiffs' copyrighted works were made without permission, license, or consent and violated Plaintiffs' exclusive rights under the Copyright Act.

161. Defendant's infringement was willful. As alleged above, Defendant knowingly trained its models on and/or optimized its product with datasets saturated with pirated books, including Plaintiffs' Works; relied on shadow-library corpora it knew to be illegal; ignored internal and external warnings; attempted to conceal the composition of its training datasets; and continued copying after public reports, lawsuits, law-enforcement seizures, cease-and-desist notices, and industry-wide alerts made the illegality unmistakable. This conduct by Defendant meets the standard for criminal copyright infringement.[79]

162. Upon information and belief, Defendant has made and will continue to make substantial profits and gains to which it is not in law or in equity entitled.

163. Plaintiffs have been injured by Defendant's willful acts of copyright infringement. Plaintiffs are entitled to statutory damages, actual damages, restitution of profits, and/or other remedies in law or equity.

164. Plaintiffs are entitled to recover attorneys' fees and costs under 17 U.S.C. § 505.

COUNT II

Direct Copyright Infringement (17 U.S.C. § 501)—Project Vellum

165. Plaintiffs incorporate the allegations above.

166. Plaintiffs are the legal or beneficial owners of the copyrighted works listed in Exhibit A (referred to herein as the Plaintiffs' Works).

167. Beginning in or around mid-2025, Defendant, without authorization from Plaintiffs, acquired physical copies of Plaintiffs' Works, scanned those copies into digital form, ran the scanned images through optical character recognition software to extract the text of Plaintiffs' Works, and processed, cleaned, de-duplicated, tokenized, stored, and retained the resulting digital copies. Defendant then used those digital copies to train, develop, fine-tune, evaluate, and deploy its Muse family of models. Each of these acts reproduced Plaintiffs' Works in copies, in violation of Plaintiffs' exclusive rights under 17 U.S.C. § 106(1).

[79] 17 U.S.C. § 506(a)(1)(A).

168. Defendant's purchase of physical copies of Plaintiffs' Works did not authorize any of these reproductions. Ownership of a physical copy conveys no right to reproduce the copyrighted work embodied in it. 17 U.S.C. §§ 109(a), 202. Defendant never sought or obtained a license from Plaintiffs to scan, digitize, copy, or otherwise reproduce Plaintiffs' Works, or to use Plaintiffs' Works as training data.

169. Defendant's reproductions of Plaintiffs' Works through Project Vellum were not a one-for-one change of format for Defendant's own reading or reference. They were wholesale reproductions of entire works, made and retained in multiple copies, for the purpose of building a commercial training corpus and commercial AI products, in a manner that supplants the existing market for licensing Plaintiffs' Works for AI training and that threatens to dilute the market for Plaintiffs' Works.

170. Defendant's infringement through Project Vellum is independent of, and in addition to, its infringement through the acquisition, reproduction, and distribution of pirated copies of Plaintiffs' Works alleged in Count I. Each is a separate and independent basis for liability.

171. Defendant's infringement was willful. Defendant undertook Project Vellum after it had been sued for training Llama on pirated books; with actual knowledge that the books it scanned bore copyright notices; with knowledge of an existing market for licensing books for AI training; and with the deliberate purpose of avoiding the cost of that market.

172. Upon information and belief, Defendant has made and will continue to make substantial profits and gains to which it is not in law or in equity entitled.

173. Plaintiffs have been injured by Defendant's willful acts of copyright infringement. Plaintiffs are entitled to statutory damages, actual damages, restitution of profits, and/or other remedies in law or equity.

174. Plaintiffs are entitled to recover attorneys' fees and costs under 17 U.S.C. § 505.

COUNT III

Contributory Copyright Infringement (17 U.S.C. § 501)

175. Plaintiffs incorporate the allegations above.

176. Defendant used torrenting software, programs, or protocols to download datasets containing pirated copies of works, including the Plaintiffs' Works.

177. In connection with its torrenting of datasets that contained copyrighted works, Defendant uploaded and distributed, either through "seeding" and/or "leeching," copyrighted materials, including the Plaintiffs' Works, thereby making those works available to third parties for downloading on peer-to-peer networks.

178. Defendant knowingly participated in peer-to-peer sharing networks that it knew trafficked in pirated copies of copyrighted materials. In other words, Defendant knew that others on these networks were infringing copyrighted materials through reproduction and/or distribution. There was no substantial or commercially significant non-infringing use of the copyrighted materials that Meta uploaded and distributed. Nor was there substantial or commercially significant non-infringing use of Defendant's uploading and distribution of Plaintiffs' copyrighted works. By participating in these networks, and by further uploading and distributing Plaintiffs' copyrighted works, Defendant materially contributed to and induced further infringement of Plaintiffs' Works.

179. By knowingly inducing and materially contributing to others' infringement of Plaintiffs' Works, Defendant is liable for contributory copyright infringement.

180. As a direct and proximate result of Defendant's conduct, Plaintiffs were injured and are entitled to statutory damages, actual damages, restitution of profits, and/or other remedies in law or equity.

181. Plaintiffs are entitled to recover attorneys' fees and costs under 17 U.S.C. § 505.

COUNT IV

Removal of Copyright Management Information (17 U.S.C. § 1202(b)(1))

182. Plaintiffs incorporate the allegations above.

183. Plaintiffs' Works contain information that constitutes "copyright management information" as that term is defined in 17 U.S.C. § 1202(c). This includes but is not limited to author information, information about the copyright owner, and copyright notices.

184. Upon downloading copyrighted materials, including Plaintiffs' Works, Defendant processed the data, and in doing so, removed and altered certain text and information, including

copyright management information found in and on Plaintiffs' Works. When it removed this information, Defendant did so without the authority of the copyright owners.

185. Defendant's removal of the copyright management information was intentional—it did so to, among other things, create high-quality LLM training data and, through the creation and use of high-quality training data, ultimately create high-quality LLM models.

186. Defendant removed copyright management information from Plaintiffs' Works knowing or having reasonable grounds to believe that it was enabling, facilitating, and concealing acts of copyright infringement. As to concealment, Defendant knew or had reasonable grounds to believe that by stripping copyrighted works of copyright management information it would be harder for others to discover the true sources—*e.g.*, copyrighted works—of Defendant's training data.

187. Plaintiffs were harmed by Defendant's removal of copyright management information from their works and are entitled to statutory damages, actual damages, restitution of profits, and other remedies provided by law. Plaintiffs are entitled to recover attorneys' fees and costs under 17 U.S.C. § 1203(b)(5).

PRAYER FOR RELIEF

WHEREFORE, Plaintiffs request that the Court enter judgment on their behalf by ordering:

a. Judgment in favor of Plaintiffs against the Defendant;

b. A declaration that the Defendant has infringed Plaintiffs' exclusive copyrights under the Copyright Act;

c. A declaration that such infringement is willful;

d. A declaration that Defendant violated 17 U.S.C. § 1202(b) through its removal of copyright management information;

e. A permanent injunction enjoining the Defendant and all those acting in concert with it from engaging in the infringing conduct alleged herein;

f. That the Defendant be directed to account to Plaintiffs for all gains, profits, and advantages derived from their unlawful acts;

g. An award of statutory damages under the Copyright Act;

h. An award of statutory or actual damages under 17 U.S.C. § 1203(c);

i. An award of restitution, disgorgement, costs, expenses, and attorneys' fees as permitted by law (including those allowable under 17 U.S.C. § 505 and/or 17 U.S.C. § 1203(b)(4)–(5)).

j. Pre- and post-judgment interest on the damages awarded to Plaintiffs; and

k. Further relief for Plaintiffs as the Court may deem just and proper.

JURY TRIAL DEMANDED

Under Federal Rule of Civil Procedure 38(b), Plaintiffs demand a trial by jury.

Dated: October 19, 2026

Respectfully submitted,

/s/ Elizabeth Brannen

Elizabeth Brannen (SBN 226234)
Lauren Martin (SBN 294367)
STRIS & MAHER LLP
17785 Center Court Dr N, Ste 600
Cerritos, CA 90703
T: (213) 995-6800
F: (213) 261-0299
ebrannen@stris.com
lmartin@stris.com

Kyle Roche (*pro hac vice* forthcoming)
FREEDMAN NORMAND FRIEDLAND LLP
155 E. 44th Street, Ste 915
New York NY 10017
T: (646) 494-2900
kroche@fnf.law

Counsel for Plaintiffs