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

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

Counsel for Plaintiffs

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

JACQUELINE WINSPEAR, BLAISE AGUIRRE, JOHN BAIR, ROBERT BELL, ALEXANDRA BERLINA, GOLDIE BROWNING, AARON BUBBICO, JEFF CAROLLO, CHRISTINE CARPENITO, JOSEPHINE CAUST, GENE COGORNO, JESSA CRISPIN, HARVEY DEUTSCHENDORF, STEPHEN DIAMOND, ROGER DOOLEY, JOHN EBERT, CAROL EKARIUS, JULEN ETXABE, ROBERT FICOCIELLO, DONALD FIRESMITH, SILVIA FOTI, ERIC GALE, DAWN GILPIN, LAURA GOMPERTZ, REYNA GRANDE, JEANNETTE GRAULAU, RAYMOND GREENWELL, DANIEL GUMBINER, OWEN GUTFREUND, ANDREW HARGREAVES, PAUL HERMAN, GORDON INKELES, MARY KARLIN, DAVE KERPEN, JOSEPH KERTES, PHILIPPE KRUCHTEN, GARY LAMBERTI, MICHELLE LEDER, ELISABETH LLOYD, JEANNE MARECEK, OLEKSANDR MASLOV, ROB MCENTARFFER, MARSHA J. MEANS, RACHEL NUWER, JOSHUA PATRICK, MINXIN PEI, EDWARD POLLACK, DAVID ROLLINS, BOB ROSNER, TONY ROTHMAN, THEODORE SAVAS, TODD SHUSTER, JANET SORENSEN, MARK SPRAGG, JUSTIN

Civil Case No.:

COMPLAINT

DEMAND FOR JURY TRIAL

VELEZ-HAGAN, RICHARD WATERS, ROBERT WETTEMANN, BETTINA WILLIFORD, DAVID WOLMAN, and J. DYLAN YATES,

Plaintiffs,

v.

META PLATFORMS, INC.,

Defendant.

1. Plaintiffs Jacqueline Winspear, Blaise Aguirre, John Bair, Robert Bell, Alexandra Berlina, Goldie Browning, Aaron Bubbico, Jeff Carollo, Christine Carpenito, Josephine Caust, Gene Cogorno, Jessa Crispin, Harvey Deutschendorf, Stephen Diamond, Roger Dooley, John Ebert, Carol Ekarius, Julen Etxabe, Robert Ficociello, Donald Firesmith, Silvia Foti, Eric Gale, Dawn Gilpin, Laura Gompertz, Reyna Grande, Jeannette Graulau, Raymond Greenwell, Daniel Gumbiner, Owen Gutfreund, Andrew Hargreaves, Paul Herman, Gordon Inkeles, Mary Karlin, Dave Kerpen, Joseph Kertes, Philippe Kruchten, Gary Lamberti, Michelle Leder, Elisabeth Lloyd, Jeanne Marecek, Oleksandr Maslov, Rob McEntarffer, Marsha J. Means, Rachel Nuwer, Joshua Patrick, Minxin Pei, Edward Pollack, David Rollins, Bob Rosner, Tony Rothman, Theodore Savas, Todd Shuster, Janet Sorensen, Mark Spragg, Justin Velez-Hagan, Richard Waters, Robert Wettemann, Bettina Williford, David Wolman, and J. Dylan Yates (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 Jacqueline Winspear is an author who resides in Washington. Jacqueline Winspear is the author of *A Dangerous Place*; *A Sunlit Weapon*; *Among the Mad*; *An Incomplete Revenge*; *An Incomplete Revenge: A Maisie Dobbs Novel*; *Birds of a Feather: A Novel*; *Elegy for Eddie*; *In This Grave Hour: A Maisie Dobbs Novel*; *Journey to Munich*; *Leaving Everything Most Loved*; *Maisie Debbs*[3]; *Maisie Dobbs*; *Messenger of Truth: A Maisie Dobbs Novel*; *Pardonable Lies: A Maisie Dobbs Novel*; *The American Agent*; *The Care and Management of Lies*; *The Comfort of Ghosts*; *The Consequences of Fear*; *The Mapping of Love and Death*; *The White Lady*; *This Time Next Year We'll Be Laughing*; *To Die but Once*; and *What Would Maisie Do?* One or more of Winspear'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 Winspear's works. Winspear's works were also published

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

[3] This title reflects an error included in the U.S. Copyright Office registration data; the correct title is *Maisie Dobbs*.

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 Blaise Aguirre is an author who resides in Massachusetts. Blaise Aguirre is the author of *Borderline Personality Disorder in Adolescents*; *Coping with BPD: DBT and CBT Skills to Soothe the Symptoms of Borderline Personality Disorder*; *Depression*; *Fighting Back: What an Olympic Champion's Story Can Teach Us About Recognizing and Preventing Child Sexual Abuse—and Helping Kids Recover*; *Helping Your Troubled Teen*; *Parenting Your Child with Autism*; and *The Unwanted Thoughts and Intense Emotions Workbook: CBT and DBT Skills to Break the Cycle of Intrusive Thoughts and Emotional Overwhelm*. One or more of Aguirre'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 Aguirre's works. Aguirre'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 John Bair is an author who resides in Washington. John Bair is the author of *Hiding Behind the Keyboard, First Edition* and *Seeking the Truth from Mobile Evidence: 1st Edition*. One or more of Bair'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 Bair's works. Bair'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.

20. Plaintiff Robert Bell is an author who resides in North Carolina. Robert Bell is the author of *America's Disaster Culture: The Production of Natural Disasters in Literature and Pop Culture* and *Eco Culture: Disaster, Narrative, Discourse*. One or more of Bell'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 Bell's works. Bell'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 Alexandra Berlina is an author who resides in Germany. Alexandra Berlina is the author of *Viktor Shklovsky: A Reader*. Berlina'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 Berlina's work. Berlina'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.

22. Plaintiff Goldie Browning is an author who resides in Kentucky. Goldie Browning is the author of *Night Journey*. Browning'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 Browning's work. Browning'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.

23. Plaintiff Aaron Bubbico is an author who resides in New Mexico. Aaron Bubbico is the author of *Essentials of Eccentric Training*. Bubbico'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 Bubbico's work. Bubbico'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 Jeff Carollo is an author who resides in California. Jeff Carollo is the author of *How Google Tests Software*. Carollo'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 Carollo's work. Carollo'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 Christine Carpenito is an author who resides in Massachusetts. Christine Carpenito is the author of *Back to the Dream*; *Dreamlands*; *I'm Not Who You Think I Am*; *The Secret of the Sheikh's Betrothed*; *Twisted Dreams*; and *Uneasy Dreams*. One or more of Carpenito's works are contained in pirated online libraries, including Books3 (and thus The Pile), 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 Carpenito's works. Carpenito'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 Josephine Caust is an author who resides in Australia. Josephine Caust is the author of *Arts and Cultural Leadership in Asia*. Caust'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 Caust's work. Caust'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.

27. Plaintiff Gene Cogorno is an author who resides in California. Gene Cogorno is the author of *Geometric Dimensioning and Tolerancing for Mechanical Design*; *Geometric Dimensioning and Tolerancing for Mechanical Design, Second Edition*; and *Geometric Dimensioning and Tolerancing for Mechanical Design, Third Edition*. One or more of Cogorno'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 Cogorno's works. Cogorno'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 Jessa Crispin is an author who resides in Pennsylvania. Jessa Crispin is the author of *My Three Dads: Patriarchy on the Great Plains* and *The Creative Tarot*. One or more of Crispin'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 Crispin's works. Crispin'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 Harvey Deutschendorf is an author who resides in Canada. Harvey Deutschendorf is the author of *Of Work and Men: How Men Can Become More Than Their Careers* and *The Other Kind of Smart: Simple Ways to Boost Your Emotional Intelligence for Greater Personal Effectiveness and Success*. One or more of Deutschendorf'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 Deutschendorf's works. Deutschendorf'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.

30. Plaintiff Stephen Diamond is an author who resides in California. Stephen Diamond is the author of *Anger, Madness, and the Daimonic: The Psychological Genesis of Violence, Evil, and Creativity*. Diamond'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 Diamond's work. Diamond'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 Roger Dooley is an author who resides in Texas. Roger Dooley is the author of *Brainfluence* and *FRICTION - The Untapped Force That Can Be Your Most Powerful Advantage*. One or more of Dooley'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 Dooley's works. Dooley'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.

32. Plaintiff John Ebert is an author who resides in New Mexico. John Ebert is the author of *Dead Celebrities, Living Icons: Tragedy and Fame in the Age of the Multimedia Superstar*; *The Age of Catastrophe: Disaster and Humanity in Modern Times*; *The New Media Invasion: Digital Technologies and the World They Unmake*; and *Twilight of the Clockwork God: Conversations on Science and Spirituality at the End of an Age*. One or more of Ebert'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 Ebert's works. Ebert'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 Carol Ekarius is an author who resides in Colorado. Carol Ekarius is the author of *Hobby Farm: Living Your Rural Dream for Pleasure and Profit*; *How to Build Animal Housing: 60 Plans for Coops, Hutches, Barns, Sheds, Pens, Nest Boxes, Feeders, Stanchions, and Much More*; *Pocketful of Poultry*; *Poultry Breeds*; *Small-Scale Livestock Farming*; *Storey's Guide to Raising Sheep, 5th Edition*; *Storey's Guide to Raising Sheep: Breeding, Care, Facilities*; *Storey's Illustrated Breed Guide to Sheep, Goats, Cattle and Pigs*; *Storey's Illustrated Guide to Poultry Breeds*; *The Field Guide to Fleece*; and *The Fleece & Fiber Sourcebook*. One or more of Ekarius'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 Ekarius's works. Ekarius'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.

34. Plaintiff Julen Etxabe is an author who resides in Oregon. Julen Etxabe is the author of *A Cultural History of Law: Volumes 1-6* and *Ranciere and the Law*. One or more of Etxabe'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 Etxabe's works. Etxabe'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 Robert Ficociello is an author who resides in Pennsylvania. Robert Ficociello is the author of *America's Disaster Culture: The Production of Natural Disasters in Literature and Pop Culture* and *Eco Culture: Disaster, Narrative, Discourse*. One or more of Ficociello'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 Ficociello's works. Ficociello'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.

36. Plaintiff Donald Firesmith is an author who resides in Pennsylvania. Donald Firesmith is the author of *Common System and Software Testing Pitfalls: How to Prevent and Mitigate Them*; *Dictionary of Object Technology: The Definitive Desk Reference*; *Hell Holes: What Lurks Below*; *Object-Oriented Requirements Analysis and Logical Design: A Software Engineering Approach*; and *The Method Framework for Engineering System Architectures*. One or more of Firesmith'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 Firesmith's works. Firesmith'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.

37. Plaintiff Silvia Foti is an author who resides in Illinois. Silvia Foti is the author of *The Nazi's Granddaughter: How I Discovered My Grandfather Was a War Criminal*. Foti'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 Foti's work. Foti'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.

38. Plaintiff Eric Gale is an author who resides in Illinois. Eric Gale is the author of *The Bully Book*; *The Wizard's Dog*; *The Wizard's Dog Fetches the Grail*; and *The Zoo at the Edge of the World*. One or more of Gale's works are contained in pirated online libraries, including Books3 (and thus The Pile), 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 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.

39. Plaintiff Dawn Gilpin is an author who resides in Arizona. Dawn Gilpin is the author of *Art, Activism, and Sexual Violence*; *Crisis Management in a Complex World*; and *Organically Grown: Emergent Power and Influence in Complex Issue Networks*. One or more of Gilpin'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 Gilpin's works. Gilpin'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 Laura Gompertz is an author who resides in Texas. Laura Gompertz is the author of *Bayou Dreams*; *Bayou Loup*; *Common Powers 1: Soul Bonds*; *Common Powers 3: Edward Unconditionally*; *In the Company of Men: Baymore's Heir*; *In the Company of Men: The Mercenary's Tale*; *Rougaroux Social Club 2: Bayou's End*; *Rush in the Dark Common Powers 2*; *Silent Lodge*; and *Soul Bonds*. One or more of Gompertz'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 Gompertz's works. Gompertz'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 Reyna Grande is an author who resides in California. Reyna Grande is the author of *A Ballad of Love and Glory*; *A Dream Called Home*; *Across a Hundred Mountains*; *Dancing with Butterflies*; *Spare Parts (Young Readers' Edition)*; *The Distance Between Us*; and *The Distance Between Us: Young Readers Edition*. One or more of Grande'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 Grande's works. Grande'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 Jeannette Graulau is an author who resides in New York. Jeannette Graulau is the author of *Development, Gold and Mining in the Brazilian Amazon: Women's Labor in Garimpagem* and *The Underground Wealth of Nations: On the Capitalist Origins of Silver Mining, A.D. 1150-1450*. One or more of Graulau'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 Graulau's works. Graulau'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 Raymond Greenwell is an author who resides in Colorado. Raymond Greenwell is the author of *Calculus for the Life Sciences, 2/E*; *Calculus with Applications*; *Calculus with Applications for the Life Sciences*; *Calculus with Applications Ninth Edition*; *Calculus with Applications, 11/E*; *Calculus with Applications, 12th Edition*; *Calculus with Applications, Brief*

Version; *Calculus with Applications, Brief Version, 12th Edition*; *Finite Mathematics*; *Finite Mathematics and Calculus with Applications*; *Finite Mathematics Ninth Edition*; *Finite Mathematics, 11/E*; and *Finite Mathematics, 12th Edition*. One or more of Greenwell'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 Greenwell's works. Greenwell'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 Daniel Gumbiner is an author who resides in California. Daniel Gumbiner is the author of *Fire in the Canyon*; *The Best American Nonrequired Reading 2014 - Editors' Note*; *The Best American Nonrequired Reading 2015 - Editor's Note*; *The Best American Nonrequired Reading 2016 - Editor's Note*; *The Best American Nonrequired Reading 2017 - Foreword*; and *The Boatbuilder*. One or more of Gumbiner'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 Gumbiner's works. Gumbiner'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 Owen Gutfreund is an author who resides in New York. Owen Gutfreund is the author of *Twentieth-Century Sprawl: Highways and the Reshaping of the American Landscape*. Gutfreund'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 Gutfreund's work. Gutfreund'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 Andrew Hargreaves is an author who resides in the United Kingdom. Andrew Hargreaves is the author of *Special Operations in World War II: British and American Irregular Warfare.* Hargreaves'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 Hargreaves's work. Hargreaves'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.

47. Plaintiff Paul Herman is an author who resides in California. Paul Herman is the author of *Tales from the Script*. Herman's work is contained in pirated online libraries, including Books3 (and thus The Pile) 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 Herman's work. Herman'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.

48. Plaintiff Gordon Inkeles is an author who resides in California. Gordon Inkeles is the author of *Ergonomic Living: How to Create a User-Friendly Home and Office*; *Massage & Peaceful Pregnancy: A Daily Book for New Mothers and Fathers*; *Massage for a Peaceful Pregnancy*; *Sensual Massage Made Simple*; *Sensual Massage on a String*; *The Art of Sensual Massage*; *The Art of Sensual Massage. by Gordon Inkeles & Mary Todris*[4]; *The New Massage: Total Body Conditioning for People Who Exercise*; *The New Sensual Massage 2013 Edition*; and *Unwinding: Super Massage for Stress Control*. One or more of Inkeles'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 Inkeles's works. Inkeles's

[4] This title reflects an error included in the U.S. Copyright Office registration data; the correct co-author name is Murray Todris.

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 Mary Karlin is an author who resides in Arizona. Mary Karlin is the author of *ARTISAN CHEESE MAKING AT HOME: Techniques & Recipes for Mastering World-Class Cheeses*; *MASTERING FERMENTATION: Recipes for Making and Cooking with Fermented Foods*; and *Wood-Fired Cooking*. One or more of Karlin'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 Karlin's works. Karlin'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 Dave Kerpen is an author who resides in New York. Dave Kerpen is the author of *Get Over Yourself: How to Lead and Delegate Effectively for More Time, More Freedom, and More Success*; *Likeable Social Media, Third Edition: How to Delight Your Customers, Create an Irresistible Brand, & Be Generally Amazing on All Social Networks That Matter, Third Edition*; *Likeable Social Media: How to Delight Your Customers, Create an Irresistible Brand, and Be Generally Amazing on Facebook (and Other Social Networks)*; and *The Art of People: The 11 Simple People Skills That Will Get You Everything You Want*. One or more of Kerpen'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 Kerpen's works. Kerpen'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 Joseph Kertes is an author who resides in Canada. Joseph Kertes is the author of *Gratitude* and *The Afterlife of Stars*. One or more of Kertes'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 Kertes's works. Kertes'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 Philippe Kruchten is an author who resides in Canada. Philippe Kruchten is the author of *The Rational Unified Process - An Introduction* and *The Rational Unified Process Made Easy: A Practitioner's Guide to the RUP*. One or more of Kruchten'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 Kruchten's works. Kruchten'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 Gary Lamberti is an author who resides in Indiana. Gary Lamberti is the author of *An Introduction to the Aquatic Insects of North America, Fifth Edition*; *Methods in Stream Ecology*; *Methods in Stream Ecology, 3rd Edition*; and *Methods in Stream Ecology, Third Edition*. One or more of Lamberti'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 Lamberti's works. Lamberti'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 Michelle Leder is an author who resides in California. Michelle Leder is the author of *Financial Fine Print: Uncovering a Company's True Value*. Leder'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 Leder's work. Leder'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.

55. Plaintiff Elisabeth Lloyd is an author who resides in Indiana. Elisabeth Lloyd is the author of *A Semantic Approach to the Structure of Evolutionary Theory*; *Keywords in Evolutionary Biology*; *Normality and Variation: The Human Genome Project and the Ideal Human Type*; *The Case of the Female Orgasm: Bias in the Science of Evolution*; and *The Structure and Confirmation of Evolutionary Theory*. One or more of Lloyd'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 Lloyd's works. Lloyd'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 Jeanne Marecek is an author who resides in California. Jeanne Marecek is the author of *APA Handbook of the Psychology of Women Volume 1 History, Theory, and Battlegrounds*; *Gender, Sex, and Sexualities: Psychological Perspectives*; *Making a Difference: Psychology and the Construction of Gender*; *Transnational Psychology of Women: Expanding International and Intersectional Approaches*; and *Women and Sex Roles: A Psychological Perspective*. One or more of Marecek'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 Marecek's works. Marecek'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.

57. Plaintiff Oleksandr Maslov is an author who resides in New York. Oleksandr Maslov is the author of *Veterans: Faces of World War II*. Maslov'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 Maslov's work. Maslov'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.

58. Plaintiff Rob McEntarffer is an author who resides in Nebraska. Rob McEntarffer is the author of *Activities for Teaching Statistics and Research Methods a Guide for Psychology Instructors*; *Assessing Undergraduate Learning in Psychology: Strategies for Measuring and Improving Student Performance*; *Barron's How to Prepare for the AP Psychology Advanced Placement Exam*; *Barron's How to Prepare for the AP Psychology, Advanced Placement Examination*; and *Blair-Broeker Thinking About Psychology 4E Downloadable Test Bank*. One or more of McEntarffer'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 McEntarffer's works. McEntarffer'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 Marsha J. Means is an author who resides in Arizona. Marsha J. Means is the author of *Your Sexually Addicted Spouse: How Partners Can Cope and Heal*. Means's work is 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 Means's work. Means'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 Rachel Nuwer is an author who resides in New York. Rachel Nuwer is the author of *I Feel Love: MDMA and the Quest for Connection in a Fractured World* and *Poached: Inside the Dark World of Wildlife Trafficking*. One or more of Nuwer'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 Nuwer's works. Nuwer'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 Joshua Patrick is an author who resides in North Carolina. Joshua Patrick is the author of *Angelos Odyssey*. Patrick'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 Patrick's work. Patrick'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 Minxin Pei is an author who resides in California. Minxin Pei is the author of *China's Crony Capitalism: The Dynamics of Regime Decay*; *China's Trapped Transition: The Limits of Developmental Autocracy*; *From Reform to Revolution: The Demise of Communism in China and the Soviet Union*; and *The Sentinel State: Surveillance and the Survival of Dictatorship in China*. One or more of Pei'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 Pei's works. Pei'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 Edward Pollack is an author who resides in New York. Edward Pollack is the author of *Dynamic SQL: Applications, Performance, and Security*. Pollack'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 Pollack's work. Pollack'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.

64. Plaintiff David Rollins is an author who resides in Australia. David Rollins is the author of *A Knife Edge*; *Hard Rain*; and *The Death Trust*. One or more of Rollins's works are contained in pirated online libraries, including Books3 (and thus The Pile), 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 Rollins's works. Rollins'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.

65. Plaintiff Bob Rosner is an author who resides in Washington. Bob Rosner is the author of *Gray Matters*; *International Business: Competing in the Global Marketplace, Eighth Edition*; *The Boss's Survival Guide*; *The Boss's Survival Guide: Workplace 911 for the Toughest Problems Today's Managers Face, Second Edition*; and *Working Wounded: Advice That Adds Insight to Injury*. One or more of Rosner'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 Rosner's works. Rosner'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 Tony Rothman is an author who resides in New Jersey. Tony Rothman is the author of *A Little Book About the Big Bang by Tony Rothman*; *A Physicist on Madison Avenue*; *Doubt and Certainty*; *Everything's Relative: And Other Fables from Science and Technology*[5]; *Frontiers of Modern Physics: New Perspectives on Cosmology, Relativity, Black Holes, and Extraterrestrial Intelligence*; *Instant Physics: From Aristotle to Einstein, and Beyond*; *Sacred Mathematics: Japanese Temple Geometry*; *Science a La Mode: Physical Fashions and Fictions*; and *The World Is Round*. One or more of Rothman's works are contained in pirated online libraries,

[5] This title reflects an error included in the U.S. Copyright Office registration data; the correct title is *Everything's Relative and Other Fables from Science and Technology*.

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 Rothman's works. Rothman'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 Theodore Savas is an author who resides in South Carolina. Theodore Savas is the author of *Hunt and Kill: U-505 and the U-Boat War in the Atlantic*; *Never for Want of Powder: The Confederate Powder Works in Augusta, Georgia*; and *The New American Revolution Handbook: Facts and Artwork for Readers of All Ages, 1775-1783*. One or more of Savas'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 Savas's works. Savas'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.

68. Plaintiff Todd Shuster is an author who resides in New York. Todd Shuster is the author of *Real Boys' Voices* and *The First Day of Peace*. One or more of Shuster'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 Shuster's works. Shuster'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.

69. Plaintiff Janet Sorensen is an author who resides in California. Janet Sorensen is the author of *Strange Vernaculars: How Eighteenth-Century Slang, Cant, Provincial Languages, and Nautical Jargon Became English*. Sorensen'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 Sorensen's work. Sorensen'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.

70. Plaintiff Mark Spragg is an author who resides in Wyoming. Mark Spragg is the author of *An Unfinished Life: A Novel*; *Bone Fire*; *The Fruit of Stone*; *Thunder of the Mustangs*; and *Where Rivers Change Direction*. One or more of Spragg'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 Spragg's works. Spragg'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.

71. Plaintiff Justin Velez-Hagan is an author who resides in Virginia. Justin Velez-Hagan is the author of *The Common Sense Behind Basic Economics: A Guide for Budding Economists, Students, and Voters* and *The Impact of Fiscal Consolidation Policies on Economic Performance*. One or more of Velez-Hagan'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 Velez-Hagan's works. Velez-Hagan'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 Richard Waters is an author who resides in Massachusetts. Richard Waters is the author of *Readings in Artificial Intelligence and Software Engineering* and *The Programmer's Apprentice*. One or more of Waters'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 Waters's works. Waters'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. Plaintiff Robert Wettemann is an author who resides in Colorado. Robert Wettemann is the author of *Privilege vs. Equality: Civil-Military Relations in the Jacksonian Era, 1815-1845* and *The Texas You Expect: The Story of Buffalo Gap Historic Village*. One or more of Wettemann'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 Wettemann's works. Wettemann'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.

74. Plaintiff Bettina Williford is an author who resides in Texas. Bettina Williford is the author of *Illegal* and *Moose and Magpie*. One or more of Williford'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 Williford's works. Williford'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.

75. Plaintiff David Wolman is an author who resides in Hawaii. David Wolman is the author of *Aloha Rodeo*; *End of Money*; *Left-Hand Turn Around the World: Chasing the Mystery and Meaning of All Things Southpaw*; *Righting the Mother Tongue*; and *The End of Money*. One or more of Wolman'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 Wolman's works. Wolman'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.

76. Plaintiff J. Dylan Yates is an author who resides in California. J. Dylan Yates is the author of *Taste*; *The Belief in Angels*; and *The Belief in Angels: Jules*. One or more of Yates'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 Yates's works. Yates'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.

77. A non-exhaustive list of registered copyrights owned by Plaintiffs is included as Exhibit A (herein, the "Plaintiffs' Works").[6]

B. Defendant

78. 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

79. 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.

80. 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 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.

81. 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.

[6] 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.

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

82. 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.

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

84. 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.

85. 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"

86. "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.

87. 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.

88. 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."[7] 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."[8] For purposes of this Complaint, "training" refers to all stages of the LLM training process.

89. 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 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

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

[8] *Id.*

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.

90. 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."[9]

91. 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."[10]

92. 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.'"[11]

93. In terms of data quality, "[r]ecent research from major developers suggests that quality may even be a more important consideration than quantity."[12] "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 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.

---

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

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

[11] *Id.* at 10.

[12] *Id.* at 11.

94. 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.[13] And, as the Court found in that litigation, Meta considered spending up to $100 million to license copyrighted materials in connection with LLM training.[14]

95. 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.

96. 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.[15] For these companies, staying ahead of competitors is "code red."[16] 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 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.

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

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

[15] *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).

[16] *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.

97. 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."[17]

98. 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

99. 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.[18] And they do all of this without authorization from authors or publishers.[19] In short, shadow libraries facilitate the reproduction and distribution of unauthorized copies of copyrighted works at industrial scale.

100. 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 States trade authorities. As centralized shadow libraries increasingly faced enforcement actions,

---

[17] 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.

[18] *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.

[19] *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.

including the seizure of domains, third parties responded by creating full mirrored copies of those repositories for decentralized redistribution.[20]

101. In 2018, an OpenAI employee downloaded pirated copies of books from Library Genesis, or "LibGen"—a shadow library repeatedly enjoined by federal courts[21]—and used those books to create two internal datasets OpenAI called "LibGen1" and "LibGen2," which OpenAI publicly referred to as "Books1" and "Books2."[22] 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.

102. 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.[23]

103. 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 approximately 196,640 books.[24] Books3 comprises about 12 percent (just over 100 gigabytes) of *The Pile*.[25]

---

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

[21] *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).

[22] *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.

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

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

[25] *Id.*

104. 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,"[26] 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."[27]

105. 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.[28]

106. 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."[29] 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.

107. 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 Books3," he said. "But the alternative is that, without Books3, only OpenAI can do what they're doing."[30]

---

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

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

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

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

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

108. 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.[31]

109. 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.

110. 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.[32] 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.[33] LibGen has been repeatedly enjoined by federal courts for copyright infringement and has been designated a "notorious market" by the United States Trade Representative.[34] Despite enforcement actions, LibGen has remained accessible through shifting domains, mirrors, and downloadable archives. Its persistence is a function of deliberate decentralization designed to evade shutdown.[35]

---

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

[32] 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.").

[33] *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.

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

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

111. 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.[36] Z-Library offered premium features—including faster downloads and higher volume limits—in exchange for payment, operating in effect as a commercial piracy service.[37] In 2022, Z-Library's domains were seized by law-enforcement authorities, and its operators were arrested and later indicted for criminal copyright infringement.[38] 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.[39]

112. 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.[40]

113. 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 complete collections of LibGen, Z-Library, and other pirated sources.[41] Anna's Archive functions

---

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

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

[38] 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.

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

[40] *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.

[41] *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.

as a meta-library: it indexes, mirrors, and redistributes multiple shadow libraries simultaneously, offering users unified access to millions of pirated books.[42]

114. Anna's Archive offers paid tiers that provide high-speed or priority access to its pirated collections.[43] 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.[44]

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

116. 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."[45] 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."[46]

117. 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 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")

---

[42] *Id.*

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

[44] *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.

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

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

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

118. 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,"[47] and it in turn expects Llama to generate somewhere in the realm of $460 billion to $1.4 trillion in revenue.[48] None of the money that Meta invested into the development of its Llama models was used to pay for Plaintiffs' copyrighted works.[49]

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

120. 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.[50] To accomplish that, Meta needed to train on large quantities of high-quality books and other published materials.

121. 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."[51]

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

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

[49] *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").

[50] 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.").

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

122. 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."[52]

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

124. 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.

125. In connection with Llama 1's release, Meta disclosed the composition of the model's "pre-training data."[53] Among the datasets used to train Llama 1 were "two book corpora," including "the Books3 section of ThePile."[54] 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."[55] This distinction was deliberate and notable, as Meta *knew* that Books3 was *not* a corpus of books in the "public domain."

126. 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 available C4 dataset," comprising 15 percent of the data mix.[56] "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

---

[52] *Id.*

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

[54] *Id.*

[55] *Id.*

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

month."[57] The C4 dataset itself is derived from CommonCrawl.[58] In short, a massive amount (and the majority) of data used by Meta to train Llama was "scraped" from the Internet.

127. 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.[59] 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."[60]

128. 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.

129. "In early 2024, Meta also downloaded Anna's Archive, a compilation of shadow libraries including LibGen, Z-Library, and others."[61]

130. Meta likewise downloaded copyrighted materials from Z-Library, Sci-Hub, and other shadow libraries.[62] Indeed, by April 2024, Meta had downloaded over 25 TB of data from Z-Library and 10 TB of data from LibGen.[63]

[57] 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.

[58] *Id.*

[59] *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").

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

[61] *Id.*

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

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

131. To obtain the massive quantities of data from these shadow libraries, Meta used torrenting protocols.[64] 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."[65]

132. 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.

133. 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.

134. 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.

135. 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

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

[65] *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.").

materials from shadow libraries. And on information and belief, certain employees implicitly or explicitly raised concerns with torrenting from shadow libraries.

136. 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.

137. 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.

138. 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.

139. 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.

140. As Meta has admitted elsewhere, Meta "created a script to remove repetitive text from the training data for its Llama models."[66] As admitted elsewhere, Meta's "program was designed to remove lines that contain the word 'copyright.'"[67]

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

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

[67] *Kadrey*, ECF No. 485, ¶ 89.

142. As admitted elsewhere, Meta used at least Books3 and data it obtained from LibGen and Z-Library to train the Llama family of models.[68] 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.

143. 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.[69] 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."[70] In other words, if a book was in LibGen, Meta downloaded it.

144. 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.[71]

145. 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

---

[68] *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").

[69] 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.").

[70] *Id.* at 12–14.

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

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

146. 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.[72] Meta internally referred to this book-acquisition and scanning operation as "Project Vellum," and it is referred to as such herein.[73]

147. 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.[74]

148. 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, 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

[72] Meta Interrog. Resps. at 23.

[73] *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.

[74] 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).

other third parties.[75] Meta has admitted that books it acquired and scanned as part of Project Vellum were used to train the Muse Spark family of models.[76]

149. 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.

150. 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).

151. 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.[77] Meta has admitted that "[a] number of transitory and persistent copies of each of these 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.[78]

---

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

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

[77] *Id.* at 18.

[78] *Id.* at 24.

152. 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."[79] In other words, Meta scanned each book's copyright page, ingested the copyright notice, and trained on the work anyway.

153. 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.

154. 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.

155. 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.

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

156. 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

[79] *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").

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.

157. *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.

158. *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."[80]

159. *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 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.

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

V. CLAIMS FOR RELIEF

COUNT I

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

160. Plaintiffs incorporate the allegations above.

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

162. 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.

163. 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.

164. 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.

165. 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.[81]

166. 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.

167. 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.

168. 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

169. Plaintiffs incorporate the allegations above.

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

171. 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).

172. 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

[81] 17 U.S.C. § 506(a)(1)(A).

Plaintiffs to scan, digitize, copy, or otherwise reproduce Plaintiffs' Works, or to use Plaintiffs' Works as training data.

173. 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.

174. 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.

175. 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.

176. 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.

177. 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.

178. 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)

179. Plaintiffs incorporate the allegations above.

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

181. 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.

182. 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.

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

184. 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.

185. 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))

186. Plaintiffs incorporate the allegations above.

187. 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.

188. 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.

189. 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.

190. 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.

191. 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