

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

Counsel for Plaintiffs

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

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

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

Civil Case No.:

COMPLAINT

DEMAND FOR JURY TRIAL

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

4 Plaintiffs,

5 v.

6 META PLATFORMS, INC.,

7 Defendant.

8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

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

13 **I. INTRODUCTION**

14 2. This case concerns Defendant Meta’s exploitation of Plaintiffs’ copyrighted works
 15 to build its “Llama” and “Muse” families of large language models (“LLMs”). Defendant is in open
 16 and active competition with its tech company rivals to win what many have deemed the “generative
 17 AI arms race.” Participants in this race are under immense pressure to build more, better, and faster,
 18 with an eye towards ensuring that *its* generative AI models are widely adopted. Broad adoption of
 19 a company’s generative AI models will yield great profits, the thinking goes; on the converse, failure
 20 to move quickly means missing the boat on this trillion-dollar industry. To remain competitive in
 21 this race, Meta turned to notorious online “shadow libraries” and similar pirated datasets to obtain
 22 vast quantities of copyrighted materials—books, articles, training materials, and the like—which it
 23 desperately wanted to train and optimize its LLM models. Meta could have—and *should have*—
 24 paid copyright owners, including Plaintiffs, for licenses to use their copyrighted works in connection
 25 with training its Llama and Muse models. Instead, Meta fed from, and then back into, the dark
 26 market for digitally available copyrighted materials. Ultimately, Meta downloaded, reproduced,
 27 distributed, and defaced these copyrighted works, including Plaintiffs’ Works, to get ahead in the
 28

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

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

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

19 5. For its Llama training datasets, Meta needed gold standard training data. To get its
20 hands on published content, Meta considered and explored the possibility of licensing copyrighted
21 materials for use in its LLM training. The company’s head of generative AI even discussed spending
22 up to \$100 million on licensing such materials. Ultimately, however, Meta abandoned this idea in
23 favor of a faster and cheaper approach: downloading copyrighted materials from Internet-based
24 “shadow libraries” and other datasets that contained the contents of these shadow libraries. These
25 shadow libraries, each of which contained tens or hundreds of thousands of unauthorized digital
26

27 ¹ Unless otherwise indicated, references to Meta’s “Llama” and “Muse” refer to all versions of
28 Llama and Muse, respectively, in any stage of their development or deployment.

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

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

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

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

22 9. Meta’s infringement did not stop at shadow libraries. Beginning in or around mid-
23 2025—after it had been sued repeatedly for training Llama on pirated books—Meta launched a
24 book-scanning operation, referred to herein as “Project Vellum,” through which it acquired physical
25 copies of books en masse, scanned them into digital form, ran the scanned images through optical
26 character recognition software to extract their text, and fed the resulting digital copies into the
27 training of its next-generation “Muse” family of LLMs, including Muse Spark (released in April
28 2026) and Muse Glimmer (released in August 2026).

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

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

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

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

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

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

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

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

10 **II. PARTIES**

11 **A. Plaintiffs**

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

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

27
 28 ² *Feltner v. Columbia Pictures Television, Inc.*, 523 U.S. 340, 353 (1998).

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

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

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

19 21. Plaintiff Matthew David Brozik is an author who resides in New York. Matthew
20 David Brozik is the author of *Danger ...with a Hard G*; *Odder Space*; *Spelled Wrong*; *Taking Ivy*
21 *Seriously*; *The Admirable Vice Provost and Other Stories*; *The Government Manual for New Pirates*;
22 *The Government Manual for New Superheroes*; *The Government Manual for New Wizards*; *The*
23 *Vowels of the Earth*; *Twin Freaks*; *Whimsy & Soda*; and *Xavier, the Creative Little Gator*. One or
24 more of Brozik's works are contained in pirated online libraries, including Z-Library. Meta has
25 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
26 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
27 Brozik's works. Brozik's works were also published in print and were commercially available for
28

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

21 37. Plaintiff Wayne Gladstone is an author who resides in New York. Wayne Gladstone
22 is the author of *Agents of the Internet Apocalypse*; *Notes from the Internet Apocalypse*; and *Reports*
23 *on the Internet Apocalypse*. One or more of Gladstone's works are contained in pirated online
24 libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has
25 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
26 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
27 Gladstone's works. Gladstone's works were also published in print and were commercially available
28

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

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

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

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

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

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

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

28

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

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

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

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

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

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

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

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

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

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

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

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

24 54. Plaintiff Robert Ostergren is an author who resides in Wisconsin. Robert Ostergren
 25 is the author of *A Community Transplanted: The Trans-Atlantic Experience of a Swedish Immigrant*
 26 *Settlement in the Upper Middle West, 1835-1915*; *Building Nazi Germany: Place, Space,*
 27 *Architecture, and Ideology*; *Europeans, Second Edition: A Geography of People, Culture, and*
 28

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

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

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

25
26
27 _____
28 ³ This title reflects an error included in the U.S. Copyright Office registration data; the correct title
is *The Europeans, Second Edition: A Geography of People, Culture, and Environment*.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

11 **B. Defendant**

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

19 **III. JURISDICTION AND VENUE**

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

23 76. This Court has personal jurisdiction over the Defendant. The Defendant has
 24 purposefully availed itself of the privilege of conducting business in this District and the State of
 25 California. The Defendant committed acts of copyright infringement in this District, directed
 26

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

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

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

10 **IV. FACTUAL ALLEGATIONS**

11 **A. Plaintiffs and their Protected Works**

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

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

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

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

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

4 **B. LLMs and the Generative AI "Arms Race"**

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

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

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

23 85. LLMs typically are trained, at least initially, by feeding them massive amounts of
 24 text data from which they can "learn" statistical relationships between and among text. In the process
 25

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

28 ⁶ *Id.*

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

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

13 87. In determining the corpus of training data for an LLM, a model’s developer typically
 14 considers “the quantity of data, its quality, and the ultimate purpose(s) of the model.”⁸

15 88. In terms of data quantity, the LLMs that have been developed and are being
 16 developed by leading technology companies are trained on enormous datasets, typically on the scale
 17 of terabytes. This is because AI researchers have found that “increasing the quantity of training data
 18 typically increases a model’s ‘performance.’”⁹

19 89. In terms of data quality, “[r]ecent research from major developers suggests that
 20 quality may even be a more important consideration than quantity.”¹⁰ “Garbage in, garbage out,”
 21 the saying goes. AI researchers quickly discovered that published content—books in particular—
 22 makes for some of the best training material for LLMs. Unlike much content on the Internet and
 23

24 ⁷ James Betker, *The “it” in AI Models is the Dataset*, (June 10, 2023), <https://perma.cc/ZCH9-H53S>.

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

27 ⁹ *Id.* at 10.

28 ¹⁰ *Id.* at 11.

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

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

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

15 92. As discussed in the introduction, technology companies have treated generative AI
 16 as the next foundational layer of the digital economy. Industry leaders publicly describe an “AI arms
 17 race,” in which they have redirected their corporate strategies to seize control of what they believe
 18 will become a new infrastructure layer for commerce, communication, and knowledge work.¹³ For
 19 these companies, staying ahead of competitors is “code red.”¹⁴ Among other things, being too slow
 20 out of the blocks could mean ending up in last place; if by the time you publicly released *your* LLM
 21

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

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

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

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

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

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

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

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

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

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

20 _____
21 ¹⁵ Mike Proulx, *Meta: So Long, Metaverse; Hello, Superintelligence*, FORRESTER (July 30, 2025),
22 <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>.

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

26 ¹⁷ See Riddhi Setty, *Rampant 'Shadow Libraries' Drive Calls for Anti-Piracy Action*, BLOOMBERG
27 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,
28 2022, 11:38 AM), <https://perma.cc/K9FA-VLPW>.

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

4 97. In 2018, an OpenAI employee downloaded pirated copies of books from Library
 5 Genesis, or “LibGen”—a shadow library repeatedly enjoined by federal courts¹⁹—and used those
 6 books to create two internal datasets OpenAI called “LibGen1” and “LibGen2,” which OpenAI
 7 publicly referred to as “Books1” and “Books2.”²⁰ OpenAI used those pirated corpora to train GPT-
 8 3, which it released in June 2020 to widespread commercial acclaim. OpenAI’s use of a books
 9 corpora from a pirate library established the model for how to quickly and cheaply obtain published
 10 materials for use as LLM training data.

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

16 99. Because OpenAI did not disclose the precise makeup of its training datasets,
 17 members of EleutherAI constructed a book corpus of their own: “Books3,” consisting of
 18
 19
 20
 21

22 ¹⁸ Woodcock, ‘Shadow Libraries’ Are Moving Their Pirated Books to the Dark Web After Fed
 Crackdowns.

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

25 ²⁰ *In re OpenAI, Inc., Copyright Infringement Litig.*, Case No. 1:2025-md-03143 (S.D.N.Y. 2025),
 26 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>.

27 ²¹ Leo Gao et al., *The Pile: An 800GB Dataset of Diverse Text for Language Modeling*, arXiv, 1
 28 (2020), <https://perma.cc/NHV6-R8YE>.

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

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

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

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

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

18 [Presser] created Books3 in the hope that it would allow any developer to create
 19 generative-AI tools. “It would be better if it wasn’t necessary to have something like
 20
 21

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

23 ²³ *Id.*

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

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

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

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

Books3,” he said. “But the alternative is that, without Books3, only OpenAI can do what they’re doing.”²⁸

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

³⁷ See Woodcock, *‘Shadow Libraries’ Are Moving Their Pirated Books to The Dark Web*.

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

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

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

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

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

13 113. Many of these shadow libraries and datasets can be downloaded using “torrent,” a
 14 file-sharing method used in peer-to-peer networks. Torrenting works by breaking a file into many
 15 small pieces and distributing those pieces across a network of participating computers. A user who
 16 torrents a shadow-library repository does not receive a single copy from a single source; rather, the
 17 user downloads portions of the library from numerous other computers that already possess the
 18

19 ³⁹ See Ernesto Van de Sar, “*Anna’s Archive*” *Opens the Door to Z-Library and Other Pirate*
 20 *Libraries*, TORRENTFREAK (Nov. 19, 2022), <https://perma.cc/U88R-WTR4>.

21 ⁴⁰ *Id.*

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

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

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

27 ⁴⁴ See *If you’re an LLM, please read this*, ANNA’S ARCHIVE (Feb. 18, 2026), [https://perma.cc/MJ7Z-](https://perma.cc/MJ7Z-3ZCL)
 28 [3ZCL](https://perma.cc/MJ7Z-3ZCL).

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

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

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

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

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

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

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

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

27 ⁴⁸ Jon Russell, *Mark Zuckerberg Announces New Team at Meta Working on A.I. Products for*
 28 *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.").

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

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

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

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

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

20 122. Among the other datasets used by Meta to train Llama 1 were “five CommonCrawl
21 dumps, ranging from 2017 to 2020,” comprising 67 percent of the overall data mix, and “the publicly
22

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

24 ⁵⁰ *Id.*

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

27 ⁵² *Id.*

28 ⁵³ *Id.*

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

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

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

15 125. “In early 2024, Meta also downloaded Anna’s Archive, a compilation of shadow
 16 libraries including LibGen, Z-Library, and others.”⁵⁹

17
18
19
20
21 ⁵⁴ Touvron et al., *LLaMA: Open and Efficient Foundation Language Models*, arXiv, 2.

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

24 ⁵⁶ *Id.*

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

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

28 ⁵⁹ *Id.*

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

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

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

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

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

21 _____
 22 ⁶⁰ *Cambronne*, ECF No. 109 (Meta Answer), ¶ 83.

23 ⁶¹ *Kadrey*, ECF No. 472 (Pls.’ Partial Mot. for Summ. J.), at 11.

24 ⁶² *Cambronne*, ECF No. 109 at ¶ 83 (admitting that “it used BitTorrent to download certain portions
 25 of these publicly available datasets”).

26 ⁶³ *See Kadrey*, 788 F. Supp. 3d at 1041 (“To download these large datasets more quickly and without
 27 unnecessarily slowing down its networks, Meta torrented them. . . . There is no dispute that Meta
 28 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.”).

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

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

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

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

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

21 136. As Meta has admitted elsewhere, Meta “created a script to remove repetitive text
22 from the training data for its Llama models.”⁶⁴ As admitted elsewhere, Meta’s “program was
23 designed to remove lines that contain the word ‘copyright.’”⁶⁵

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

26
27 ⁶⁴ *Kadrey*, ECF No. 485, ¶ 88 (Meta’s Answer to Third Am. Compl.).

28 ⁶⁵ *Kadrey*, ECF No. 485, ¶ 89.

1 reasonable inference that Meta downloaded, manipulated, reproduced, and distributed the Plaintiffs’
2 Works.

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

9 139. Meta has since confirmed, in interrogatory responses served in other litigation, the
10 specific pirated datasets on which each Llama model was trained. According to Meta, the Books3
11 dataset (which Meta refers to internally as “b3g”) was used to train Llama 1, Llama 2, Llama 3, and
12 Llama 4; the LibGen “Scitech” and LibGen “Fiction” datasets were used to train Llama 3 and Llama
13 4; and the Z-Library dataset, which Meta obtained by torrenting Anna’s Archive, was used to train
14 Llama 4.⁶⁷ Meta has further admitted that, beginning in or about Spring 2023, a Meta engineer
15 downloaded the Fiction and Scitech branches of LibGen using a direct-download script, and that
16 Meta’s downloads of both branches “are believed to have been substantially complete.”⁶⁸ In other
17 words, if a book was in LibGen, Meta downloaded it.

18 140. On information and belief, Meta, like Anthropic, downloaded, maintained, and
19 retained pirated copies of works that were not specifically tied to training its LLM models.⁶⁹
20
21

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

24 ⁶⁷ Def. Meta Platforms, Inc.’s Second Supp. Resps. & Objs. to Pl.’s First Set of Interrogs. at 17–18,
25 *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.”).

26 ⁶⁸ *Id.* at 12–14.

27 ⁶⁹ *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1014 (N.D. Cal. 2025) (Anthropic’s amassing of
28 “a central library”).

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

6 **E. Project Vellum: Meta’s Mass Scanning of Physical Books to Train Its Muse**
 7 **Models**

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

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

19 144. Meta used the digital copies it created through Project Vellum to train its “Muse”
 20 family of LLMs. In April 2026, Meta introduced the initial version of its Muse Spark model, which
 21 Meta made available to the public through certain Meta products and services and, in later versions,
 22

23 ⁷⁰ Meta Interrog. Resps. at 23.

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

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

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

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

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

16 147. Scanning a book is reproduction. Each time Meta scanned a physical copy of a book,
 17 it created a new digital copy of the entire work. Each time Meta ran the scanned images through
 18 OCR, it created another copy—this time as machine-readable text. Each time Meta processed,
 19 “cleaned,” de-duplicated, and tokenized that text, it created still more copies. And each time Meta
 20 used that text in a training run—including in the “number of internal models” that Meta trains “for
 21 experimental and evaluation purposes” before each public release, and in the “purpose-built
 22 variants” of its LLMs that Meta uses for discrete internal purposes—it created additional copies of
 23 the work.⁷⁵ Meta has admitted that “[a] number of transitory and persistent copies of each of these
 24

25 _____
 26 ⁷³ Meta Interrog. Resps. at 23, 25.

27 ⁷⁴ *Id.* at 18 & Ex. D.

28 ⁷⁵ *Id.* at 18.

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

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

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

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

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

25
26 ⁷⁶ *Id.* at 24.

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

1 **F. Meta’s Conduct Harms the Market for Plaintiffs’ Copyrighted Materials**

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

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

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

23 155. *Third*, even if Llama and Muse models are restricted from outputting extended
24 portions of verbatim text from copyrighted works, they are nevertheless capable of producing nearly
25
26

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

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

3 **V. CLAIMS FOR RELIEF**

4 **COUNT I**

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

7 156. Plaintiffs incorporate the allegations above.

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

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

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

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

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

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

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

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

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

14 **COUNT II**

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

16 165. Plaintiffs incorporate the allegations above.

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

19 167. Beginning in or around mid-2025, Defendant, without authorization from Plaintiffs,
 20 acquired physical copies of Plaintiffs' Works, scanned those copies into digital form, ran the
 21 scanned images through optical character recognition software to extract the text of Plaintiffs'
 22 Works, and processed, cleaned, de-duplicated, tokenized, stored, and retained the resulting digital
 23 copies. Defendant then used those digital copies to train, develop, fine-tune, evaluate, and deploy
 24 its Muse family of models. Each of these acts reproduced Plaintiffs' Works in copies, in violation
 25 of Plaintiffs' exclusive rights under 17 U.S.C. § 106(1).
 26
 27

28 ⁷⁹ 17 U.S.C. § 506(a)(1)(A).

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

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

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

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

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

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

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

COUNT III

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

175. Plaintiffs incorporate the allegations above.

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

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

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

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

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

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

21 **COUNT IV**

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

23 182. Plaintiffs incorporate the allegations above.

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

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

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

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

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

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

16 **PRAYER FOR RELIEF**

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

- 18 a. Judgment in favor of Plaintiffs against the Defendant;
- 19 b. A declaration that the Defendant has infringed Plaintiffs' exclusive copyrights
- 20 under the Copyright Act;
- 21 c. A declaration that such infringement is willful;
- 22 d. A declaration that Defendant violated 17 U.S.C. § 1202(b) through its removal
- 23 of copyright management information;
- 24 e. A permanent injunction enjoining the Defendant and all those acting in concert
- 25 with it from engaging in the infringing conduct alleged herein;
- 26 f. That the Defendant be directed to account to Plaintiffs for all gains, profits, and
- 27 advantages derived from their unlawful acts;
- 28 g. An award of statutory damages under the Copyright Act;

- 1 h. An award of statutory or actual damages under 17 U.S.C. § 1203(c);
2 i. An award of restitution, disgorgement, costs, expenses, and attorneys' fees as
3 permitted by law (including those allowable under 17 U.S.C. § 505 and/or 17
4 U.S.C. § 1203(b)(4)–(5)).
5 j. Pre- and post-judgment interest on the damages awarded to Plaintiffs; and
6 k. Further relief for Plaintiffs as the Court may deem just and proper.

7 **JURY TRIAL DEMANDED**

8 Under Federal Rule of Civil Procedure 38(b), Plaintiffs demand a trial by jury.
9

10 Dated: October 19, 2026

Respectfully submitted,

11 /s/ Elizabeth Brannen

12 Elizabeth Brannen (SBN 226234)

13 Lauren Martin (SBN 294367)

STRIS & MAHER LLP

14 17785 Center Court Dr N, Ste 600

Cerritos, CA 90703

15 T: (213) 995-6800

16 F: (213) 261-0299

ebrannen@stris.com

17 lmartin@stris.com

18 Kyle Roche (*pro hac vice* forthcoming)

FREEDMAN NORMAND FRIEDLAND LLP

19 155 E. 44th Street, Ste 915

20 New York NY 10017

T: (646) 494-2900

21 kroche@fnf.law

22 *Counsel for Plaintiffs*
23
24
25
26
27
28