

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

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

Civil Case No.:

COMPLAINT

DEMAND FOR JURY TRIAL

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

5 Plaintiffs,

6 v.

7 META PLATFORMS, INC.,

8 Defendant.

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

1 1. Plaintiffs Jacqueline Winspear, Blaise Aguirre, John Bair, Robert Bell, Alexandra
 2 Berlina, Goldie Browning, Aaron Bubbico, Jeff Carollo, Christine Carpenito, Josephine Caust, Gene
 3 Cogorno, Jessa Crispin, Harvey Deutschendorf, Stephen Diamond, Roger Dooley, John Ebert, Carol
 4 Ekarius, Julen Etxabe, Robert Ficociello, Donald Firesmith, Silvia Foti, Eric Gale, Dawn Gilpin,
 5 Laura Gompertz, Reyna Grande, Jeannette Graulau, Raymond Greenwell, Daniel Gumbiner, Owen
 6 Gutfreund, Andrew Hargreaves, Paul Herman, Gordon Inkeles, Mary Karlin, Dave Kerpen, Joseph
 7 Kertes, Philippe Kruchten, Gary Lamberti, Michelle Leder, Elisabeth Lloyd, Jeanne Marecek,
 8 Oleksandr Maslov, Rob McEntarffer, Marsha J. Means, Rachel Nuwer, Joshua Patrick, Minxin Pei,
 9 Edward Pollack, David Rollins, Bob Rosner, Tony Rothman, Theodore Savas, Todd Shuster, Janet
 10 Sorensen, Mark Spragg, Justin Velez-Hagan, Richard Waters, Robert Wettemann, Bettina Williford,
 11 David Wolman, and J. Dylan Yates (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 Jacqueline Winspear is an author who resides in Washington. Jacqueline
 13 Winspear is the author of *A Dangerous Place*; *A Sunlit Weapon*; *Among the Mad*; *An Incomplete*
 14 *Revenge*; *An Incomplete Revenge: A Maisie Dobbs Novel*; *Birds of a Feather: A Novel*; *Elegy for*
 15 *Eddie*; *In This Grave Hour: A Maisie Dobbs Novel*; *Journey to Munich*; *Leaving Everything Most*
 16 *Loved*; *Maisie Debbs*³; *Maisie Dobbs*; *Messenger of Truth: A Maisie Dobbs Novel*; *Pardonable*
 17 *Lies: A Maisie Dobbs Novel*; *The American Agent*; *The Care and Management of Lies*; *The Comfort*
 18 *of Ghosts*; *The Consequences of Fear*; *The Mapping of Love and Death*; *The White Lady*; *This Time*
 19 *Next Year We’ll Be Laughing*; *To Die but Once*; and *What Would Maisie Do?* One or more of
 20 Winspear’s works are contained in pirated online libraries, including Books3 (and thus The Pile),
 21 LibGen, Z-Library, and Anna’s Archive. Meta has directly or indirectly downloaded books illegally
 22 contained in The Pile, LibGen, Z-Library, or Anna’s Archive, and there is accordingly a reasonable
 23 inference that Meta illegally downloaded Winspear’s works. Winspear’s works were also published
 24
 25

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

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

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

3 18. Plaintiff Blaise Aguirre is an author who resides in Massachusetts. Blaise Aguirre is
4 the author of *Borderline Personality Disorder in Adolescents; Coping with BPD: DBT and CBT*
5 *Skills to Soothe the Symptoms of Borderline Personality Disorder; Depression; Fighting Back: What*
6 *an Olympic Champion's Story Can Teach Us About Recognizing and Preventing Child Sexual*
7 *Abuse—and Helping Kids Recover; Helping Your Troubled Teen; Parenting Your Child with*
8 *Autism; and The Unwanted Thoughts and Intense Emotions Workbook: CBT and DBT Skills to*
9 *Break the Cycle of Intrusive Thoughts and Emotional Overwhelm*. One or more of Aguirre's works
10 are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library,
11 and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The
12 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
13 Meta illegally downloaded Aguirre's works. Aguirre's works were also published in print and were
14 commercially available for purchase in physical form during the period in which Meta conducted
15 Project Vellum, as alleged below.

16 19. Plaintiff John Bair is an author who resides in Washington. John Bair is the author
17 of *Hiding Behind the Keyboard, First Edition* and *Seeking the Truth from Mobile Evidence: 1st*
18 *Edition*. One or more of Bair's works are contained in pirated online libraries, including Books3
19 (and thus The Pile), LibGen, and Z-Library. Meta has directly or indirectly downloaded books
20 illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a
21 reasonable inference that Meta illegally downloaded Bair's works. Bair's works were also published
22 in print and were commercially available for purchase in physical form during the period in which
23 Meta conducted Project Vellum, as alleged below.

24 20. Plaintiff Robert Bell is an author who resides in North Carolina. Robert Bell is the
25 author of *America's Disaster Culture: The Production of Natural Disasters in Literature and Pop*
26 *Culture* and *Eco Culture: Disaster, Narrative, Discourse*. One or more of Bell's works are contained
27 in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or
28 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,

1 and there is accordingly a reasonable inference that Meta illegally downloaded Bell's works. Bell'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 21. Plaintiff Alexandra Berlina is an author who resides in Germany. Alexandra Berlina
5 is the author of *Viktor Shklovsky: A Reader*. Berlina's work is contained in pirated online libraries,
6 including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly
7 or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's
8 Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Berlina's
9 work. Berlina's work was also published in print and was commercially available for purchase in
10 physical form during the period in which Meta conducted Project Vellum, as alleged below.

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

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

25 24. Plaintiff Jeff Carollo is an author who resides in California. Jeff Carollo is the author
26 of *How Google Tests Software*. Carollo's work is contained in pirated online libraries, including
27 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
28 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable

1 inference that Meta illegally downloaded Carollo's work. Carollo's work was also published in print
2 and was commercially available for purchase in physical form during the period in which Meta
3 conducted Project Vellum, as alleged below.

4 25. Plaintiff Christine Carpenito is an author who resides in Massachusetts. Christine
5 Carpenito is the author of *Back to the Dream; Dreamlands; I'm Not Who You Think I Am; The Secret*
6 *of the Sheikh's Betrothed; Twisted Dreams; and Uneasy Dreams*. One or more of Carpenito's works
7 are contained in pirated online libraries, including Books3 (and thus The Pile), Z-Library, and
8 Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile,
9 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
10 illegally downloaded Carpenito's works. Carpenito's works were also published in print and were
11 commercially available for purchase in physical form during the period in which Meta conducted
12 Project Vellum, as alleged below.

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

20 27. Plaintiff Gene Cogorno is an author who resides in California. Gene Cogorno is the
21 author of *Geometric Dimensioning and Tolerancing for Mechanical Design; Geometric*
22 *Dimensioning and Tolerancing for Mechanical Design, Second Edition; and Geometric*
23 *Dimensioning and Tolerancing for Mechanical Design, Third Edition*. One or more of Cogorno's
24 works are contained in pirated online libraries, including LibGen and Z-Library. Meta has directly
25 or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's
26 Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Cogorno's
27 works. Cogorno's works were also published in print and were commercially available for purchase
28 in physical form during the period in which Meta conducted Project Vellum, as alleged below.

1 28. Plaintiff Jessa Crispin is an author who resides in Pennsylvania. Jessa Crispin is the
2 author of *My Three Dads: Patriarchy on the Great Plains* and *The Creative Tarot*. One or more of
3 Crispin'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 Crispin's works. Crispin'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 29. Plaintiff Harvey Deutschendorf is an author who resides in Canada. Harvey
10 Deutschendorf is the author of *Of Work and Men: How Men Can Become More Than Their Careers*
11 and *The Other Kind of Smart: Simple Ways to Boost Your Emotional Intelligence for Greater*
12 *Personal Effectiveness and Success*. One or more of Deutschendorf's works are contained in pirated
13 online 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 Deutschendorf's works.
16 Deutschendorf's works were also published in print and were commercially available for purchase
17 in physical form during the period in which Meta conducted Project Vellum, as alleged below.

18 30. Plaintiff Stephen Diamond is an author who resides in California. Stephen Diamond
19 is the author of *Anger, Madness, and the Daimonic: The Psychological Genesis of Violence, Evil,*
20 *and Creativity*. Diamond's work is contained in pirated online libraries, including LibGen and Z-
21 Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen,
22 Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
23 downloaded Diamond's work. Diamond's work was also published in print and was commercially
24 available for purchase in physical form during the period in which Meta conducted Project Vellum,
25 as alleged below.

26 31. Plaintiff Roger Dooley is an author who resides in Texas. Roger Dooley is the author
27 of *Brainfluence* and *FRICITION - The Untapped Force That Can Be Your Most Powerful Advantage*.
28 One or more of Dooley's works are contained in pirated online libraries, including Books3 (and thus

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

6 32. Plaintiff John Ebert is an author who resides in New Mexico. John Ebert is the author
7 of *Dead Celebrities, Living Icons: Tragedy and Fame in the Age of the Multimedia Superstar*; *The*
8 *Age of Catastrophe: Disaster and Humanity in Modern Times*; *The New Media Invasion: Digital*
9 *Technologies and the World They Unmake*; and *Twilight of the Clockwork God: Conversations on*
10 *Science and Spirituality at the End of an Age*. One or more of Ebert's works are contained in pirated
11 online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
12 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
13 is accordingly a reasonable inference that Meta illegally downloaded Ebert's works. Ebert's works
14 were also published in print and were commercially available for purchase in physical form during
15 the period in which Meta conducted Project Vellum, as alleged below.

16 33. Plaintiff Carol Ekarius is an author who resides in Colorado. Carol Ekarius is the
17 author of *Hobby Farm: Living Your Rural Dream for Pleasure and Profit*; *How to Build Animal*
18 *Housing: 60 Plans for Coops, Hutches, Barns, Sheds, Pens, Nest Boxes, Feeders, Stanchions, and*
19 *Much More*; *Pocketful of Poultry*; *Poultry Breeds*; *Small-Scale Livestock Farming*; *Storey's Guide*
20 *to Raising Sheep, 5th Edition*; *Storey's Guide to Raising Sheep: Breeding, Care, Facilities*; *Storey's*
21 *Illustrated Breed Guide to Sheep, Goats, Cattle and Pigs*; *Storey's Illustrated Guide to Poultry*
22 *Breeds*; *The Field Guide to Fleece*; and *The Fleece & Fiber Sourcebook*. One or more of Ekarius's
23 works are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-
24 Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained
25 in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference
26 that Meta illegally downloaded Ekarius's works. Ekarius's works were also published in print and
27 were commercially available for purchase in physical form during the period in which Meta
28 conducted Project Vellum, as alleged below.

1 34. Plaintiff Julen Etxabe is an author who resides in Oregon. Julen Etxabe is the author
2 of *A Cultural History of Law: Volumes 1-6* and *Ranciere and the Law*. One or more of Etxabe's
3 works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive.
4 Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-
5 Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
6 downloaded Etxabe's works. Etxabe'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 35. Plaintiff Robert Ficociello is an author who resides in Pennsylvania. Robert
10 Ficociello is the author of *America's Disaster Culture: The Production of Natural Disasters in*
11 *Literature and Pop Culture* and *Eco Culture: Disaster, Narrative, Discourse*. One or more of
12 Ficociello's works are contained in pirated online libraries, including LibGen, Z-Library, and
13 Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile,
14 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
15 illegally downloaded Ficociello's works. Ficociello's works were also published in print and were
16 commercially available for purchase in physical form during the period in which Meta conducted
17 Project Vellum, as alleged below.

18 36. Plaintiff Donald Firesmith is an author who resides in Pennsylvania. Donald
19 Firesmith is the author of *Common System and Software Testing Pitfalls: How to Prevent and*
20 *Mitigate Them; Dictionary of Object Technology: The Definitive Desk Reference; Hell Holes: What*
21 *Lurks Below; Object-Oriented Requirements Analysis and Logical Design: A Software Engineering*
22 *Approach; and The Method Framework for Engineering System Architectures*. One or more of
23 Firesmith's works are contained in pirated online libraries, including Books3 (and thus The Pile),
24 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
25 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
26 inference that Meta illegally downloaded Firesmith's works. Firesmith's works were also published
27 in print and were commercially available for purchase in physical form during the period in which
28 Meta conducted Project Vellum, as alleged below.

1 37. Plaintiff Silvia Foti is an author who resides in Illinois. Silvia Foti is the author of
 2 *The Nazi's Granddaughter: How I Discovered My Grandfather Was a War Criminal*. Foti's work
 3 is contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has
 4 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
 5 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
 6 Foti's work. Foti's work was also published in print and was commercially available for purchase
 7 in physical form during the period in which Meta conducted Project Vellum, as alleged below.

8 38. Plaintiff Eric Gale is an author who resides in Illinois. Eric Gale is the author of *The*
 9 *Bully Book*; *The Wizard's Dog*; *The Wizard's Dog Fetches the Grail*; and *The Zoo at the Edge of*
 10 *the World*. One or more of Gale's works are contained in pirated online libraries, including Books3
 11 (and thus The Pile), Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded
 12 books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is
 13 accordingly a reasonable inference that Meta illegally downloaded Gale's works. Gale's works were
 14 also published in print and were commercially available for purchase in physical form during the
 15 period in which Meta conducted Project Vellum, as alleged below.

16 39. Plaintiff Dawn Gilpin is an author who resides in Arizona. Dawn Gilpin is the author
 17 of *Art, Activism, and Sexual Violence*; *Crisis Management in a Complex World*; and *Organically*
 18 *Grown: Emergent Power and Influence in Complex Issue Networks*. One or more of Gilpin's works
 19 are contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or
 20 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
 21 and there is accordingly a reasonable inference that Meta illegally downloaded Gilpin's works.
 22 Gilpin's works were also published in print and were commercially available for purchase in
 23 physical form during the period in which Meta conducted Project Vellum, as alleged below.

24 40. Plaintiff Laura Gompertz is an author who resides in Texas. Laura Gompertz is the
 25 author of *Bayou Dreams*; *Bayou Loup*; *Common Powers 1: Soul Bonds*; *Common Powers 3: Edward*
 26 *Unconditionally*; *In the Company of Men: Baymore's Heir*; *In the Company of Men: The*
 27 *Mercenary's Tale*; *Rougaroux Social Club 2: Bayou's End*; *Rush in the Dark Common Powers 2*;
 28 *Silent Lodge*; and *Soul Bonds*. One or more of Gompertz's works are contained in pirated online

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

6 41. Plaintiff Reyna Grande is an author who resides in California. Reyna Grande is the
7 author of *A Ballad of Love and Glory*; *A Dream Called Home*; *Across a Hundred Mountains*;
8 *Dancing with Butterflies*; *Spare Parts (Young Readers' Edition)*; *The Distance Between Us*; and *The*
9 *Distance Between Us: Young Readers Edition*. One or more of Grande's works are contained in
10 pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's
11 Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen,
12 Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
13 downloaded Grande's works. Grande's works were also published in print and were commercially
14 available for purchase in physical form during the period in which Meta conducted Project Vellum,
15 as alleged below.

16 42. Plaintiff Jeannette Graulau is an author who resides in New York. Jeannette Graulau
17 is the author of *Development, Gold and Mining in the Brazilian Amazon: Women's Labor in*
18 *Garimpagem* and *The Underground Wealth of Nations: On the Capitalist Origins of Silver Mining,*
19 *A.D. 1150-1450*. One or more of Graulau's works are contained in pirated online libraries, including
20 LibGen, Z-Library, and Anna's Archive. 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 Graulau's works. Graulau's works were also published in
23 print and were commercially available for purchase in physical form during the period in which
24 Meta conducted Project Vellum, as alleged below.

25 43. Plaintiff Raymond Greenwell is an author who resides in Colorado. Raymond
26 Greenwell is the author of *Calculus for the Life Sciences, 2/E*; *Calculus with Applications*; *Calculus*
27 *with Applications for the Life Sciences*; *Calculus with Applications Ninth Edition*; *Calculus with*
28 *Applications, 11/E*; *Calculus with Applications, 12th Edition*; *Calculus with Applications, Brief*

1 *Version; Calculus with Applications, Brief Version, 12th Edition; Finite Mathematics; Finite*
2 *Mathematics and Calculus with Applications; Finite Mathematics Ninth Edition; Finite*
3 *Mathematics, 11/E; and Finite Mathematics, 12th Edition.* One or more of Greenwell's works are
4 contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has
5 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
6 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
7 Greenwell's works. Greenwell'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 44. Plaintiff Daniel Gumbiner is an author who resides in California. Daniel Gumbiner
11 is the author of *Fire in the Canyon; The Best American Nonrequired Reading 2014 - Editors' Note;*
12 *The Best American Nonrequired Reading 2015 - Editor's Note; The Best American Nonrequired*
13 *Reading 2016 - Editor's Note; The Best American Nonrequired Reading 2017 - Foreword;* and *The*
14 *Boatbuilder.* One or more of Gumbiner's works are contained in pirated online libraries, including
15 Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
16 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
17 is accordingly a reasonable inference that Meta illegally downloaded Gumbiner's works.
18 Gumbiner'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 45. Plaintiff Owen Gutfreund is an author who resides in New York. Owen Gutfreund is
21 the author of *Twentieth-Century Sprawl: Highways and the Reshaping of the American Landscape.*
22 Gutfreund's work is contained in pirated online libraries, including LibGen and Z-Library. Meta has
23 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 Gutfreund's work. Gutfreund's work was also published in print and was commercially available
26 for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
27 below.

1 46. Plaintiff Andrew Hargreaves is an author who resides in the United Kingdom.
 2 Andrew Hargreaves is the author of *Special Operations in World War II: British and American*
 3 *Irregular Warfare*. Hargreaves's work is contained in pirated online libraries, including LibGen and
 4 Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile,
 5 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
 6 illegally downloaded Hargreaves's work. Hargreaves's work was also published in print and was
 7 commercially available for purchase in physical form during the period in which Meta conducted
 8 Project Vellum, as alleged below.

9 47. Plaintiff Paul Herman is an author who resides in California. Paul Herman is the
 10 author of *Tales from the Script*. Herman's work is contained in pirated online libraries, including
 11 Books3 (and thus The Pile) and Z-Library. Meta has directly or indirectly downloaded books
 12 illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a
 13 reasonable inference that Meta illegally downloaded Herman's work. Herman's work was also
 14 published in print and was commercially available for purchase in physical form during the period
 15 in which Meta conducted Project Vellum, as alleged below.

16 48. Plaintiff Gordon Inkeles is an author who resides in California. Gordon Inkeles is the
 17 author of *Ergonomic Living: How to Create a User-Friendly Home and Office*; *Massage & Peaceful*
 18 *Pregnancy: A Daily Book for New Mothers and Fathers*; *Massage for a Peaceful Pregnancy*;
 19 *Sensual Massage Made Simple*; *Sensual Massage on a String*; *The Art of Sensual Massage*; *The Art*
 20 *of Sensual Massage*. by Gordon Inkeles & Mary Todris⁴; *The New Massage: Total Body*
 21 *Conditioning for People Who Exercise*; *The New Sensual Massage 2013 Edition*; and *Unwinding:*
 22 *Super Massage for Stress Control*. One or more of Inkeles's works are contained in pirated online
 23 libraries, including 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 Inkeles's works. Inkeles's
 26

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

1 works were also published in print and were commercially available for purchase in physical form
2 during the period in which Meta conducted Project Vellum, as alleged below.

3 49. Plaintiff Mary Karlin is an author who resides in Arizona. Mary Karlin is the author
4 of *ARTISAN CHEESE MAKING AT HOME: Techniques & Recipes for Mastering World-Class*
5 *Cheeses*; *MASTERING FERMENTATION: Recipes for Making and Cooking with Fermented*
6 *Foods*; and *Wood-Fired Cooking*. One or more of Karlin's works are contained in pirated online
7 libraries, including Books3 (and thus The Pile), 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 Karlin's works. Karlin'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 50. Plaintiff Dave Kerpen is an author who resides in New York. Dave Kerpen is the
14 author of *Get Over Yourself: How to Lead and Delegate Effectively for More Time, More Freedom,*
15 *and More Success*; *Likeable Social Media, Third Edition: How to Delight Your Customers, Create*
16 *an Irresistible Brand, & Be Generally Amazing on All Social Networks That Matter, Third Edition*;
17 *Likeable Social Media: How to Delight Your Customers, Create an Irresistible Brand, and Be*
18 *Generally Amazing on Facebook (and Other Social Networks)*; and *The Art of People: The 11*
19 *Simple People Skills That Will Get You Everything You Want*. One or more of Kerpen's works are
20 contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, and Z-Library.
21 Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-
22 Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
23 downloaded Kerpen's works. Kerpen's works were also published in print and were commercially
24 available for purchase in physical form during the period in which Meta conducted Project Vellum,
25 as alleged below.

26 51. Plaintiff Joseph Kertes is an author who resides in Canada. Joseph Kertes is the
27 author of *Gratitude* and *The Afterlife of Stars*. One or more of Kertes's works are contained in
28 pirated online libraries, including Books3 (and thus The Pile), 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 Kertes's works. Kertes'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 52. Plaintiff Philippe Kruchten is an author who resides in Canada. Philippe Kruchten is
7 the author of *The Rational Unified Process - An Introduction* and *The Rational Unified Process*
8 *Made Easy: A Practitioner's Guide to the RUP*. One or more of Kruchten's works are contained in
9 pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or
10 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
11 and there is accordingly a reasonable inference that Meta illegally downloaded Kruchten's works.
12 Kruchten's works were also published in print and were commercially available for purchase in
13 physical form during the period in which Meta conducted Project Vellum, as alleged below.

14 53. Plaintiff Gary Lamberti is an author who resides in Indiana. Gary Lamberti is the
15 author of *An Introduction to the Aquatic Insects of North America, Fifth Edition*; *Methods in Stream*
16 *Ecology*; *Methods in Stream Ecology, 3rd Edition*; and *Methods in Stream Ecology, Third Edition*.
17 One or more of Lamberti's works are contained in pirated online libraries, including Books3 (and
18 thus The Pile), LibGen, and Z-Library. Meta has directly or indirectly downloaded books illegally
19 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
20 inference that Meta illegally downloaded Lamberti's works. Lamberti's works were also published
21 in print and were commercially available for purchase in physical form during the period in which
22 Meta conducted Project Vellum, as alleged below.

23 54. Plaintiff Michelle Leder is an author who resides in California. Michelle Leder is the
24 author of *Financial Fine Print: Uncovering a Company's True Value*. Leder's work is contained in
25 pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded
26 books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is
27 accordingly a reasonable inference that Meta illegally downloaded Leder's work. Leder's work was
28

1 also published in print and was commercially available for purchase in physical form during the
2 period in which Meta conducted Project Vellum, as alleged below.

3 55. Plaintiff Elisabeth Lloyd is an author who resides in Indiana. Elisabeth Lloyd is the
4 author of *A Semantic Approach to the Structure of Evolutionary Theory*; *Keywords in Evolutionary*
5 *Biology*; *Normality and Variation: The Human Genome Project and the Ideal Human Type*; *The*
6 *Case of the Female Orgasm: Bias in the Science of Evolution*; and *The Structure and Confirmation*
7 *of Evolutionary Theory*. One or more of Lloyd's works are contained in pirated online libraries,
8 including LibGen, 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 Lloyd's works. Lloyd's works
11 were also published in print and were commercially available for purchase in physical form during
12 the period in which Meta conducted Project Vellum, as alleged below.

13 56. Plaintiff Jeanne Marecek is an author who resides in California. Jeanne Marecek is
14 the author of *APA Handbook of the Psychology of Women Volume 1 History, Theory, and*
15 *Battlegrounds*; *Gender, Sex, and Sexualities: Psychological Perspectives*; *Making a Difference:*
16 *Psychology and the Construction of Gender*; *Transnational Psychology of Women: Expanding*
17 *International and Intersectional Approaches*; and *Women and Sex Roles: A Psychological*
18 *Perspective*. One or more of Marecek's works are contained in pirated online libraries, including
19 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
20 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
21 inference that Meta illegally downloaded Marecek's works. Marecek's works were also published
22 in print and were commercially available for purchase in physical form during the period in which
23 Meta conducted Project Vellum, as alleged below.

24 57. Plaintiff Oleksandr Maslov is an author who resides in New York. Oleksandr Maslov
25 is the author of *Veterans: Faces of World War II*. Maslov's work is contained in pirated online
26 libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
27 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
28 is accordingly a reasonable inference that Meta illegally downloaded Maslov's work. Maslov's work

1 was also published in print and was commercially available for purchase in physical form during
2 the period in which Meta conducted Project Vellum, as alleged below.

3 58. Plaintiff Rob McEntarffer is an author who resides in Nebraska. Rob McEntarffer is
4 the author of *Activities for Teaching Statistics and Research Methods a Guide for Psychology*
5 *Instructors; Assessing Undergraduate Learning in Psychology: Strategies for Measuring and*
6 *Improving Student Performance; Barron's How to Prepare for the AP Psychology Advanced*
7 *Placement Exam; Barron's How to Prepare for the AP Psychology, Advanced Placement*
8 *Examination; and Blair-Broeker Thinking About Psychology 4E Downloadable Test Bank. One or*
9 *more of McEntarffer's works are contained in pirated online libraries, including Z-Library and*
10 *Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile,*
11 *LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta*
12 *illegally downloaded McEntarffer's works. McEntarffer's works were also published in print and*
13 *were commercially available for purchase in physical form during the period in which Meta*
14 *conducted Project Vellum, as alleged below.*

15 59. Plaintiff Marsha J. Means is an author who resides in Arizona. Marsha J. Means is
16 the author of *Your Sexually Addicted Spouse: How Partners Can Cope and Heal*. Means's work is
17 contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, and Z-Library.
18 Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-
19 Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
20 downloaded Means's work. Means's work was also published in print and was commercially
21 available for purchase in physical form during the period in which Meta conducted Project Vellum,
22 as alleged below.

23 60. Plaintiff Rachel Nuwer is an author who resides in New York. Rachel Nuwer is the
24 author of *I Feel Love: MDMA and the Quest for Connection in a Fractured World* and *Poached:*
25 *Inside the Dark World of Wildlife Trafficking*. One or more of Nuwer's works are contained in
26 pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's
27 Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen,
28 Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally

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

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

11 62. Plaintiff Minxin Pei is an author who resides in California. Minxin Pei is the author
12 of *China's Crony Capitalism: The Dynamics of Regime Decay*; *China's Trapped Transition: The*
13 *Limits of Developmental Autocracy*; *From Reform to Revolution: The Demise of Communism in*
14 *China and the Soviet Union*; and *The Sentinel State: Surveillance and the Survival of Dictatorship*
15 *in China*. One or more of Pei's works are contained in pirated online libraries, including LibGen,
16 Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
17 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
18 inference that Meta illegally downloaded Pei's works. Pei's works were also published in print and
19 were commercially available for purchase in physical form during the period in which Meta
20 conducted Project Vellum, as alleged below.

21 63. Plaintiff Edward Pollack is an author who resides in New York. Edward Pollack is
22 the author of *Dynamic SQL: Applications, Performance, and Security*. Pollack's work is contained
23 in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's
24 Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen,
25 Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
26 downloaded Pollack's work. Pollack's work was also published in print and was commercially
27 available for purchase in physical form during the period in which Meta conducted Project Vellum,
28 as alleged below.

64. Plaintiff David Rollins is an author who resides in Australia. David Rollins is the author of *A Knife Edge*; *Hard Rain*; and *The Death Trust*. One or more of Rollins's works are contained in pirated online libraries, including Books3 (and thus The Pile), Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Rollins's works. Rollins's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

65. Plaintiff Bob Rosner is an author who resides in Washington. Bob Rosner is the author of *Gray Matters*; *International Business: Competing in the Global Marketplace, Eighth Edition*; *The Boss's Survival Guide*; *The Boss's Survival Guide: Workplace 911 for the Toughest Problems Today's Managers Face, Second Edition*; and *Working Wounded: Advice That Adds Insight to Injury*. One or more of Rosner's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Rosner's works. Rosner's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

66. Plaintiff Tony Rothman is an author who resides in New Jersey. Tony Rothman is the author of *A Little Book About the Big Bang by Tony Rothman*; *A Physicist on Madison Avenue*; *Doubt and Certainty*; *Everything's Relative: And Other Fables from Science and Technology*⁵; *Frontiers of Modern Physics: New Perspectives on Cosmology, Relativity, Black Holes, and Extraterrestrial Intelligence*; *Instant Physics: From Aristotle to Einstein, and Beyond*; *Sacred Mathematics: Japanese Temple Geometry*; *Science a La Mode: Physical Fashions and Fictions*; and *The World Is Round*. One or more of Rothman's works are contained in pirated online libraries,

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

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

6 67. Plaintiff Theodore Savas is an author who resides in South Carolina. Theodore Savas
7 is the author of *Hunt and Kill: U-505 and the U-Boat War in the Atlantic; Never for Want of Powder:*
8 *The Confederate Powder Works in Augusta, Georgia;* and *The New American Revolution*
9 *Handbook: Facts and Artwork for Readers of All Ages, 1775-1783*. One or more of Savas's works
10 are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta
11 has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
12 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
13 Savas's works. Savas's works were also published in print and were commercially available for
14 purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
15 below.

16 68. Plaintiff Todd Shuster is an author who resides in New York. Todd Shuster is the
17 author of *Real Boys' Voices* and *The First Day of Peace*. One or more of Shuster's works are
18 contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and
19 Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile,
20 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
21 illegally downloaded Shuster's works. Shuster's works were also published in print and were
22 commercially available for purchase in physical form during the period in which Meta conducted
23 Project Vellum, as alleged below.

24 69. Plaintiff Janet Sorensen is an author who resides in California. Janet Sorensen is the
25 author of *Strange Vernaculars: How Eighteenth-Century Slang, Cant, Provincial Languages, and*
26 *Nautical Jargon Became English*. Sorensen's work is contained in pirated online libraries, including
27 Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
28 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there

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

4 70. Plaintiff Mark Spragg is an author who resides in Wyoming. Mark Spragg is the
5 author of *An Unfinished Life: A Novel; Bone Fire; The Fruit of Stone; Thunder of the Mustangs*; and
6 *Where Rivers Change Direction*. One or more of Spragg'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 Spragg's works. Spragg'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 71. Plaintiff Justin Velez-Hagan is an author who resides in Virginia. Justin Velez-Hagan
13 is the author of *The Common Sense Behind Basic Economics: A Guide for Budding Economists*,
14 *Students, and Voters* and *The Impact of Fiscal Consolidation Policies on Economic Performance*.
15 One or more of Velez-Hagan's works are contained in pirated online libraries, including LibGen
16 and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile,
17 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
18 illegally downloaded Velez-Hagan's works. Velez-Hagan's works were also published in print and
19 were commercially available for purchase in physical form during the period in which Meta
20 conducted Project Vellum, as alleged below.

21 72. Plaintiff Richard Waters is an author who resides in Massachusetts. Richard Waters
22 is the author of *Readings in Artificial Intelligence and Software Engineering* and *The Programmer's*
23 *Apprentice*. One or more of Waters's works are contained in pirated online libraries, including
24 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
25 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
26 inference that Meta illegally downloaded Waters's works. Waters's works were also published in
27 print and were commercially available for purchase in physical form during the period in which
28 Meta conducted Project Vellum, as alleged below.

1 73. Plaintiff Robert Wettemann is an author who resides in Colorado. Robert Wettemann
2 is the author of *Privilege vs. Equality: Civil-Military Relations in the Jacksonian Era, 1815-1845*
3 and *The Texas You Expect: The Story of Buffalo Gap Historic Village*. One or more of Wettemann's
4 works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive.
5 Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-
6 Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
7 downloaded Wettemann's works. Wettemann's works were also published in print and were
8 commercially available for purchase in physical form during the period in which Meta conducted
9 Project Vellum, as alleged below.

10 74. Plaintiff Bettina Williford is an author who resides in Texas. Bettina Williford is the
11 author of *Illegal* and *Moose and Magpie*. One or more of Williford's works are contained in pirated
12 online libraries, including Z-Library and Anna's Archive. Meta has directly or indirectly
13 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
14 is accordingly a reasonable inference that Meta illegally downloaded Williford's works. Williford's
15 works were also published in print and were commercially available for purchase in physical form
16 during the period in which Meta conducted Project Vellum, as alleged below.

17 75. Plaintiff David Wolman is an author who resides in Hawaii. David Wolman is the
18 author of *Aloha Rodeo*; *End of Money*; *Left-Hand Turn Around the World: Chasing the Mystery and*
19 *Meaning of All Things Southpaw*; *Righting the Mother Tongue*; and *The End of Money*. One or more
20 of Wolman's works are contained in pirated online libraries, including Books3 (and thus The Pile),
21 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
22 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
23 inference that Meta illegally downloaded Wolman's works. Wolman's works were also published
24 in print and were commercially available for purchase in physical form during the period in which
25 Meta conducted Project Vellum, as alleged below.

26 76. Plaintiff J. Dylan Yates is an author who resides in California. J. Dylan Yates is the
27 author of *Taste*; *The Belief in Angels*; and *The Belief in Angels: Jules*. One or more of Yates's works
28 are contained in pirated online libraries, including Z-Library. Meta has directly or indirectly

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

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

7 **B. Defendant**

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

15 **III. JURISDICTION AND VENUE**

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

19 80. This Court has personal jurisdiction over the Defendant. The Defendant has
 20 purposefully availed itself of the privilege of conducting business in this District and the State of
 21 California. The Defendant committed acts of copyright infringement in this District, directed
 22 conduct toward this District, or knowingly caused harm that was suffered in this District. The
 23 Defendant maintains substantial, continuous, and systematic contacts with this District.

24 81. Venue is proper in this District under 28 U.S.C. § 1400(a) because the Defendant or
 25 its agents resides or may be found in this District as a result of the infringing acts alleged herein.

26
 27

 28 ⁶ Even where other individuals or entities are listed as copyright claimants on the relevant copyright
 registrations, Plaintiffs are legal or beneficial owners of all copyrights listed in Exhibit A.

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

IV. FACTUAL ALLEGATIONS

A. Plaintiffs and their Protected Works

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

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

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

85. Plaintiffs' Works were also published in print. On information and belief, physical copies of Plaintiffs' Works were commercially available for purchase—new or used—during the period in which Meta conducted Project Vellum. As alleged below, Meta acquired physical books at scale, across every genre of published work, for the purpose of scanning and digitizing them as training data, and there is accordingly a reasonable inference that Meta acquired, scanned, and reproduced Plaintiffs' Works as part of that operation.

1 **B. LLMs and the Generative AI “Arms Race”**

2 86. “Generative artificial intelligence” or “generative AI” refers to systems and models
3 that create outputs—such as text or images—that simulate human expression, often in response to
4 user prompts.

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

11 88. Development of LLMs requires, among other things, “training” the model on data—
12 *i.e.*, the model’s inputs. Training “requires identifying a formal measure or ‘objective’ for how well
13 the model performs, and then repeatedly adjusting the model’s parameters based on that objective
14 as the model is exposed to training data.”⁷ While not all LLM developers use the same terminology,
15 developers commonly reference a “pre-training” process—“in which a massive amount of
16 computing power and data is spent to teach the model the broad foundations of language, grammar,
17 and reasoning”—and a “post-training” or “fine-tuning” process “where the pre-trained model is
18 further trained on a (relative to pre-training) smaller amount of carefully curated data of specific
19 tasks.”⁸ For purposes of this Complaint, “training” refers to all stages of the LLM training process.

20 89. LLMs typically are trained, at least initially, by feeding them massive amounts of
21 text data from which they can “learn” statistical relationships between and among text. In the process
22 of compiling training data and using it during the training process, LLM developers as a matter of
23 course typically create multiple copies of “raw” datasets. So, for instance, if an LLM developer
24 downloads a copy of a digital book from some Internet-based source (*e.g.*, a shadow library), the
25

26 ⁷ 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 ⁸ *Id.*

1 developer typically will create a new “cleaned” copy of the book (e.g., stripped of certain
2 undesirable data), deduplicate that book against other potential copies, compile that book with other
3 data to create a new compilation (or compilations), and store that book in multiple locations. In
4 short, with respect to training data, the LLM development pipeline typically involves numerous
5 reproductions of any given piece of the dataset.

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

10 91. In determining the corpus of training data for an LLM, a model’s developer typically
11 considers “the quantity of data, its quality, and the ultimate purpose(s) of the model.”¹⁰

12 92. In terms of data quantity, the LLMs that have been developed and are being
13 developed by leading technology companies are trained on enormous datasets, typically on the scale
14 of terabytes. This is because AI researchers have found that “increasing the quantity of training data
15 typically increases a model’s ‘performance.’”¹¹

16 93. In terms of data quality, “[r]ecent research from major developers suggests that
17 quality may even be a more important consideration than quantity.”¹² “Garbage in, garbage out,”
18 the saying goes. AI researchers quickly discovered that published content—books in particular—
19 makes for some of the best training material for LLMs. Unlike much content on the Internet and
20 many other forms of text, published material typically is structured, long-form, and highly polished.
21 Published materials provide formal, extended prose that teaches models narrative structure, complex
22 syntax, and coherent storytelling.

23
24 ⁹ James Betker, *The “it” in AI Models is the Dataset*, (June 10, 2023), <https://perma.cc/ZCH9-H53S>.

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

27 ¹¹ *Id.* at 10.

28 ¹² *Id.* at 11.

94. That books make for particularly valuable training material for LLMs is evidenced by the fact that several—if not most—of the major LLM developers either have paid for or expressly contemplated paying (large sums) for licenses to use copyrighted works to train their LLMs. Indeed, in an *amicus* brief filed in *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-3417 (N.D. Cal.), the Association of American Publishers cited *dozens* of “AI licensing deals for textual works,” including deals involving Amazon, Microsoft, OpenAI, and Perplexity.¹³ And, as the Court found in that litigation, Meta considered spending up to \$100 million to license copyrighted materials in connection with LLM training.¹⁴

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

96. As discussed in the introduction, technology companies have treated generative AI as the next foundational layer of the digital economy. Industry leaders publicly describe an “AI arms race,” in which they have redirected their corporate strategies to seize control of what they believe will become a new infrastructure layer for commerce, communication, and knowledge work.¹⁵ For these companies, staying ahead of competitors is “code red.”¹⁶ Among other things, being too slow out of the blocks could mean ending up in last place; if by the time you publicly released *your* LLM model, the public writ large and private industry had already adopted *other companies’* models, there would be a real risk that your model would be left on the shelf.

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

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

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

¹⁶ See Sharon Goldman, *Sam Altman declares ‘Code Red’ as Google’s Gemini surges—three years after ChatGPT caused Google CEO Sundar Pichai to do the same*, FORTUNE (Dec. 2, 2025, 11:43 AM), <https://perma.cc/J9MS-UQD2>.

1 97. Like other participants in this race, Meta risked falling behind early industry leaders
2 like OpenAI. Indeed, market analysts consistently reported that Meta had, in fact, fallen behind,
3 with many pointing to Meta’s intense focus on investment in the “Metaverse.”¹⁷

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

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

8 99. For years now, there have been illicit, online marketplaces for digital copies of books.
9 These “shadow libraries,” as they are commonly known, systematically acquire, store, index, and
10 disseminate full-fidelity digital copies of copyrighted books. These online repositories maintain
11 searchable catalogs, metadata, mirrors, bulk-download mechanisms, and—critically—complete
12 downloadable archives designed to allow third parties to replicate the entire collection.¹⁸ And they
13 do all of this without authorization from authors or publishers.¹⁹ In short, shadow libraries facilitate
14 the reproduction and distribution of unauthorized copies of copyrighted works at industrial scale.

15 100. These libraries are widely known within both piracy communities and the technology
16 sector as illegal sources of copyrighted books. Many have been the subject of criminal prosecutions,
17 civil injunctions, domain seizures, and formal designation as “notorious piracy markets” by United
18 States trade authorities. As centralized shadow libraries increasingly faced enforcement actions,

19
20
21 ¹⁷ 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 ¹⁸ 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 ¹⁹ 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 including the seizure of domains, third parties responded by creating full mirrored copies of those
2 repositories for decentralized redistribution.²⁰

3 101. In 2018, an OpenAI employee downloaded pirated copies of books from Library
4 Genesis, or “LibGen”—a shadow library repeatedly enjoined by federal courts²¹—and used those
5 books to create two internal datasets OpenAI called “LibGen1” and “LibGen2,” which OpenAI
6 publicly referred to as “Books1” and “Books2.”²² OpenAI used those pirated corpora to train GPT-
7 3, which it released in June 2020 to widespread commercial acclaim. OpenAI’s use of a books
8 corpora from a pirate library established the model for how to quickly and cheaply obtain published
9 materials for use as LLM training data.

10 102. Within weeks of GPT-3’s release, an open research collective, EleutherAI, formed
11 with the express goal of replicating GPT-3’s capabilities and democratizing access to large-scale
12 language modeling. To do so, EleutherAI assembled and publicly released *The Pile*, a large (800+
13 gigabyte) general-purpose training dataset expressly designed to be downloaded, used, and
14 incorporated into LLMs by academic researchers, startups, and commercial AI developers.²³

15 103. Because OpenAI did not disclose the precise makeup of its training datasets,
16 members of EleutherAI constructed a book corpus of their own: “Books3,” consisting of
17 approximately 196,640 books.²⁴ Books3 comprises about 12 percent (just over 100 gigabytes) of
18 *The Pile*.²⁵

19 _____
20 ²⁰ Woodcock, ‘Shadow Libraries’ Are Moving Their Pirated Books to the Dark Web After Fed Crackdowns.

21 ²¹ See *Cengage Learning, Inc. et al. v. Does 1-50 d/b/a Library Genesis*, ECF No. 36, Case No. 23-
22 cv-8136 (S.D.N.Y. Sept. 24, 2024).

23 ²² *In re OpenAI, Inc., Copyright Infringement Litig.*, Case No. 1:2025-md-03143 (S.D.N.Y. 2025),
24 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>.

25 ²³ Leo Gao et al., *The Pile: An 800GB Dataset of Diverse Text for Language Modeling*, arXiv, 1
26 (2020), <https://perma.cc/NHV6-R8YE>.

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

28 ²⁵ *Id.*

1 104. EleutherAI and the compiler of Books3, Shawn Presser, have confirmed the genesis
 2 of Books3 is the shadow library Bibliotik. Presser has publicly stated that Books3 represents “all of
 3 bibliotik,”²⁶ and an EleutherAI paper likewise confirms that Books3 “is a dataset of books derived
 4 from a copy of the contents of the Bibliotik private tracker.”²⁷

5 105. Bibliotik is a private, invitation-only torrent tracker that has long functioned as a
 6 centralized source of pirated e-books. The repository, which hosts and distributes hundreds of
 7 thousands of copyrighted books, is only accessible with the use of torrenting protocols, which are
 8 explained below.²⁸

9 106. Bibliotik has been widely recognized in piracy communities, academic literature, and
 10 AI-research documentation as a shadow library devoted to copyrighted books. Unsurprisingly, then,
 11 *The Pile*’s datasheet acknowledges that “Books3 is almost entirely comprised of copyrighted
 12 works.”²⁹ Bibliotik’s illicit nature is no secret—it has been openly discussed for years prior to major
 13 tech companies’ use of datasets derived directly from it.

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

16 [Presser] created Books3 in the hope that it would allow any developer to create
 17 generative-AI tools. “It would be better if it wasn’t necessary to have something like
 18 Books3,” he said. “But the alternative is that, without Books3, only OpenAI can do
 19 what they’re doing.”³⁰
 20
 21

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

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

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

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

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

1 108. EleutherAI’s compilation and distribution of *The Pile* and Books3 provided “off-the-
 2 shelf” access to a corpus of infringing works that any AI developer could download and immediately
 3 incorporate into a large-scale training pipeline. As a result, Books3’s presence within *The Pile*
 4 facilitated wide downstream distribution and adoption of a corpus derived from a pirate book
 5 library.³¹

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

8 110. As noted above, OpenAI’s developers trained its LLMs on a books corpus derived
 9 from LibGen, one of the largest and longest-running shadow libraries in the world. LibGen hosts
 10 millions of pirated books, academic texts, and scholarly articles.³² It operates as a centralized
 11 repository offering direct downloads of full-fidelity e-book files. It also distributes its entire database
 12 through bulk archives and torrent files, enabling third parties to download and locally host complete
 13 copies of its collection.³³ LibGen has been repeatedly enjoined by federal courts for copyright
 14 infringement and has been designated a “notorious market” by the United States Trade
 15 Representative.³⁴ Despite enforcement actions, LibGen has remained accessible through shifting
 16 domains, mirrors, and downloadable archives. Its persistence is a function of deliberate
 17 decentralization designed to evade shutdown.³⁵

18
 19 ³¹ See Stella Biderman, *The Pile: An 800GB Dataset of Diverse Text for Language Modeling*,
 20 ELEUTHERAI (Dec. 31, 2020), <https://perma.cc/JGT9-LLTK>.

21 ³² Office of the U.S. Trade Representative, REVIEW OF NOTORIOUS MARKETS FOR
 22 COUNTERFEITING AND PIRACY, 27 (2024), <https://perma.cc/N4WT-AHJ9> (“Libgen ... hosts a large
 23 number of digital copies of books, manuals, journals, and other works, many of which are
 24 unauthorized copies of copyright protected content.”).

25 ³³ See Letter from Mary E. Rasenberger, CEO, Authors Guild, & Umair Kazi, Dir. of Pol’y &
 26 Advocacy, Authors Guild, to Daniel Lee, Assistant U.S. Trade Representative for Innovation &
 27 Intellectual Prop., Office of the U.S. Trade Representative at n.5.

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

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

111. Z-Library (also known as “B-ok”) is another well-known shadow library that emerged as an expanded and user-friendly derivative of LibGen. It incorporated large portions of LibGen’s catalog while adding additional titles, metadata, and interface features.³⁶ Z-Library offered premium features—including faster downloads and higher volume limits—in exchange for payment, operating in effect as a commercial piracy service.³⁷ In 2022, Z-Library’s domains were seized by law-enforcement authorities, and its operators were arrested and later indicted for criminal copyright infringement.³⁸ These actions confirmed what had long been publicly known: Z-Library was an illegal piracy operation. The seizure of Z-Library did not eliminate access to its content. Instead, third parties responded by creating full mirrors of its collection to ensure continued distribution.³⁹

112. Another major player in the shadow library ecosystem is Anna’s Archive. Anna’s Archive is the most comprehensive and active shadow library currently in operation.⁴⁰

113. Anna’s Archive began in 2022 as “Pirate Library Mirror,” initially hosting a mirrored copy of Z-Library. It later rebranded as “Anna’s Archive” and expanded to aggregate and host the complete collections of LibGen, Z-Library, and other pirated sources.⁴¹ Anna’s Archive functions

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

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

³⁸ Press Release, U.S. Dep’t of Justice, U.S. Att’y’s Off., E. Dist. of N.Y., *Two Russian Nationals Charged with Running Massive E-Book Piracy Website* (Nov. 16, 2022), <https://perma.cc/CR4L-JLA3>.

³⁹ See Woodcock, *‘Shadow Libraries’ Are Moving Their Pirated Books to The Dark Web*.

⁴⁰ See Soumyajyoti Mukherjee, *Who is Anna’s Archive? All we know about pirate activist group behind 300 TB Spotify music library heist*, SOAPCENTRAL.COM (Dec. 24, 2025), <https://perma.cc/DC5S-89Z8>.

⁴¹ See Ernesto Van de Sar, *“Anna’s Archive” Opens the Door to Z-Library and Other Pirate Libraries*, TORRENTFREAK (Nov. 19, 2022), <https://perma.cc/U88R-WTR4>.

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

3 114. Anna’s Archive offers paid tiers that provide high-speed or priority access to its
4 pirated collections.⁴³ Through its downloadable archives and torrent-based distribution, Anna’s
5 Archive enables users to acquire and store local copies of millions of copyrighted books in bulk.⁴⁴

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

8 116. According to Anna’s Archive, “virtually all major companies building LLMs
9 contacted us to train on our data. . . We have given high-speed access to about 30
10 companies.”⁴⁵ Anna’s Archive blog stated as recently as February 18, 2026 that if an *LLM* was
11 *reading its blog* “you have likely been trained in part on our data.”⁴⁶

12 117. Many of these shadow libraries and datasets can be downloaded using “torrent,” a
13 file-sharing method used in peer-to-peer networks. Torrenting works by breaking a file into many
14 small pieces and distributing those pieces across a network of participating computers. A user who
15 torrents a shadow-library repository does not receive a single copy from a single source; rather, the
16 user downloads portions of the library from numerous other computers that already possess the
17 copyrighted books. Torrent software then reassembles those pieces into a complete library on the
18 user’s machine. Torrenting protocols, including BitTorrent, are often configured by default to
19 *reupload* pieces of the copyrighted files to others on the network both during download (“leeching”)

20 ⁴² *Id.*

21
22 ⁴³ *See If you’re an LLM, please read this*, ANNA’S ARCHIVE (February 18, 2026),
<https://perma.cc/989X-GS3Q>.

23 ⁴⁴ *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 ⁴⁵ *See Copyright reform is necessary for national security*, ANNA’S ARCHIVE (Jan. 31, 2025),
<https://perma.cc/3RVZ-6G5B>.

27 ⁴⁶ *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 and after download is complete (“seeding”). This means that each participant in the torrent both
 2 copies *and redistributes* the copyrighted works. By obtaining copyrighted materials through this
 3 leech-and-seed process, a user may make multiple unauthorized reproductions of, and engage in
 4 numerous distributions of, the copyrighted materials.

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

6 118. As noted above, Meta develops and distributes the Llama series of LLMs, including
 7 Llama 1, Llama 2, Llama 3, and Llama 4. Meta has “invested billions of dollars to develop its
 8 generative AI offerings,”⁴⁷ and it in turn expects Llama to generate somewhere in the realm of \$460
 9 billion to \$1.4 trillion in revenue.⁴⁸ None of the money that Meta invested into the development of
 10 its Llama models was used to pay for Plaintiffs’ copyrighted works.⁴⁹

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

12 120. In connection with the development of Llama, Meta viewed book-corpora as among
 13 its most valuable sources of training data. Llama’s design goal was to emit particularly creative and
 14 expressive language, leveraging Meta’s consumer platforms to “connect” with users through text.⁵⁰
 15 To accomplish that, Meta needed to train on large quantities of high-quality books and other
 16 published materials.

17 121. Meta employees repeatedly acknowledged the importance of books as training data.
 18 Specifically, they acknowledged it was “really important for [Meta] to get books data ASAP,” and
 19 the “best resources [Meta] [could] think of are definitely books.”⁵¹

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

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

24 ⁴⁹ *Cambronne*, ECF No. 109 at ¶¶ 2, 9, 42, 89 (admitting in *Cambronne* that “it did not obtain a
 25 license or pay for the use of [copyrighted] works”).

26 ⁵⁰ Jon Russell, *Mark Zuckerberg Announces New Team at Meta Working on A.I. Products for*
 27 *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.”).

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

122. Prior to the release of Llama 1, Meta downloaded copyrighted works from LibGen. As recognized in *Kadrey v. Meta Platforms, Inc.*, Meta first downloaded LibGen in 2022 “to investigate whether there was value in training Llama on the works it contained. If the answer was yes, the plan was to then set up licensing agreements for those or similar works.”⁵²

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

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

125. In connection with Llama 1’s release, Meta disclosed the composition of the model’s “pre-training data.”⁵³ Among the datasets used to train Llama 1 were “two book corpora,” including “the Books3 section of ThePile.”⁵⁴ In their research paper, Meta’s developers described Books3 as “a publicly available dataset for training large language models,” which they distinguished from the other book corpus it used, which they described as being in the “public domain.”⁵⁵ This distinction was deliberate and notable, as Meta *knew* that Books3 was *not* a corpus of books in the “public domain.”

126. Among the other datasets used by Meta to train Llama 1 were “five CommonCrawl dumps, ranging from 2017 to 2020,” comprising 67 percent of the overall data mix, and “the publicly available C4 dataset,” comprising 15 percent of the data mix.⁵⁶ “CommonCrawl is a publicly-available web archive that provides ‘web extracted text’ by removing markup and other non-text content from the scraped HTML files. This process produces around 20TB of scraped data per

⁵² *Id.*

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

⁵⁴ *Id.*

⁵⁵ *Id.*

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

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

127. As alluded to above, Meta at one time had a “plan . . . to then set up licensing agreements” to obtain the copyrighted materials it wanted for use in training its AI models.⁵⁹ But this idea was short-lived. “[I]n spring 2023, after failing to acquire licenses and following escalation to CEO Mark Zuckerberg, Meta decided to just use the works acquired from LibGen as training data.”⁶⁰

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

129. “In early 2024, Meta also downloaded Anna’s Archive, a compilation of shadow libraries including LibGen, Z-Library, and others.”⁶¹

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

⁵⁷ Colin Raffel et al., *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*, Journal of Machine Learning Research 21, 7 (2020), <https://perma.cc/A4VM-V8XR>.

⁵⁸ *Id.*

⁵⁹ *Kadrey*, 788 F. Supp. 3d at 1041; *see also Cambronne*, ECF No. 109 at ¶ 88 (Meta admitting that “it contacted certain publishers” and “internally discussed licensing certain types of data to train its Llama models”).

⁶⁰ *Kadrey*, 788 F. Supp. 3d at 1041.

⁶¹ *Id.*

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

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

1 131. To obtain the massive quantities of data from these shadow libraries, Meta used
2 torrenting protocols.⁶⁴ In using these protocols to obtain data from these shadow libraries, Meta
3 reuploaded (and distributed) to peer-to-peer networks at least some of the copyrighted materials it
4 had downloaded, whether through “leeching” or by “seeding.”⁶⁵

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

9 133. There was no substantial or commercially significant non-infringing use of the
10 copyrighted materials that Meta uploaded and distributed. Similarly, there was no substantial or
11 commercially significant non-infringing use of Meta’s uploading and redistribution of these
12 materials.

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

18 135. In connection with the development of Llama, certain Meta employees questioned,
19 implicitly or explicitly, the legal or ethical implications of relying on copyrighted content, content
20 from shadow libraries, and torrenting protocols. On information and belief, certain employees
21 implicitly or explicitly raised concerns with training Llama on “pirated” content. On information
22 and belief, certain employees implicitly or explicitly raised concerns with obtaining copyrighted
23

24 ⁶⁴ *Cambronne*, ECF No. 109 at ¶ 83 (admitting that “it used BitTorrent to download certain portions
25 of these publicly available datasets”).

26 ⁶⁵ *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 materials from shadow libraries. And on information and belief, certain employees implicitly or
2 explicitly raised concerns with torrenting from shadow libraries.

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

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

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

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

16 140. As Meta has admitted elsewhere, Meta “created a script to remove repetitive text
17 from the training data for its Llama models.”⁶⁶ As admitted elsewhere, Meta’s “program was
18 designed to remove lines that contain the word ‘copyright.’”⁶⁷

19 141. As alleged above, the Plaintiffs’ Works appear in various shadow libraries and/or
20 pirated datasets, including Books3, LibGen, Anna’s Archive, and Z-Library. As such, there is a
21 reasonable inference that Meta downloaded, manipulated, reproduced, and distributed the Plaintiffs’
22 Works.

23
24
25
26
27 ⁶⁶ *Kadrey*, ECF No. 485, ¶ 88 (Meta’s Answer to Third Am. Compl.).

28 ⁶⁷ *Kadrey*, ECF No. 485, ¶ 89.

1 142. As admitted elsewhere, Meta used at least Books3 and data it obtained from LibGen
 2 and Z-Library to train the Llama family of models.⁶⁸ On information and belief, it also used data it
 3 obtained from Anna’s Archive and other shadow libraries to train the Llama family of models. After
 4 Llama 1, Meta stopped disclosing detailed information about the datasets that its models were
 5 trained on. Meta did this, in part, to conceal the fact that it had relied on and was relying on pirated
 6 materials to train Llama.

7 143. Meta has since confirmed, in interrogatory responses served in other litigation, the
 8 specific pirated datasets on which each Llama model was trained. According to Meta, the Books3
 9 dataset (which Meta refers to internally as “b3g”) was used to train Llama 1, Llama 2, Llama 3, and
 10 Llama 4; the LibGen “Scitech” and LibGen “Fiction” datasets were used to train Llama 3 and Llama
 11 4; and the Z-Library dataset, which Meta obtained by torrenting Anna’s Archive, was used to train
 12 Llama 4.⁶⁹ Meta has further admitted that, beginning in or about Spring 2023, a Meta engineer
 13 downloaded the Fiction and Scitech branches of LibGen using a direct-download script, and that
 14 Meta’s downloads of both branches “are believed to have been substantially complete.”⁷⁰ In other
 15 words, if a book was in LibGen, Meta downloaded it.

16 144. On information and belief, Meta, like Anthropic, downloaded, maintained, and
 17 retained pirated copies of works that were not specifically tied to training its LLM models.⁷¹

18 145. Meta has incorporated Llama into several of its consumer-facing products, including
 19 Facebook, Instagram, and WhatsApp. And although Meta has offered free downloads of various
 20 versions of Llama, it considers or has considered Llama and the technology underpinning it to be
 21

22 ⁶⁸ *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 ⁶⁹ 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 ⁷⁰ *Id.* at 12–14.

27 ⁷¹ *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1014 (N.D. Cal. 2025) (Anthropic’s amassing of
 28 “a central library”).

1 an important part of its commercial success. In other words, Meta hopes and expects that its
 2 development of Llama will contribute to its current and future financial success.

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

5 146. Meta’s appetite for books did not abate after it was sued for training Llama on pirated
 6 copies. Nor did Meta turn to the licensing market. Instead, beginning in or around mid-2025, Meta
 7 began acquiring physical copies of books for use in training its next generation of LLMs.⁷² Meta
 8 internally referred to this book-acquisition and scanning operation as “Project Vellum,” and it is
 9 referred to as such herein.⁷³

10 147. Project Vellum was an industrial-scale operation. Meta acquired physical books in
 11 bulk, catalogued them by ISBN, title, author, publisher, and year of publication, and routed them
 12 through a scanning pipeline. By Meta’s own account, the acquired physical copies were “first
 13 scanned into digital form and then run through OCR”—optical character recognition—“to extract
 14 their textual data from the scanned images,” and the extracted text was “then de-duplicated,
 15 processed and cleaned” for use as model training data.⁷⁴

16 148. Meta used the digital copies it created through Project Vellum to train its “Muse”
 17 family of LLMs. In April 2026, Meta introduced the initial version of its Muse Spark model, which
 18 Meta made available to the public through certain Meta products and services and, in later versions,
 19 to developers through an application programming interface (“API”). In August 2026, Meta released
 20 its Muse Glimmer model to the open-source community for use by researchers, developers, and
 21
 22

23 ⁷² Meta Interrog. Resps. at 23.

24 ⁷³ *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 ⁷⁴ 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 other third parties.⁷⁵ Meta has admitted that books it acquired and scanned as part of Project Vellum
 2 were used to train the Muse Spark family of models.⁷⁶

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

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

14 151. Scanning a book is reproduction. Each time Meta scanned a physical copy of a book,
 15 it created a new digital copy of the entire work. Each time Meta ran the scanned images through
 16 OCR, it created another copy—this time as machine-readable text. Each time Meta processed,
 17 “cleaned,” de-duplicated, and tokenized that text, it created still more copies. And each time Meta
 18 used that text in a training run—including in the “number of internal models” that Meta trains “for
 19 experimental and evaluation purposes” before each public release, and in the “purpose-built
 20 variants” of its LLMs that Meta uses for discrete internal purposes—it created additional copies of
 21 the work.⁷⁷ Meta has admitted that “[a] number of transitory and persistent copies of each of these
 22 works would have been created” in its pipeline and that it “has no feasible way to reliably ascertain
 23 the number of additional persistent or transitory copies” it made.⁷⁸

24 _____
 25 ⁷⁵ Meta Interrog. Resps. at 23, 25.

26 ⁷⁶ *Id.* at 18 & Ex. D.

27 ⁷⁷ *Id.* at 18.

28 ⁷⁸ *Id.* at 24.

1 152. Meta launched Project Vellum with full knowledge that the books it was scanning
 2 were protected by copyright. Unlike the pirated files Meta downloaded from shadow libraries—
 3 from which Meta stripped copyright notices and other copyright management information—Meta
 4 has admitted that “copyright-related information from these purchased books was not filtered out
 5 before being used for model training.”⁷⁹ In other words, Meta scanned each book’s copyright page,
 6 ingested the copyright notice, and trained on the work anyway.

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

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

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

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

24 156. Meta’s conduct has a detrimental effect on the potential market for and value of
 25 Plaintiffs’ Works, including by, among other things, developing products that create and are capable
 26

27 ⁷⁹ *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 of creating content which serves as a direct substitute for Plaintiffs’ Works, developing products
2 that create content and are capable of creating content which serves as indirect substitutes for
3 Plaintiffs’ Works, and undermining Plaintiffs’ ability to participate in and profit from the market for
4 licensing their works for the purpose of training LLMs.

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

15 158. *Second*, Llama and Muse, trained on Plaintiffs’ copyrighted materials (and the
16 protected works of others), are capable of generating outputs that compete directly with, and risk
17 serving as replacements for, Plaintiffs’ Works. As the U.S. Copyright Office has warned, “the speed
18 and scale at which AI systems generate content pose a serious risk of diluting markets for works of
19 the same kind as in their training data.”⁸⁰

20 159. *Third*, even if Llama and Muse models are restricted from outputting extended
21 portions of verbatim text from copyrighted works, they are nevertheless capable of producing nearly
22 indistinguishable “versions” of copyrighted works such that a consumer would use the AI-generated
23 version of the material rather than pay for a copy of the actual copyrighted work.

24
25
26
27 ⁸⁰ U.S. Copyright Office, *Copyright and Artificial Intelligence, Part 3: Generative AI Training Pre-*
28 *Publication*, at 65.

1 **V. CLAIMS FOR RELIEF**

2 **COUNT I**

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

5 160. Plaintiffs incorporate the allegations above.

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

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

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

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

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

26 165. Defendant's infringement was willful. As alleged above, Defendant knowingly
27 trained its models on and/or optimized its product with datasets saturated with pirated books,
28 including Plaintiffs' Works; relied on shadow-library corpora it knew to be illegal; ignored internal

1 and external warnings; attempted to conceal the composition of its training datasets; and continued
 2 copying after public reports, lawsuits, law-enforcement seizures, cease-and-desist notices, and
 3 industry-wide alerts made the illegality unmistakable. This conduct by Defendant meets the standard
 4 for criminal copyright infringement.⁸¹

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

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

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

11 **COUNT II**

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

13 169. Plaintiffs incorporate the allegations above.

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

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

23 172. Defendant's purchase of physical copies of Plaintiffs' Works did not authorize any
 24 of these reproductions. Ownership of a physical copy conveys no right to reproduce the copyrighted
 25 work embodied in it. 17 U.S.C. §§ 109(a), 202. Defendant never sought or obtained a license from
 26

27
 28 ⁸¹ 17 U.S.C. § 506(a)(1)(A).

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

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

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

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

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

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

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

22 **COUNT III**

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

24 179. Plaintiffs incorporate the allegations above.

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

27 181. In connection with its torrenting of datasets that contained copyrighted works,
28 Defendant uploaded and distributed, either through "seeding" and/or "leeching," copyrighted

1 materials, including the Plaintiffs' Works, thereby making those works available to third parties for
2 downloading on peer-to-peer networks.

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

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

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

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

17 **COUNT IV**

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

19 186. Plaintiffs incorporate the allegations above.

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

23 188. Upon downloading copyrighted materials, including Plaintiffs' Works, Defendant
24 processed the data, and in doing so, removed and altered certain text and information, including
25 copyright management information found in and on Plaintiffs' Works. When it removed this
26 information, Defendant did so without the authority of the copyright owners.

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

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

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

PRAYER FOR RELIEF

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

- a. Judgment in favor of Plaintiffs against the Defendant;
- b. A declaration that the Defendant has infringed Plaintiffs' exclusive copyrights under the Copyright Act;
- c. A declaration that such infringement is willful;
- d. A declaration that Defendant violated 17 U.S.C. § 1202(b) through its removal of copyright management information;
- e. A permanent injunction enjoining the Defendant and all those acting in concert with it from engaging in the infringing conduct alleged herein;
- f. That the Defendant be directed to account to Plaintiffs for all gains, profits, and advantages derived from their unlawful acts;
- g. An award of statutory damages under the Copyright Act;
- h. An award of statutory or actual damages under 17 U.S.C. § 1203(c);

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

6 **JURY TRIAL DEMANDED**

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

8 Dated: October 19, 2026

Respectfully submitted,

9 /s/ Elizabeth Brannen

10 Elizabeth Brannen (SBN 226234)

11 Lauren Martin (SBN 294367)

STRIS & MAHER LLP

12 17785 Center Court Dr N, Ste 600

Cerritos, CA 90703

13 T: (213) 995-6800

14 F: (213) 261-0299

ebrannen@stris.com

15 lmartin@stris.com

16 Kyle Roche (*pro hac vice* forthcoming)

FREEDMAN NORMAND FRIEDLAND LLP

17 155 E. 44th Street, Ste 915

New York, NY 10017

18 T: (646) 494-2900

19 kroche@fnf.law

20 *Counsel for Plaintiffs*