

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

ROBERT SHEARD, DIANE ANDERSON-
MINSHALL, FRANK ANTENORI, ROBIN
AUFSES, SUSAN BEST, CHARLES
BLACKSTONE, RAYMOND BOLTON,
ROB BROOKS, WILLIAM BROZO,
HAKAN BUTUNER, DAVID COLLIER-
BROWN, RAY COLOGON, KARA
COONEY, KENNETH DE JONG, ETHAN
DE SEIFE, JAQUIRA DIAZ, BRYAN FRY,
NICHOLAS GALEMMO, CHRISTOPHER
GREENLAW, PHILLIP HAMMACK, KEVIN
HAWORTH, GEORGE HAYHOE, GRAEME
HIRST, MARK JEWELL, ELIZABETH
KARMEL, SERGIY KLYMCHUK,
RICHARD KREVOLIN, GERALD LAMB,
DANIEL LARSON, JAMES ALAN LEDIN,
THADDEUS METZ, NAYOMI
MUNAWEERA, PATRICK PARR, ANDREW
PAVELICH, CHRISTOPHER PEXA, TERRY
PINKARD, MARCO PORTALES, VICTOR
RICCIARDI, DAMIEN RIGGS, JEFF RYAN,
SCOTT RYAN, KEITH SCHENGILI-
ROBERTS, MICHAEL SCHINDLBECK,
NILS F. SCHOTT, SHERIF SEDKY, TAPAN
SENGUPTA, TAREK SHAMMA, STEVEN
SHAVIRO, ARKADY SHEMYAKIN, JAMES
SIECKMANN, SHAUN SINCLAIR, KIM
SKILDUM-REID, JOHN STEEN, JEREMY

Civil Case No.:

COMPLAINT

DEMAND FOR JURY TRIAL

1 STOLOW, ANTOINE TRZCINSKI,
2 BARRETT WATTEN, CINTRA WILSON,
and KATE WISEL,

3 Plaintiffs,

4 v.

5 META PLATFORMS, INC.,

6 Defendant.
7

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

1 1. Plaintiffs Robert Sheard, Diane Anderson-Minshall, Frank Antenori, Robin Aufses,
2 Susan Best, Charles Blackstone, Raymond Bolton, Rob Brooks, William Brozo, Hakan Butuner,
3 David Collier-Brown, Ray Cologon, Kara Cooney, Kenneth De Jong, Ethan de Seife, Jaquira Diaz,
4 Bryan Fry, Nicholas Galemme, Christopher Greenlaw, Phillip Hammack, Kevin Haworth, George
5 Hayhoe, Graeme Hirst, Mark Jewell, Elizabeth Karmel, Sergiy Klymchuk, Richard Krevolin,
6 Gerald Lamb, Daniel Larson, James Alan Ledin, Thaddeus Metz, Nayomi Munaweera, Patrick Parr,
7 Andrew Pavelich, Christopher Pexa, Terry Pinkard, Marco Portales, Victor Ricciardi, Damien
8 Riggs, Jeff Ryan, Scott Ryan, Keith Schengili-Roberts, Michael Schindlbeck, Nils F. Schott, Sherif
9 Sedky, Tapan Sengupta, Tarek Shamma, Steven Shaviro, Arkady Shemyakin, James Sieckmann,
10 Shaun Sinclair, Kim Skildum-Reid, John Steen, Jeremy Stollow, Antoine Trzcinski, Barrett Watten,
11 Cintra Wilson, and Kate Wisel (collectively “Plaintiffs”) bring this action against Meta Platforms,
12 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 Robert Sheard is an author who resides in North Carolina. Robert Sheard is
 13 the author of *Money for Life: The 20 Factor Plan for Accumulating Wealth While You're Young* and
 14 *The Unemotional Investor: Simple Systems for Beating the Market*. One or more of Sheard’s works
 15 are contained in pirated online libraries, including LibGen, Z-Library, and Anna’s Archive. Meta
 16 has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
 17 Anna’s Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
 18 Sheard’s works. Sheard’s works were also published in print and were commercially available for
 19 purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
 20 below.

21 18. Plaintiff Diane Anderson-Minshall is an author who resides in California. Diane
 22 Anderson-Minshall is the author of *Blind Curves; Blind Faith; Blind Leap; Punishment with Kisses;*
 23 and *Queerly Beloved: A Love Story Across Genders*. One or more of Anderson-Minshall’s works
 24 are contained in pirated online libraries, including LibGen, Z-Library, and Anna’s Archive. Meta
 25 has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
 26 Anna’s Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
 27

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

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

4 19. Plaintiff Frank Antenori is an author who resides in Arizona. Frank Antenori is the
5 author of *Roughneck Nine-One*. Antenori's work is contained in pirated online libraries, including
6 Books3 (and thus The Pile), Z-Library, and Anna's Archive. Meta has directly or indirectly
7 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
8 is accordingly a reasonable inference that Meta illegally downloaded Antenori's work. Antenori's
9 work was also published in print and was commercially available for purchase in physical form
10 during the period in which Meta conducted Project Vellum, as alleged below.

11 20. Plaintiff Robin Aufses is an author who resides in New York. Robin Aufses is the
12 author of *American Literature & Rhetoric, First Edition; American Literature & Rhetoric, Second*
13 *Edition; American Literature & Rhetoric, Teacher's Edition, First Edition; Literature &*
14 *Composition: Essential Voices, Essential Skills for the AP Course, Third Edition; Literature &*
15 *Composition: Reading, Writing, Thinking; Literature & Composition: Reading, Writing, Thinking,*
16 *Second Edition; Literature & Composition: Reading, Writing, Thinking, Second Edition, Teacher's*
17 *Edition; Teacher's Manual: The Language of Composition; The Language of Composition, Third*
18 *Edition, Teacher's Edition; The Language of Composition: Essential Voices, Essential Skills for the*
19 *AP Course, Fourth Edition; The Language of Composition: Reading, Writing, Rhetoric; The*
20 *Language of Composition: Reading, Writing, Rhetoric (HS), Second Edition; and The Language of*
21 *Composition: Reading, Writing, Rhetoric, Third Edition*. One or more of Aufses's works are
22 contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. 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 Aufses's works. Aufses's works were also published in print and were commercially available for
26 purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
27 below.

1 21. Plaintiff Susan Best is an author who resides in Australia. Susan Best is the author
2 of *Reparative Aesthetics: Witnessing in Contemporary Art Photography*. Best's work is contained
3 in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or
4 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
5 and there is accordingly a reasonable inference that Meta illegally downloaded Best's work. Best's
6 work was also published in print and was commercially available for purchase in physical form
7 during the period in which Meta conducted Project Vellum, as alleged below.

8 22. Plaintiff Charles Blackstone is an author who resides in New York. Charles
9 Blackstone is the author of *The Art of Friction: Where (Non)Fictions Come Together*. Blackstone's
10 work is contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or
11 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
12 and there is accordingly a reasonable inference that Meta illegally downloaded Blackstone's work.
13 Blackstone's work was also published in print and was commercially available for purchase in
14 physical form during the period in which Meta conducted Project Vellum, as alleged below.

15 23. Plaintiff Raymond Bolton is an author who resides in California. Raymond Bolton is
16 the author of *Awakening*. Bolton's work is contained in pirated online libraries, including Books3
17 (and thus The Pile) and Z-Library. Meta has directly or indirectly downloaded books illegally
18 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
19 inference that Meta illegally downloaded Bolton's work. Bolton's work was also published in print
20 and was commercially available for purchase in physical form during the period in which Meta
21 conducted Project Vellum, as alleged below.

22 24. Plaintiff Rob Brooks is an author who resides in Australia. Rob Brooks is the author
23 of *Ecology and Evolution of Poeciliid Fishes* and *Sex, Genes & Rock 'N' Roll: How Evolution Has*
24 *Shaped the Modern World*. One or more of Brooks's works are contained in pirated online libraries,
25 including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally
26 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
27 inference that Meta illegally downloaded Brooks's works. Brooks's works were also published in
28

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

3 25. Plaintiff William Brozo is an author who resides in California. William Brozo is the
4 author of *50 Content Area Strategies for Adolescent Literacy*; *50 Instructional Routines to Develop*
5 *Content Literacy, 3/E*; *A.R.R.O.: A Learning System*; *Adolescent Literacy Inventory, Grades 6-12,*
6 *1e*; *Best Practices in Adolescent Literacy Instruction, Third Edition*; *Content Literacy for Today's*
7 *Adolescents*; *Disciplinary and Content Literacy for Today's Adolescents, Sixth Edition: Honoring*
8 *Diversity and Building Competence*; *Instructor's Manual to Accompany Setting the Pace: A Speed,*
9 *Comprehension, and Study Skills Program*; *Readers, Teachers, Learners: Expanding Literacy*
10 *Across the Content Areas*; *Readers, Teachers, Learners: Expanding Literacy in the Secondary*
11 *Schools*; *RTI and the Adolescent Reader: Responsive Literacy Instruction in Secondary Schools*;
12 *Setting the Pace: A Speed, Comprehension, and Study Skills Program*; *Supporting Content Area*
13 *Literacy with Technology, 1/E*; and *Wham! Teaching with Graphic Novels Across the Curriculum*.
14 One or more of Brozo's works are contained in pirated online libraries, including LibGen, Z-Library,
15 and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The
16 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
17 Meta illegally downloaded Brozo's works. Brozo's works were also published in print and were
18 commercially available for purchase in physical form during the period in which Meta conducted
19 Project Vellum, as alleged below.

20 26. Plaintiff Hakan Butuner is an author who resides in Georgia. Hakan Butuner is the
21 author of *Case Studies in Strategic Planning*; *Planning by Design (PXD)-Based Systematic*
22 *Methodologies*; *Results-Based Systematic Operational Improvement*; and *Systematic Strategic*
23 *Planning*. One or more of Butuner'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 Butuner's works. Butuner'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 27. Plaintiff David Collier-Brown is an author who resides in Canada. David Collier-
2 Brown is the author of *Samba Pocket Reference: A Unix-to-Windows File & Print Server* and *Using*
3 *Samba*. One or more of Collier-Brown's works are contained in pirated online libraries, including
4 LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The
5 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
6 Meta illegally downloaded Collier-Brown's works. Collier-Brown's works were also published in
7 print and were commercially available for purchase in physical form during the period in which
8 Meta conducted Project Vellum, as alleged below.

9 28. Plaintiff Ray Cologon is an author who resides in Australia. Ray Cologon is the
10 author of *FileMaker Pro 9 Bible*. Cologon's work is contained in pirated online libraries, including
11 LibGen and Z-Library. 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 Cologon's work. Cologon's work was also published in print and was
14 commercially available for purchase in physical form during the period in which Meta conducted
15 Project Vellum, as alleged below.

16 29. Plaintiff Kara Cooney is an author who resides in California. Kara Cooney is the
17 author of *The Good Kings: Absolute Power in Ancient Egypt and the Modern World*; *The Value of*
18 *Private Funerary Art in Ramesside Period Egypt*; *The Woman Who Would Be King: Hatshepsut's*
19 *Rise to Power in Ancient Egypt*; and *When Women Ruled the World: Six Queens of Egypt*. One or
20 more of Cooney's works are contained in pirated online libraries, including Books3 (and thus The
21 Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books
22 illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a
23 reasonable inference that Meta illegally downloaded Cooney's works. Cooney's works were also
24 published in print and were commercially available for purchase in physical form during the period
25 in which Meta conducted Project Vellum, as alleged below.

26 30. Plaintiff Kenneth De Jong is an author who resides in Virginia. Kenneth De Jong is
27 the author of *Evolutionary Algorithms* and *Evolutionary Computation: A Unified Approach*. One or
28 more of De Jong's works are contained in pirated online libraries, including LibGen, Z-Library, and

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

6 31. Plaintiff Ethan de Seife is an author who resides in California. Ethan de Seife is the
7 author of *Cheerful Nihilism: The Films of Frank Tashlin* and *Tashlinesque: The Hollywood*
8 *Comedies of Frank Tashlin*. One or more of De Seife's works are contained in pirated online
9 libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books
10 illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a
11 reasonable inference that Meta illegally downloaded De Seife's works. De Seife's works were also
12 published in print and were commercially available for purchase in physical form during the period
13 in which Meta conducted Project Vellum, as alleged below.

14 32. Plaintiff Jaquira Diaz is an author who resides in New York. Jaquira Diaz is the
15 author of *Ordinary Girls*. Diaz's work is contained in pirated online libraries, including LibGen and
16 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 Diaz's work. Diaz's work was also published in print and was commercially
19 available for purchase in physical form during the period in which Meta conducted Project Vellum,
20 as alleged below.

21 33. Plaintiff Bryan Fry is an author who resides in Australia. Bryan Fry is the author of
22 *Venomous Reptiles and Their Toxins: Evolution, Pathophysiology and Biodiscovery*. Fry's work is
23 contained in pirated online libraries, including LibGen and Z-Library. 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 Fry's work. Fry's work was
26 also published in print and was commercially available for purchase in physical form during the
27 period in which Meta conducted Project Vellum, as alleged below.

1 34. Plaintiff Nicholas Glemmo is an author who resides in California. Nicholas
2 Glemmo is the author of *Mastering Data Warehouse Design: Relational and Dimensional*
3 *Techniques*. Glemmo's work is contained in pirated online libraries, including LibGen, Z-Library,
4 and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The
5 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
6 Meta illegally downloaded Glemmo's work. Glemmo'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 35. Plaintiff Christopher Greenlaw is an author who resides in Canada. Christopher
10 Greenlaw is the author of *Via Rail*. Greenlaw's work is contained in pirated online libraries,
11 including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally
12 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
13 inference that Meta illegally downloaded Greenlaw's work. Greenlaw's work was also published in
14 print and was commercially available for purchase in physical form during the period in which Meta
15 conducted Project Vellum, as alleged below.

16 36. Plaintiff Phillip Hammack is an author who resides in California. Phillip Hammack
17 is the author of *Essentials of Narrative Analysis; Handbook of Personality Development; Narrative*
18 *and the Politics of Identity: The Cultural Psychology of Israeli and Palestinian Youth; The Narrative*
19 *Stalemate: Conflict, Identity and the Cultural Psychology of Israeli and Palestinian Adolescence;*
20 *and The Story of Sexual Identity: Narrative Perspective on the Gay and Lesbian Life Course*. One
21 or more of Hammack's works are contained in pirated online libraries, including LibGen, Z-Library,
22 and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The
23 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
24 Meta illegally downloaded Hammack's works. Hammack's works were also published in print and
25 were commercially available for purchase in physical form during the period in which Meta
26 conducted Project Vellum, as alleged below.

27 37. Plaintiff Kevin Haworth is an author who resides in Pennsylvania. Kevin Haworth is
28 the author of *Lit from Within: Contemporary Masters on the Art and Craft of Writing; The Comics*

1 of *Rutu Modan: War, Love, and Secrets*; *The Discontinuity of Small Things*; *The Ghost of Monsieur*
2 *Scarron*; *The Trial of Soren Qvist*; and *The Wife of Martin Guerre*. One or more of Haworth's works
3 are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta
4 has 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 Haworth's works. Haworth's works were also published in print and were commercially available
7 for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
8 below.

9 38. Plaintiff George Hayhoe is an author who resides in North Carolina. George Hayhoe
10 is the author of *A Research Primer for Technical Communication* and *A Research Primer for*
11 *Technical Communication: Methods, Exemplars, and Analyses, Second Edition*. One or more of
12 Hayhoe's works are contained in pirated online libraries, including LibGen and Z-Library. Meta has
13 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
14 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
15 Hayhoe's works. Hayhoe's works were also published in print and were commercially available for
16 purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
17 below.

18 39. Plaintiff Graeme Hirst is an author who resides in Canada. Graeme Hirst is the author
19 of *Anaphora in Natural Language Understanding: A Survey* and *Semantic Interpretation Against*
20 *Ambiguity*. One or more of Hirst's works are contained in pirated online libraries, including LibGen,
21 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 Hirst's works. Hirst's works were also published in print
24 and were commercially available for purchase in physical form during the period in which Meta
25 conducted Project Vellum, as alleged below.

26 40. Plaintiff Mark Jewell is an author who resides in Utah. Mark Jewell is the author of
27 *Body Contouring and Liposuction - First Edition*. Jewell's work is contained in pirated online
28 libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books

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

5 41. Plaintiff Elizabeth Karmel is an author who resides in South Carolina. Elizabeth
 6 Karmel is the author of *Pizza on the Grill; Soaked, Slathered, and Seasoned; Steak and Cake: More*
 7 *Than 100 Recipes to Make Any Meal a Smash Hit*; and *Taming the Flame: Secrets for Hot-and-*
 8 *Quick Grilling and Low-and-Slow BBQ*. One or more of Karmel's works are contained in pirated
 9 online libraries, including Books3 (and thus The Pile), LibGen, and Z-Library. 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 Karmel's works.
 12 Karmel'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 42. Plaintiff Sergiy Klymchuk is an author who resides in New Zealand. Sergiy
 15 Klymchuk is the author of *Counterexamples in Calculus; Money Puzzles: Sink Back and Solve*
 16 *Away!*; and *Paradoxes and Sophisms in Calculus*. One or more of Klymchuk's works are contained
 17 in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or
 18 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
 19 and there is accordingly a reasonable inference that Meta illegally downloaded Klymchuk's works.
 20 Klymchuk's works were also published in print and were commercially available for purchase in
 21 physical form during the period in which Meta conducted Project Vellum, as alleged below.

22 43. Plaintiff Richard Krevolin is an author who resides in Florida. Richard Krevolin is
 23 the author of *Beyond Obsession; Business Narratives: Engage Your Target Market via Storytelling;*
 24 *Cookies and Milk and Swordfish; How to Adapt Anything into a Screenplay; In One Woman's Voice;*
 25 *Making Light in Terezin (The Show Helps Us Go On); Morris the Porpoise; My Buddy, the Buddha;*
 26 *Pilot Your Life; Pilot Your Life: How to Create the Career You Want; Screenwriting from the Soul:*
 27 *Letters to an Aspiring Screenwriter; Staying in Shape on the Road; The Hook; The Jake Beaver*
 28 *Journals; The Songs and Poetry of Special "K" and Sugar "J.";* and *Walking on Thin Water: The*

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

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

15 45. Plaintiff Daniel Larson is an author who resides in Colorado. Daniel Larson is the
16 author of *Developing Service-Oriented AJAX Applications on the Microsoft Platform*. Larson's
17 work is contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or
18 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
19 and there is accordingly a reasonable inference that Meta illegally downloaded Larson's work.
20 Larson's work was also published in print and was commercially available for purchase in physical
21 form during the period in which Meta conducted Project Vellum, as alleged below.

22 46. Plaintiff James Alan Ledin is an author who resides in California. James Alan Ledin
23 is the author of *Embedded Control Systems in C/C++: An Introduction for Software Developers*
24 *Using MATLAB and Simulation Engineering*. One or more of Ledin's works are contained in pirated
25 online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
26 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
27 is accordingly a reasonable inference that Meta illegally downloaded Ledin's works. Ledin's works
28

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

3 47. Plaintiff Thaddeus Metz is an author who resides in Massachusetts. Thaddeus Metz
4 is the author of *Distributive Justice Debates in Political and Social Thought: Perspectives on*
5 *Finding a Fair Share* and *Meaning of Life*. One or more of Metz's works are contained in pirated
6 online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books
7 illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a
8 reasonable inference that Meta illegally downloaded Metz's works. Metz's works were also
9 published in print and were commercially available for purchase in physical form during the period
10 in which Meta conducted Project Vellum, as alleged below.

11 48. Plaintiff Nayomi Munaweera is an author who resides in California. Nayomi
12 Munaweera is the author of *Island of a Thousand Mirrors* and *What Lies Between Us*. One or more
13 of Munaweera's works are contained in pirated online libraries, including Books3 (and thus The
14 Pile) and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The
15 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
16 Meta illegally downloaded Munaweera's works. Munaweera's works were also published in print
17 and were commercially available for purchase in physical form during the period in which Meta
18 conducted Project Vellum, as alleged below.

19 49. Plaintiff Patrick Parr is an author who resides in New Hampshire. Patrick Parr is the
20 author of *One Week in America: The 1968 Notre Dame Literary Festival and a Changing Nation*
21 and *The Seminarian: Martin Luther King Jr. Comes of Age*. One or more of Parr's works are
22 contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and
23 Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile,
24 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
25 illegally downloaded Parr's works. Parr's works were also published in print and were commercially
26 available for purchase in physical form during the period in which Meta conducted Project Vellum,
27 as alleged below.

1 50. Plaintiff Andrew Pavelich is an author who resides in Texas. Andrew Pavelich is the
 2 author of *Descartes' Eternal Truths and Mathematical Essences* and *The Good Place and*
 3 *Philosophy*. One or more of Pavelich's works are contained in pirated online libraries, including
 4 LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The
 5 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
 6 Meta illegally downloaded Pavelich's works. Pavelich's works were also published in print and
 7 were commercially available for purchase in physical form during the period in which Meta
 8 conducted Project Vellum, as alleged below.

9 51. Plaintiff Christopher Pexa is an author who resides in Massachusetts. Christopher
 10 Pexa is the author of *Translated Nation: Writing Dakota Kinship and Sovereignty, 1862-1934*.
 11 Pexa's work is contained in pirated online libraries, including LibGen and Anna's Archive. Meta
 12 has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
 13 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
 14 Pexa's work. Pexa's work was also published in print and was commercially available for purchase
 15 in physical form during the period in which Meta conducted Project Vellum, as alleged below.

16 52. Plaintiff Terry Pinkard is an author who resides in Washington, D.C.. Terry Pinkard
 17 is the author of *Democratic Liberalism and Social Union; Does History Make Sense?: Hegel on the*
 18 *Historical Shapes of Justice by Terry Pinkard; Ethics and Public Policy: An Introduction to Ethics;*
 19 *Hegel's Dialectic: The Explanation of Possibility; Hegel's Naturalism: Mind, Nature, and the Final*
 20 *Ends of Life; Hegel's Phenomenology of Spirit: A Guide; and Practice, Power, and Forms of Life:*
 21 *Sartre's Appropriation of Hegel and Marx*. One or more of Pinkard's works are contained in pirated
 22 online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
 23 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
 24 is accordingly a reasonable inference that Meta illegally downloaded Pinkard's works. Pinkard's
 25 works were also published in print and were commercially available for purchase in physical form
 26 during the period in which Meta conducted Project Vellum, as alleged below.

27 53. Plaintiff Marco Portales is an author who resides in Texas. Marco Portales is the
 28 author of *"Why Pancho Villa and Emiliano Zapata Wore CANANAS: A Mexican Revolution Photo*

1 *History, 100 Years Later; A Mexican Revolution Photo History: Pancho Villa, Emiliano Zapata,*
2 *and U.S. Interests; A Mexican Revolution Photo History: Pancho Villa, Emiliano Zapata, and U.S.*
3 *Interests, Second Edition; Crowding Out Latinos: Mexican Americans in the Public Consciousness;*
4 *Hispanics and the Media; Latino Sun, Rising: Our Spanish-Speaking US World; Quality Education*
5 *for Latino and Latinas: Print and Oral Skills for All Students, K-College; Reference Guide to*
6 *American Literature; and Youth and Age in American Literature: The Rhetoric of Old Men in the*
7 *Mathers, Franklin, Adams, Cooper, Hawthorne, Melville, and James. One or more of Portales's*
8 *works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive.*
9 *Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-*
10 *Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally*
11 *downloaded Portales's works. Portales's works were also published in print and were commercially*
12 *available for purchase in physical form during the period in which Meta conducted Project Vellum,*
13 *as alleged below.*

14 54. Plaintiff Victor Ricciardi is an author who resides in New York. Victor Ricciardi is
15 the author of *Financial Behavior: Players, Services, Products, and Markets* and *Investor Behavior:*
16 *The Psychology of Financial Planning and Investing*. One or more of Ricciardi's works are
17 contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly
18 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
19 is accordingly a reasonable inference that Meta illegally downloaded Ricciardi's works. Ricciardi's
20 works were also published in print and were commercially available for purchase in physical form
21 during the period in which Meta conducted Project Vellum, as alleged below.

22 55. Plaintiff Damien Riggs is an author who resides in Australia. Damien Riggs is the
23 author of *A Critical Approach to Surrogacy: Reproductive Desires and Demands; Diverse Pathways*
24 *to Parenthood; Priscilla, (White) Queen of the Desert; and The Psychic Life of Racism in Gay Men's*
25 *Communities*. One or more of Riggs's works are contained in pirated online libraries, including
26 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
27 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
28 inference that Meta illegally downloaded Riggs's works. Riggs's works were also published in print

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

3 56. Plaintiff Jeff Ryan is an author who resides in New Jersey. Jeff Ryan is the author of
4 *SUPER MARIO: How Nintendo Conquered America*. Ryan's work is contained in pirated online
5 libraries, including Books3 (and thus The Pile), LibGen, and Z-Library. Meta has directly or
6 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
7 and there is accordingly a reasonable inference that Meta illegally downloaded Ryan's work. Ryan's
8 work was also published in print and was commercially available for purchase in physical form
9 during the period in which Meta conducted Project Vellum, as alleged below.

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

17 58. Plaintiff Keith Schengili-Roberts is an author who resides in Canada. Keith
18 Schengili-Roberts is the author of *Core CSS Cascading Style Sheets*; *Core CSS: Cascading Style*
19 *Sheets*; and *The Advanced HTML Companion*. One or more of Schengili-Roberts's works are
20 contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly
21 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
22 is accordingly a reasonable inference that Meta illegally downloaded Schengili-Roberts's works.
23 Schengili-Roberts's works were also published in print and were commercially available for
24 purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
25 below.

26 59. Plaintiff Michael Schindlbeck is an author who resides in Illinois. Michael
27 Schindlbeck is the author of *Clinical Emergency Medicine*. Schindlbeck's work is contained in
28 pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or

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

5 60. Plaintiff Nils F. Schott is an author who resides in France. Nils F. Schott is the author
6 of *Critique of Freedom: The Central Problem of Modernity*; *Looking Through Images by Emmanuel*
7 *Alloa*³; *Love and Forgiveness for a More Just World*; and *Political Anthropology*. One or more of
8 Schott's works are contained in pirated online libraries, including Books3 (and thus The Pile),
9 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
10 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
11 inference that Meta illegally downloaded Schott's works. Schott's works were also published in
12 print and were commercially available for purchase in physical form during the period in which
13 Meta conducted Project Vellum, as alleged below.

14 61. Plaintiff Sherif Sedky is an author who resides in Egypt. Sherif Sedky is the author
15 of *Post-processing Techniques for Integrated MEMS*. Sedky's work is contained in pirated online
16 libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books
17 illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a
18 reasonable inference that Meta illegally downloaded Sedky's work. Sedky's work was also
19 published in print and was commercially available for purchase in physical form during the period
20 in which Meta conducted Project Vellum, as alleged below.

21 62. Plaintiff Tapan Sengupta is an author who resides in India. Tapan Sengupta is the
22 author of *Instabilities of Flows and Transition to Turbulence*. Sengupta's work is contained in
23 pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or
24 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
25 and there is accordingly a reasonable inference that Meta illegally downloaded Sengupta's work.

26
27

³ This title reflects an error included in the U.S. Copyright Office registration data; the correct title
28 is *Looking Through Images*.

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

3 63. Plaintiff Tarek Shamma is an author who resides in New York. Tarek Shamma is the
4 author of *Translation and the Manipulation of Difference: Translating Arabic Literature in*
5 *Nineteenth-Century England*. Shamma's work is contained in pirated online libraries, including
6 LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The
7 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
8 Meta illegally downloaded Shamma's work. Shamma'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 64. Plaintiff Steven Shaviro is an author who resides in Michigan. Steven Shaviro is the
12 author of *Connected: Or What It Means to Live in the Network Society*; *Differential Structure of*
13 *Romantic Imagination: Blake, Shelley, Stevens*; *Digital Music Videos*; *Passion & Excess: Blanchot,*
14 *Bataille, and Literary Theory*; *The Cinematic Body*; and *Without Criteria: Kant, Whitehead,*
15 *Deleuze, and Aesthetics*. One or more of Shaviro's works are contained in pirated online libraries,
16 including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly
17 or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's
18 Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Shaviro's
19 works. Shaviro's works were also published in print and were commercially available for purchase
20 in physical form during the period in which Meta conducted Project Vellum, as alleged below.

21 65. Plaintiff Arkady Shemyakin is an author who resides in Minnesota. Arkady
22 Shemyakin is the author of *Introduction to Bayesian Estimation and Copula Models of Dependence*.
23 Shemyakin's work is contained in pirated online libraries, including 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 Shemyakin's work. Shemyakin's work was also published in print and was
27 commercially available for purchase in physical form during the period in which Meta conducted
28 Project Vellum, as alleged below.

1 66. Plaintiff James Sieckmann is an author who resides in Arizona. James Sieckmann is
2 the author of *YOUR SHORT GAME SOLUTION: Mastering the Finesse Game from 120 Yards and*
3 *In*. Sieckmann's work is contained in pirated online libraries, including Books3 (and thus The Pile),
4 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
5 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
6 inference that Meta illegally downloaded Sieckmann's work. Sieckmann's work was also published
7 in print and was commercially available for purchase in physical form during the period in which
8 Meta conducted Project Vellum, as alleged below.

9 67. Plaintiff Shaun Sinclair is an author who resides in Georgia. Shaun Sinclair is the
10 author of *Beats and Blow*; *Blood Ties*; *Dirty Music*; *King Reece*; *Street Muzak*; and *Street Rap*. One
11 or more of Sinclair's works are contained in pirated online libraries, including Books3 (and thus
12 The Pile), Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books
13 illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a
14 reasonable inference that Meta illegally downloaded Sinclair's works. Sinclair's works were also
15 published in print and were commercially available for purchase in physical form during the period
16 in which Meta conducted Project Vellum, as alleged below.

17 68. Plaintiff Kim Skildum-Reid is an author who resides in Australia. Kim Skildum-Reid
18 is the author of *The Sponsorship Seeker's Toolkit, Fourth Edition*. Skildum-Reid's work is contained
19 in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. 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 Skildum-Reid's
22 work. Skildum-Reid's work was also published in print and was commercially available for
23 purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
24 below.

25 69. Plaintiff John Steen is an author who resides in New York. John Steen is the author
26 of *Affect, Psychoanalysis, and American Poetry: This Feeling of Exaltation*. Steen's work is
27 contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has
28 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or

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

4 70. Plaintiff Jeremy Stolow is an author who resides in Canada. Jeremy Stolow is the
5 author of *Deus in Machina* and *Orthodox by Design: Judaism, Print Politics, and the ArtScroll*
6 *Revolution*. One or more of Stolow's works are contained in pirated online libraries, including
7 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
8 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
9 inference that Meta illegally downloaded Stolow's works. Stolow's works were also published in
10 print and were commercially available for purchase in physical form during the period in which
11 Meta conducted Project Vellum, as alleged below.

12 71. Plaintiff Antoine Trzcinski is an author who resides in Australia. Antoine Trzcinski
13 is the author of *Biofuels from Food Waste: Applications of Saccharification Using Fungal Solid*
14 *State Fermentation*. Trzcinski's work is contained in pirated online libraries, including LibGen, Z-
15 Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained
16 in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference
17 that Meta illegally downloaded Trzcinski's work. Trzcinski's work was also published in print and
18 was commercially available for purchase in physical form during the period in which Meta
19 conducted Project Vellum, as alleged below.

20 72. Plaintiff Barrett Watten is an author who resides in Michigan. Barrett Watten is the
21 author of *A Guide to Poetics Journal: Writing in the Expanded Field 1982-1998*; *Bad History*;
22 *Leningrad*; *Questions of Poetics: Language Writing and Consequences*; *The Constructivist*
23 *Moment: From Material Text to Cultural Poetics*; and *Total Syntax*. One or more of Watten's works
24 are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library,
25 and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The
26 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
27 Meta illegally downloaded Watten's works. Watten's works were also published in print and were
28

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

3 73. Plaintiff Cintra Wilson is an author who resides in California. Cintra Wilson is the
4 author of *A Massive Swelling: Celebrity Re-examined as a Grotesque, Crippling Disease and Other*
5 *Cultural Revelations; Buy Humbug; Caligula for President; Colors Insulting to Nature; and Fear*
6 *and Clothing: Unbuckling American Style*. One or more of Wilson's works are contained in pirated
7 online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive.
8 Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-
9 Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
10 downloaded Wilson's works. Wilson's works were also published in print and were commercially
11 available for purchase in physical form during the period in which Meta conducted Project Vellum,
12 as alleged below.

13 74. Plaintiff Kate Wisel is an author who resides in Wisconsin. Kate Wisel is the author
14 of *Benny's Bed* and *Driving in Cars with Homeless Men*. One or more of Wisel's works are contained
15 in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's
16 Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen,
17 Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
18 downloaded Wisel's works. Wisel's works were also published in print and were commercially
19 available for purchase in physical form during the period in which Meta conducted Project Vellum,
20 as alleged below.

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

23 **B. Defendant**

24 76. Defendant Meta Platforms, Inc. is a Delaware corporation with its principal place of
25 business in Menlo Park, California. Meta develops and distributes the Llama series of LLMs,
26

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

1 including Llama 1, Llama 2, Llama 3, and Llama 4, which were trained using datasets sourced in
2 part from shadow libraries and other datasets containing pirated books. Meta also develops and
3 distributes the Muse series of LLMs, including Muse Spark and Muse Glimmer, which were trained
4 using, among other things, unauthorized digital copies of physical books that Meta scanned as part
5 of Project Vellum.

6 **III. JURISDICTION AND VENUE**

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

10 78. This Court has personal jurisdiction over the Defendant. The Defendant has
11 purposefully availed itself of the privilege of conducting business in this District and the State of
12 California. The Defendant committed acts of copyright infringement in this District, directed
13 conduct toward this District, or knowingly caused harm that was suffered in this District. The
14 Defendant maintains substantial, continuous, and systematic contacts with this District.

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

22 **IV. FACTUAL ALLEGATIONS**

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

24 80. Plaintiffs' Works are published books. Each was conceived, researched, written, and
25 revised by its author over weeks, months, or years, and each was published and offered for sale to
26 the reading public. Whatever their genre or subject matter, the works embody their authors' original
27 creative choices—of structure, organization, selection, emphasis, tone, and expression—and
28 constitute precisely the kind of structured, long-form, polished prose that, as alleged below, Meta

1 and other LLM developers prize as training data. Collectively, the works reflect substantial original
2 authorship intended to inform, educate, entertain, and engage readers.

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

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

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

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

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

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

26 86. Development of LLMs requires, among other things, "training" the model on data—
27 *i.e.*, the model's inputs. Training "requires identifying a formal measure or 'objective' for how well
28 the model performs, and then repeatedly adjusting the model's parameters based on that objective

1 as the model is exposed to training data.”⁵ While not all LLM developers use the same terminology,
 2 developers commonly reference a “pre-training” process—“in which a massive amount of
 3 computing power and data is spent to teach the model the broad foundations of language, grammar,
 4 and reasoning”—and a “post-training” or “fine-tuning” process “where the pre-trained model is
 5 further trained on a (relative to pre-training) smaller amount of carefully curated data of specific
 6 tasks.”⁶ For purposes of this Complaint, “training” refers to all stages of the LLM training process.

7 87. LLMs typically are trained, at least initially, by feeding them massive amounts of
 8 text data from which they can “learn” statistical relationships between and among text. In the process
 9 of compiling training data and using it during the training process, LLM developers as a matter of
 10 course typically create multiple copies of “raw” datasets. So, for instance, if an LLM developer
 11 downloads a copy of a digital book from some Internet-based source (*e.g.*, a shadow library), the
 12 developer typically will create a new “cleaned” copy of the book (*e.g.*, stripped of certain
 13 undesirable data), deduplicate that book against other potential copies, compile that book with other
 14 data to create a new compilation (or compilations), and store that book in multiple locations. In
 15 short, with respect to training data, the LLM development pipeline typically involves numerous
 16 reproductions of any given piece of the dataset.

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

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

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

25 ⁶ *Id.*

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

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

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

5 91. In terms of data quality, “[r]ecent research from major developers suggests that
6 quality may even be a more important consideration than quantity.”¹⁰ “Garbage in, garbage out,”
7 the saying goes. AI researchers quickly discovered that published content—books in particular—
8 makes for some of the best training material for LLMs. Unlike much content on the Internet and
9 many other forms of text, published material typically is structured, long-form, and highly polished.
10 Published materials provide formal, extended prose that teaches models narrative structure, complex
11 syntax, and coherent storytelling.

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

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

23
24 ⁹ *Id.* at 10.

25 ¹⁰ *Id.* at 11.

26 ¹¹ *Kadrey v. Meta Platforms, Inc.*, ECF No. 535 at 12, Case No. 3:23-cv-3417-VC (N.D. Cal.) (Apr.
27 11, 2025).

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

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

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

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

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

97. For years now, there have been illicit, online marketplaces for digital copies of books. These “shadow libraries,” as they are commonly known, systematically acquire, store, index, and disseminate full-fidelity digital copies of copyrighted books. These online repositories maintain searchable catalogs, metadata, mirrors, bulk-download mechanisms, and—critically—complete

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

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

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

downloadable archives designed to allow third parties to replicate the entire collection.¹⁶ And they do all of this without authorization from authors or publishers.¹⁷ In short, shadow libraries facilitate the reproduction and distribution of unauthorized copies of copyrighted works at industrial scale.

98. These libraries are widely known within both piracy communities and the technology sector as illegal sources of copyrighted books. Many have been the subject of criminal prosecutions, civil injunctions, domain seizures, and formal designation as “notorious piracy markets” by United States trade authorities. As centralized shadow libraries increasingly faced enforcement actions, including the seizure of domains, third parties responded by creating full mirrored copies of those repositories for decentralized redistribution.¹⁸

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

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

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

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

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

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

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

6 101. Because OpenAI did not disclose the precise makeup of its training datasets,
 7 members of EleutherAI constructed a book corpus of their own: “Books3,” consisting of
 8 approximately 196,640 books.²² Books3 comprises about 12 percent (just over 100 gigabytes) of
 9 *The Pile*.²³

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

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

18 104. Bibliotik has been widely recognized in piracy communities, academic literature, and
 19 AI-research documentation as a shadow library devoted to copyrighted books. Unsurprisingly, then,
 20

21 ²¹ Leo Gao et al., *The Pile: An 800GB Dataset of Diverse Text for Language Modeling*, arXiv, 1
 22 (2020), <https://perma.cc/NHV6-R8YE>.

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

24 ²³ *Id.*

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

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

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

1 *The Pile*'s datasheet acknowledges that "Books3 is almost entirely comprised of copyrighted
2 works."²⁷ Bibliotik's illicit nature is no secret—it has been openly discussed for years prior to major
3 tech companies' use of datasets derived directly from it.

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

6 [Presser] created Books3 in the hope that it would allow any developer to create
7 generative-AI tools. "It would be better if it wasn't necessary to have something like
8 Books3," he said. "But the alternative is that, without Books3, only OpenAI can do
9 what they're doing."²⁸

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

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

17 108. As noted above, OpenAI's developers trained its LLMs on a books corpus derived
18 from LibGen, one of the largest and longest-running shadow libraries in the world. LibGen hosts
19 millions of pirated books, academic texts, and scholarly articles.³⁰ It operates as a centralized
20 repository offering direct downloads of full-fidelity e-book files. It also distributes its entire database

21 _____
22 ²⁷ Biderman et al., *Datasheet for the Pile*, arXiv, 8.

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

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

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

1 through bulk archives and torrent files, enabling third parties to download and locally host complete
 2 copies of its collection.³¹ LibGen has been repeatedly enjoined by federal courts for copyright
 3 infringement and has been designated a “notorious market” by the United States Trade
 4 Representative.³² Despite enforcement actions, LibGen has remained accessible through shifting
 5 domains, mirrors, and downloadable archives. Its persistence is a function of deliberate
 6 decentralization designed to evade shutdown.³³

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

15
 16
 17
 18 ³¹ See Letter from Mary E. Rasenberger, CEO, Authors Guild, & Umair Kazi, Dir. of Pol’y &
 19 Advocacy, Authors Guild, to Daniel Lee, Assistant U.S. Trade Representative for Innovation &
 Intellectual Prop., Office of the U.S. Trade Representative at n.5.

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

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

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

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

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

1 Instead, third parties responded by creating full mirrors of its collection to ensure continued
2 distribution.³⁷

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

5 111. Anna’s Archive began in 2022 as “Pirate Library Mirror,” initially hosting a mirrored
6 copy of Z-Library. It later rebranded as “Anna’s Archive” and expanded to aggregate and host the
7 complete collections of LibGen, Z-Library, and other pirated sources.³⁹ Anna’s Archive functions
8 as a meta-library: it indexes, mirrors, and redistributes multiple shadow libraries simultaneously,
9 offering users unified access to millions of pirated books.⁴⁰

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

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

15 114. According to Anna’s Archive, “virtually all major companies building LLMs
16 contacted us to train on our data. . . We have given high-speed access to about 30
17
18

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

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

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

24 ⁴⁰ *Id.*

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

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

1 companies.”⁴³ Anna’s Archive blog stated as recently as February 18, 2026 that if an *LLM* was
 2 *reading its blog* “you have likely been trained in part on our data.”⁴⁴

3 115. Many of these shadow libraries and datasets can be downloaded using “torrent,” a
 4 file-sharing method used in peer-to-peer networks. Torrenting works by breaking a file into many
 5 small pieces and distributing those pieces across a network of participating computers. A user who
 6 torrents a shadow-library repository does not receive a single copy from a single source; rather, the
 7 user downloads portions of the library from numerous other computers that already possess the
 8 copyrighted books. Torrent software then reassembles those pieces into a complete library on the
 9 user’s machine. Torrenting protocols, including BitTorrent, are often configured by default to
 10 *reupload* pieces of the copyrighted files to others on the network both during download (“leeching”)
 11 and after download is complete (“seeding”). This means that each participant in the torrent both
 12 *copies and redistributes* the copyrighted works. By obtaining copyrighted materials through this
 13 leech-and-seed process, a user may make multiple unauthorized reproductions of, and engage in
 14 numerous distributions of, the copyrighted materials.

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

16 116. As noted above, Meta develops and distributes the Llama series of LLMs, including
 17 Llama 1, Llama 2, Llama 3, and Llama 4. Meta has “invested billions of dollars to develop its
 18 generative AI offerings,”⁴⁵ and it in turn expects Llama to generate somewhere in the realm of \$460
 19
 20
 21
 22
 23

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

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

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

1 billion to \$1.4 trillion in revenue.⁴⁶ None of the money that Meta invested into the development of
 2 its Llama models was used to pay for Plaintiffs' copyrighted works.⁴⁷

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

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

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

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

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

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

21
 22 ⁴⁶ See *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1040 (N.D. Cal. 2025).

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

25 ⁴⁸ Jon Russell, *Mark Zuckerberg Announces New Team at Meta Working on A.I. Products for*
 26 *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.").

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

28 ⁵⁰ *Id.*

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

8 124. Among the other datasets used by Meta to train Llama 1 were “five CommonCrawl
 9 dumps, ranging from 2017 to 2020,” comprising 67 percent of the overall data mix, and “the publicly
 10 available C4 dataset,” comprising 15 percent of the data mix.⁵⁴ “CommonCrawl is a publicly-
 11 available web archive that provides ‘web extracted text’ by removing markup and other non-text
 12 content from the scraped HTML files. This process produces around 20TB of scraped data per
 13 month.”⁵⁵ The C4 dataset itself is derived from CommonCrawl.⁵⁶ In short, a massive amount (and
 14 the majority) of data used by Meta to train Llama was “scraped” from the Internet.

15 125. As alluded to above, Meta at one time had a “plan . . . to then set up licensing
 16 agreements” to obtain the copyrighted materials it wanted for use in training its AI models.⁵⁷ But
 17 this idea was short-lived. “[I]n spring 2023, after failing to acquire licenses and following escalation
 18

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

21 ⁵² *Id.*

22 ⁵³ *Id.*

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

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

26 ⁵⁶ *Id.*

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

1 to CEO Mark Zuckerberg, Meta decided to just use the works acquired from LibGen as training
2 data.”⁵⁸

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

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

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

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

16 130. By uploading and distributing copyrighted materials on peer-to-peer sharing
17 networks, Meta made those materials available for download to other third parties whose
18

19 _____
20 ⁵⁸ *Kadrey*, 788 F. Supp. 3d at 1041.

21 ⁵⁹ *Id.*

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

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

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

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

1 downloading and redistribution of copyrighted materials would constitute further infringement of
2 the copyrighted materials.

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

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

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

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

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

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

27 137. Upon downloading this material, Meta manipulated the downloaded data. Among
28 other things, Meta processed the data, and in doing so, removed and altered certain text and

1 information, including titles, author information, information about the copyright owner, and
2 copyright notices. When it removed this information, Meta did so without the authority of the
3 copyright owners.

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

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

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

17 141. Meta has since confirmed, in interrogatory responses served in other litigation, the
18 specific pirated datasets on which each Llama model was trained. According to Meta, the Books3
19 dataset (which Meta refers to internally as “b3g”) was used to train Llama 1, Llama 2, Llama 3, and
20 Llama 4; the LibGen “Scitech” and LibGen “Fiction” datasets were used to train Llama 3 and Llama
21 4; and the Z-Library dataset, which Meta obtained by torrenting Anna’s Archive, was used to train
22
23

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

26 ⁶⁵ *Kadrey*, ECF No. 485, ¶ 89.

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

1 Llama 4.⁶⁷ Meta has further admitted that, beginning in or about Spring 2023, a Meta engineer
 2 downloaded the Fiction and Scitech branches of LibGen using a direct-download script, and that
 3 Meta’s downloads of both branches “are believed to have been substantially complete.”⁶⁸ In other
 4 words, if a book was in LibGen, Meta downloaded it.

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

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

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

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

19
 20 ⁶⁷ Def. Meta Platforms, Inc.’s Second Supp. Resps. & Objs. to Pl.’s First Set of Interrogs. at 17–18,
 21 *Chicken Soup for the Soul, LLC v. Meta Platforms, Inc.*, No. 3:26-cv-02333-VC (N.D. Cal. Sept. 4,
 22 2026), filed as ECF No. 140-6 (Sept. 12, 2026) (hereinafter “Meta Interrog. Resps.”).

23 ⁶⁸ *Id.* at 12–14.

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

25 ⁷⁰ Meta Interrog. Resps. at 23.

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

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

7 146. Meta used the digital copies it created through Project Vellum to train its “Muse”
8 family of LLMs. In April 2026, Meta introduced the initial version of its Muse Spark model, which
9 Meta made available to the public through certain Meta products and services and, in later versions,
10 to developers through an application programming interface (“API”). In August 2026, Meta released
11 its Muse Glimmer model to the open-source community for use by researchers, developers, and
12 other third parties.⁷³ Meta has admitted that books it acquired and scanned as part of Project Vellum
13 were used to train the Muse Spark family of models.⁷⁴

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

19 148. It had not. The purchase of a physical copy of a book conveys ownership of that
20 particular copy and nothing more. Ownership of a material object in which a work is embodied does
21 not convey any rights in the copyrighted work itself. 17 U.S.C. § 202. And while the first-sale
22 doctrine permits the owner of a lawfully made copy to sell or otherwise dispose of that copy, 17
23
24

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

27 ⁷³ Meta Interrog. Resps. at 23, 25.

28 ⁷⁴ *Id.* at 18 & Ex. D.

1 U.S.C. § 109(a), it confers no right to reproduce the work. The exclusive right to reproduce a
 2 copyrighted work in copies belongs to the copyright owner alone. 17 U.S.C. § 106(1).

3 149. Scanning a book is reproduction. Each time Meta scanned a physical copy of a book,
 4 it created a new digital copy of the entire work. Each time Meta ran the scanned images through
 5 OCR, it created another copy—this time as machine-readable text. Each time Meta processed,
 6 “cleaned,” de-duplicated, and tokenized that text, it created still more copies. And each time Meta
 7 used that text in a training run—including in the “number of internal models” that Meta trains “for
 8 experimental and evaluation purposes” before each public release, and in the “purpose-built
 9 variants” of its LLMs that Meta uses for discrete internal purposes—it created additional copies of
 10 the work.⁷⁵ Meta has admitted that “[a] number of transitory and persistent copies of each of these
 11 works would have been created” in its pipeline and that it “has no feasible way to reliably ascertain
 12 the number of additional persistent or transitory copies” it made.⁷⁶

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

19 151. Meta also launched Project Vellum with full knowledge that a market exists for
 20 licensing books for use in AI training. By mid-2025, Meta had itself explored licensing deals with
 21 publishers, and numerous other AI developers had entered into such deals. Meta chose to purchase
 22 single copies of books rather than pay for training licenses because it was cheaper. The retail price
 23 of a book compensates the author for one reader’s copy; it does not compensate the author for the
 24

25 ⁷⁵ *Id.* at 18.

26 ⁷⁶ *Id.* at 24.

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

1 right to reproduce the work wholesale and to build a commercial product on its expressive content.
2 Project Vellum was designed to bypass the licensing market, and it did.

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

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

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

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

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

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

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

V. CLAIMS FOR RELIEF

COUNT I

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

158. Plaintiffs incorporate the allegations above.

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

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

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

- acquired Plaintiffs’ Works from shadow-library repositories and datasets containing pirated works from shadow libraries;

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

- 1 • distributed Plaintiffs' Works through the use of torrenting software, programs, or
- 2 protocols;
- 3 • reproduced additional copies during ingestion, preprocessing, storage, deduplication,
- 4 formatting, and/or tokenization; and
- 5 • while training its models, made even more copies of the text—because every training
- 6 pass (each epoch and each step of gradient descent) automatically requires creating
- 7 and working with fresh versions of that text.

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

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

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

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

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

24 **COUNT II**

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

26 167. Plaintiffs incorporate the allegations above.

27
 28 ⁷⁹ 17 U.S.C. § 506(a)(1)(A).

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

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

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

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

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

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

28

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

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

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

7 **COUNT III**

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

9 177. Plaintiffs incorporate the allegations above.

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

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

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

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

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

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

2 **COUNT IV**

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

4 184. Plaintiffs incorporate the allegations above.

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

8 186. Upon downloading copyrighted materials, including Plaintiffs' Works, Defendant
9 processed the data, and in doing so, removed and altered certain text and information, including
10 copyright management information found in and on Plaintiffs' Works. When it removed this
11 information, Defendant did so without the authority of the copyright owners.

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

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

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

25 **PRAYER FOR RELIEF**

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

27 a. Judgment in favor of Plaintiffs against the Defendant;

- b. A declaration that the Defendant has infringed Plaintiffs' exclusive copyrights under the Copyright Act;
- c. A declaration that such infringement is willful;
- d. A declaration that Defendant violated 17 U.S.C. § 1202(b) through its removal of copyright management information;
- e. A permanent injunction enjoining the Defendant and all those acting in concert with it from engaging in the infringing conduct alleged herein;
- f. That the Defendant be directed to account to Plaintiffs for all gains, profits, and advantages derived from their unlawful acts;
- g. An award of statutory damages under the Copyright Act;
- h. An award of statutory or actual damages under 17 U.S.C. § 1203(c);
- i. An award of restitution, disgorgement, costs, expenses, and attorneys' fees as permitted by law (including those allowable under 17 U.S.C. § 505 and/or 17 U.S.C. § 1203(b)(4)–(5)).
- j. Pre- and post-judgment interest on the damages awarded to Plaintiffs; and
- k. Further relief for Plaintiffs as the Court may deem just and proper.

JURY TRIAL DEMANDED

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

1 Dated: October 19, 2026

Respectfully submitted,

2 /s/ Elizabeth Brannen

3 Elizabeth Brannen (SBN 226234)

4 Lauren Martin (SBN 294367)

STRIS & MAHER LLP

5 17785 Center Court Dr N, Ste 600

Cerritos, CA 90703

6 T: (213) 995-6800

7 F: (213) 261-0299

ebrannen@stris.com

8 lmartin@stris.com

9 Kyle Roche (*pro hac vice* forthcoming)

FREEDMAN NORMAND FRIEDLAND LLP

10 155 E. 44th Street, Ste 915

New York, NY 10017

11 T: (646) 494-2900

12 kroche@fnf.law

13 *Counsel for Plaintiffs*