

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

DANIEL JOSÉ OLDER, WILLIAM C
AGOSTA, BILLY ARCEMENT, LIBERTY
BARNES, CAROLE BELLACERA, JACOB
BERNSTEIN, KRISTEN BOE, CHASITY
BOWLIN VAN WINKLE, ANGELO
CALVELLO, SARITA CARGAS, GREGG
CARUSO, SINAN CIDDI, MIKE DARLOW,
DEBORAH DEMOSS SMITH, SHAWNA
DOLANSKY, PETROS DRINEAS, JULIE
DZIEWISZ, DANA SUE FELDMAN,
RANDALL GERKE, MELISSA GOOD,
KIMBERLY A. GORDON BIDDLE,
STEPHANIE GRIEST, JOSHUA
HENDRICK, MICHAEL HENNING, LINDA
HOLLEY, DAVID HOOK, CHRIS IMPEY,
DENNIS KEZAR II, HOWARD LAHMAN,
DAVID LAVOND, ERIK LINSTRUM,
SHELLY MATTHEWS, PETER
MCCULLAGH, KEITH MCFARLAND,
SHAKA MCGLOTTEN, JOHN
MCMANAMON, PETER MILONNI, JODI
MORENO, ANDY NORMAN, BOBBY
OWSINSKI, IAIN PARDOE, HARRY
POULOS, TAMMY POWLEY, DOUGLAS
REID, ZANDRIA ROBINSON, LEANDRO
RODRIGUEZ MEDINA, CATHY LISA
SCHNEIDER, STAN SHERER, THOMAS
SPEAR, MARIA STAVRINAKI, RICHARD
ULRICH, ROB WASSMER, PAUL WELLIN,
KATE WHEELER, EBERHARD WOLFF,

Civil Case No.:

COMPLAINT

DEMAND FOR JURY TRIAL

1 XIAO-QIANG ZHAO, and SUSAN
2 ZIMMER,

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 Daniel José Older, William C Agosta, Billy Arcement, Liberty Barnes,
 2 Carole Bellacera, Jacob Bernstein, Kristen Boe, Chasity Bowlin Van Winkle, Angelo Calvello,
 3 Sarita Cargas, Gregg Caruso, Sinan Ciddi, Mike Darlow, Deborah DeMoss Smith, Shawna
 4 Dolansky, Petros Drineas, Julie Dziewisz, Dana Sue Feldman, Randall Gerke, Melissa Good,
 5 Kimberly A. Gordon Biddle, Stephanie Griest, Joshua Hendrick, Michael Henning, Linda Holley,
 6 David Hook, Chris Impey, Dennis Kezar II, Howard Lahman, David Lavond, Erik Linstrum, Shelly
 7 Matthews, Peter McCullagh, Keith McFarland, Shaka McGlotten, John McManamon, Peter
 8 Milonni, Jodi Moreno, Andy Norman, Bobby Owsinski, Iain Pardoe, Harry Poulos, Tammy Powley,
 9 Douglas Reid, Zandria Robinson, Leandro Rodriguez Medina, Cathy Lisa Schneider, Stan Sherer,
 10 Thomas Spear, Maria Stavriniaki, Richard Ulrich, Rob Wassmer, Paul Wellin, Kate Wheeler,
 11 Eberhard Wolff, Xiao-Qiang Zhao, and Susan Zimmer (collectively “Plaintiffs”) bring this action
 12 against Meta Platforms, Inc. (“Meta” or “Defendant”), and allege as follows:

13 **I. INTRODUCTION**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

10 **II. PARTIES**

11 **A. Plaintiffs**

12 17. Plaintiff Daniel José Older is an author who resides in Massachusetts. Daniel José
 13 Older is the author of *Ballad & Dagger*; *Battle Hill Bolero*; *Dactyl Hill Squad*; *Dactyl Hill Squad*
 14 *Book 2: Freedom Fire*; *Dactyl Hill Squad Book Three: Thunder Run*; *Flood City*; *Half-Resurrection*
 15 *Blues*; *Last Canto of the Dead*; *Shadowhouse Fall*; *Shadowshaper*; *Shadowshaper Legacy*; and *The*
 16 *Book of Lost Saints*. One or more of Older’s works are contained in pirated online libraries, including
 17 Books3 (and thus The Pile), Z-Library, and Anna’s Archive. 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 Older’s works. Older’s works
 20 were also published in print and were commercially available for purchase in physical form during
 21 the period in which Meta conducted Project Vellum, as alleged below.

22 18. Plaintiff William C Agosta is an author who resides in Washington. William C
 23 Agosta is the author of *Bombardier Beetles and Fever Trees: A Close-Up Look at Chemical Warfare*
 24 *and Signals in Animals and Plants*; *Chemical Communication: The Language of Pheromones*; and
 25 *Thieves Deceivers and Killers: Tales of Chemistry in Nature*. One or more of Agosta’s works are
 26 contained in pirated online libraries, including LibGen, Z-Library, and Anna’s Archive. Meta has

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

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

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

13 20. Plaintiff Liberty Barnes is an author who resides in Oregon. Liberty Barnes is the
14 author of *Conceiving Masculinity: Male Infertility, Medicine, and Identity*. Barnes's work is
15 contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has
16 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 Barnes's work. Barnes's work was also published in print and was commercially available for
19 purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
20 below.

21 21. Plaintiff Carole Bellacera is an author who resides in South Carolina. Carole
22 Bellacera is the author of *Border Crossings*; *Crossfire*; *East of the Sun, West of the Moon*; *Spotlight*;
23 *Understudy*; and *Velma & Louie*. One or more of Bellacera's works are contained in pirated online
24 libraries, including Z-Library and Anna's Archive. Meta has directly or indirectly downloaded
25 books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is
26 accordingly a reasonable inference that Meta illegally downloaded Bellacera's works. Bellacera's
27 works were also published in print and were commercially available for purchase in physical form
28 during the period in which Meta conducted Project Vellum, as alleged below.

1 22. Plaintiff Jacob Bernstein is an author who resides in California. Jacob Bernstein is
 2 the author of *30 Days to Market Mastery: A Step-by-Step Guide to Profitable Trading*; *All About*
 3 *Day Trading: The Easy Way to Get Started, 1/e*; *Beyond the Investor's Quotient: The Inner World*
 4 *of Investing*; *Commodities, Now Through 1984: A Comprehensive Analysis of the Unprecedented*
 5 *Expectations*; *Cyclic Analysis in Futures Trading: Contemporary Methods and Procedures*; *How to*
 6 *Profit from Seasonal Commodity Spreads: A Complete Guide*; *How to Profit in Precious Metals*;
 7 *Jake Bernstein's Facts on Futures: Insights and Strategies for Winning in the Futures Markets*; *Jake*
 8 *Bernstein's Seasonal Futures Spreads: High-Probability Seasonal Spreads for Futures Traders*;
 9 *Seasonal Concepts in Futures Trading: Turning Seasonality into Profits*; *Seasonality: Systems,*
 10 *Strategies, and Signals*; *Strategies for the Electronic Futures Trader*; *The Compleat Day Trader II*;
 11 *The Compleat Day Trader: Trading Systems, Strategies, Timing Indicators, and Analytical*
 12 *Methods*; *The Compleat Guide to Day Trading Stocks*; *The Handbook of Commodity Cycles: A*
 13 *Window on Time*; *The Handbook of Economic Cycles: Jake Bernstein's Comprehensive Guide to*
 14 *Repetitive Price Patterns in Stocks, Futures, and Financials*; *The Investor's Quotient*; and *The*
 15 *Investor's Quotient the Psychology of Successful Investing in Commodities and Stocks*. One or more
 16 of Bernstein's works are contained in pirated online libraries, including Books3 (and thus The Pile),
 17 LibGen, Z-Library, and Anna's Archive. 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 Bernstein's works. Bernstein's works were also published
 20 in print and were commercially available for purchase in physical form during the period in which
 21 Meta conducted Project Vellum, as alleged below.

22 23. Plaintiff Kristen Boe is an author who resides in Arizona. Kristen Boe is the author
 23 of *Choked Up*; *Shoot 'Em Up*; and *Time's Up*. One or more of Boe's works are contained in pirated
 24 online libraries, including Z-Library. 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 Boe's works. Boe's works were also published in print
 27 and were commercially available for purchase in physical form during the period in which Meta
 28 conducted Project Vellum, as alleged below.

24. Plaintiff Chasity Bowlin Van Winkle is an author who resides in Kentucky. Chasity Bowlin Van Winkle is the author of *A Love So Dark*; *A Midnight Clear*; *A Passion So Strong*; *A Rogue to Remember*; *Barefoot in Hyde Park*; *House of Shadows*; *Hyacinth*; *One Wylde Night*; *Passage of Shadows*; *Tamed by the Lyon*; *The Awakening of Lord Ambrose*; *The Beast of Bath*; *The Enticement of an Earl*; *The First Proposal*; *The Haunting of a Duke*; *The Lady Confesses*; *The Last Offer*; *The Lost Lord of Castle Black*; *The Missing Marquess of Althorn*; *The Mystery of Miss Mason*; *The Pirate's Bluestocking*; *The Redemption of a Rogue*; *The Resurrection of Lady Ramsleigh*; *The Vanishing of Lord Vale*; *The Victorian Gothic Collection*; *Veil of Shadows*; and *What Happens in Piccadilly*. One or more of Bowlin Van Winkle's works are contained in pirated online libraries, including Z-Library and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Bowlin Van Winkle's works. Bowlin Van Winkle's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

25. Plaintiff Angelo Calvello is an author who resides in Illinois. Angelo Calvello is the author of *Environmental Alpha by Angelo A. Calvello*; *Goldberg Securities C B O E Stock Manual*; *Goldberg Securities Runners' Manual*; and *Lived Body and Personal Name*. One or more of Calvello's works are contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Calvello's works. Calvello's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

26. Plaintiff Sarita Cargas is an author who resides in New Mexico. Sarita Cargas is the author of *Encyclopedia of Holocaust Literature* and *Human Rights Education: Forging an Academic Discipline*. One or more of Cargas's works are contained in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable

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

4 27. Plaintiff Gregg Caruso is an author who resides in Connecticut. Gregg Caruso is the
5 author of *Exploring the Illusion of Free Will and Moral Responsibility*; *Free Will and*
6 *Consciousness: A Determinist Account of the Illusion of Free Will*; and *Neuroexistentialism:*
7 *Meaning, Morals, and Purpose in the Age of Neuroscience*. One or more of Caruso's works are
8 contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and
9 Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile,
10 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
11 illegally downloaded Caruso's works. Caruso's works were also published in print and were
12 commercially available for purchase in physical form during the period in which Meta conducted
13 Project Vellum, as alleged below.

14 28. Plaintiff Sinan Ciddi is an author who resides in Virginia. Sinan Ciddi is the author
15 of *Georgetown Journal of International Affairs: Destabilizing Demographics; Summer/Fall 2014*
16 and *Kemalism in Turkish Politics*. One or more of Ciddi's works are contained in pirated online
17 libraries, including LibGen and Z-Library. Meta has directly or indirectly downloaded books
18 illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a
19 reasonable inference that Meta illegally downloaded Ciddi's works. Ciddi's works were also
20 published in print and were commercially available for purchase in physical form during the period
21 in which Meta conducted Project Vellum, as alleged below.

22 29. Plaintiff Mike Darlow is an author who resides in Australia. Mike Darlow is the
23 author of *Know Accounting Graphically*; *Mike Darlow's Woodturning Series: Useful Woodturning*
24 *Projects*; *Woodturning Methods*; and *Woodturning Techniques*. One or more of Darlow's works are
25 contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has
26 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
27 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
28 Darlow's works. Darlow's works were also published in print and were commercially available for

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

3 30. Plaintiff Deborah DeMoss Smith is an author who resides in Louisiana. Deborah
4 DeMoss Smith is the author of *Reflections of the Heart: What Our Animal Companions Tell Us*.
5 DeMoss Smith's work is contained in pirated online libraries, including LibGen and Z-Library. Meta
6 has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
7 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
8 DeMoss Smith's work. DeMoss Smith's work was also published in print and was commercially
9 available for purchase in physical form during the period in which Meta conducted Project Vellum,
10 as alleged below.

11 31. Plaintiff Shawna Dolansky is an author who resides in Canada. Shawna Dolansky is
12 the author of *Now You See It, Now You Don't; Now You See It, Now You Don't: Biblical Perspectives*
13 *on the Relationship Between Magic and Religion; Sacred History, Sacred Literature Essays on*
14 *Ancient Israel, the Bible, and Religion in Honor of R.E.Friedman on His Sixtieth Birthday; and The*
15 *Bible Now*. One or more of Dolansky's works are contained in pirated online libraries, including
16 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
17 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
18 inference that Meta illegally downloaded Dolansky's works. Dolansky's works were also published
19 in print and were commercially available for purchase in physical form during the period in which
20 Meta conducted Project Vellum, as alleged below.

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

28

1 33. Plaintiff Julie Dziewisz is an author who resides in New York. Julie Dziewisz is the
2 author of *Critical Reading, Critical Thinking Focusing on Contemporary Issues - Third Edition* and
3 *Critical Reading, Critical Thinking: Focusing on Contemporary Issues, Fourth Edition*. One or
4 more of Dziewisz's works are contained in pirated online libraries, including LibGen and Z-Library.
5 Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-
6 Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
7 downloaded Dziewisz's works. Dziewisz's works were also published in print and were
8 commercially available for purchase in physical form during the period in which Meta conducted
9 Project Vellum, as alleged below.

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

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

24 36. Plaintiff Melissa Good is an author who resides in Florida. Melissa Good is the author
25 of *Tropical Storm*. Good's work is contained in pirated online libraries, including Z-Library and
26 Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile,
27 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
28 illegally downloaded Good's work. Good's work was also published in print and was commercially

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

3 37. Plaintiff Kimberly A. Gordon Biddle is an author who resides in California.
4 Kimberly A. Gordon Biddle is the author of *Early Childhood Education: Becoming a Professional*.
5 Gordon Biddle's work is contained in pirated online libraries, including LibGen and Z-Library. Meta
6 has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
7 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
8 Gordon Biddle's work. Gordon Biddle's work was also published in print and was commercially
9 available for purchase in physical form during the period in which Meta conducted Project Vellum,
10 as alleged below.

11 38. Plaintiff Stephanie Griest is an author who resides in North Carolina. Stephanie
12 Griest is the author of *100 Places Every Woman Should Go*; *All the Agents and Saints: Dispatches*
13 *from the U.S. Borderlands*; *Around the Bloc: My Life in Moscow, Beijing and Havana*; and *Mexican*
14 *Enough*. One or more of Griest's works are contained in pirated online libraries, including Books3
15 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
16 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
17 is accordingly a reasonable inference that Meta illegally downloaded Griest's works. Griest's works
18 were also published in print and were commercially available for purchase in physical form during
19 the period in which Meta conducted Project Vellum, as alleged below.

20 39. Plaintiff Joshua Hendrick is an author who resides in Maryland. Joshua Hendrick is
21 the author of *Gulen: The Ambiguous Politics of Market Islam in Turkey and the World* and *Making*
22 *the New Middle East: Politics, Culture, and Human Rights*. One or more of Hendrick's works are
23 contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has
24 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
25 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
26 Hendrick's works. Hendrick's works were also published in print and were commercially available
27 for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
28 below.

1 40. Plaintiff Michael Henning is an author who resides in South Africa. Michael Henning
2 is the author of *Total Domination in Graphs*. Henning's work is contained in pirated online libraries,
3 including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally
4 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
5 inference that Meta illegally downloaded Henning's work. Henning's work was also published in
6 print and was commercially available for purchase in physical form during the period in which Meta
7 conducted Project Vellum, as alleged below.

8 41. Plaintiff Linda Holley is an author who resides in Florida. Linda Holley is the author
9 of *Tipis, Tepees, Teepees: History and Design of the Cloth Tipi*. Holley's work is contained in
10 pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's
11 Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen,
12 Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
13 downloaded Holley's work. Holley's work was also published in print and was commercially
14 available for purchase in physical form during the period in which Meta conducted Project Vellum,
15 as alleged below.

16 42. Plaintiff David Hook is an author who resides in Australia. David Hook is the author
17 of *Beginning Cryptography with Java*. Hook's work is contained in pirated online libraries,
18 including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded
19 books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is
20 accordingly a reasonable inference that Meta illegally downloaded Hook's work. Hook's work was
21 also published in print and was commercially available for purchase in physical form during the
22 period in which Meta conducted Project Vellum, as alleged below.

23 43. Plaintiff Chris Impey is an author who resides in Arizona. Chris Impey is the author
24 of *Astronomy—the Cosmic Journey; Astronomy: The Cosmic Journey; Beyond: Our Future in*
25 *Space; Dreams of Other Worlds: The Amazing Story of Unmanned Space Exploration; Encountering*
26 *Life in the Universe: Ethical Foundations and Social Implications of Astrobiology; How It Began:*
27 *A Time-Traveler's Guide to the Universe; How It Ends: From You to the Universe; Humble Before*
28 *the Void: A Western Astronomer, His Journey East, and a Remarkable Encounter Between Western*

1 *Science and Tibetan Buddhism; The Living Cosmos; The Universe Revealed; and Worlds Without*
2 *End: Exoplanets, Habitability, and the Future of Humanity*. One or more of Impey's works are
3 contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and
4 Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile,
5 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
6 illegally downloaded Impey's works. Impey's works were also published in print and were
7 commercially available for purchase in physical form during the period in which Meta conducted
8 Project Vellum, as alleged below.

9 44. Plaintiff Dennis Kezar II is an author who resides in Utah. Dennis Kezar II is the
10 author of *Guilty Creatures: Renaissance Poetry and the Ethics of Authorship* and *Solon and Thespis*.
11 One or more of Kezar's works are contained in pirated online libraries, including LibGen, Z-Library,
12 and Anna's Archive. 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 Kezar's works. Kezar's works were also published in print and were
15 commercially available for purchase in physical form during the period in which Meta conducted
16 Project Vellum, as alleged below.

17 45. Plaintiff Howard Lahman is an author who resides in Massachusetts. Howard
18 Lahman is the author of *Model-Based Development: Applications*. Lahman's work is contained in
19 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 Lahman's work.
22 Lahman's work was also published in print and was commercially available for purchase in physical
23 form during the period in which Meta conducted Project Vellum, as alleged below.

24 46. Plaintiff David Lavond is an author who resides in California. David Lavond is the
25 author of *Handbook of Classical Conditioning*. Lavond's work is contained in pirated online
26 libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
27 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
28 is accordingly a reasonable inference that Meta illegally downloaded Lavond's work. Lavond's

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

3 47. Plaintiff Erik Linstrum is an author who resides in Virginia. Erik Linstrum is the
4 author of *Age of Emergency: Living with Violence at the End of the British Empire; Making Minds*
5 *Modern: The Politics of Psychology in the British Empire, 1898–1970*; and *Ruling Minds:*
6 *Psychology in the British Empire*. One or more of Linstrum’s works are contained in pirated online
7 libraries, including LibGen, Z-Library, and Anna’s Archive. Meta has directly or indirectly
8 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna’s Archive, and there
9 is accordingly a reasonable inference that Meta illegally downloaded Linstrum’s works. Linstrum’s
10 works were also published in print and were commercially available for purchase in physical form
11 during the period in which Meta conducted Project Vellum, as alleged below.

12 48. Plaintiff Shelly Matthews is an author who resides in Illinois. Shelly Matthews is the
13 author of *Feminist Biblical Studies in the Twentieth Century: Scholarship and Movement; First*
14 *Converts: Rich Pagan Women and the Rhetoric of Mission in Early Judaism and Christianity; High-*
15 *Standing Women and Mission and Conversion: A Rhetorical-Historical Analysis of the*
16 *“Antiquities” and Acts; Luke 1–9; Luke 10–24; Perfect Martyr: The Stoning of Stephen and the*
17 *Construction of Christian Identity; The Acts of the Apostles: An Introduction and Study Guide;*
18 *Taming the Tongues of Fire; and Violence in the New Testament*. One or more of Matthews’s works
19 are contained in pirated online libraries, including LibGen, Z-Library, and Anna’s Archive. Meta
20 has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
21 Anna’s Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
22 Matthews’s works. Matthews’s works were also published in print and were commercially available
23 for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
24 below.

25 49. Plaintiff Peter McCullagh is an author who resides in Illinois. Peter McCullagh is the
26 author of *Tensor Methods in Statistics*. McCullagh’s work is contained in pirated online libraries,
27 including LibGen, Z-Library, and Anna’s Archive. Meta has directly or indirectly downloaded
28 books illegally contained in The Pile, LibGen, Z-Library, or Anna’s Archive, and there is

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

4 50. Plaintiff Keith McFarland is an author who resides in Texas. Keith McFarland is the
5 author of *Louis Johnson and the Arming of America: The Roosevelt and Truman Years*; *The 1949*
6 *Revolt of the Admirals*; and *The Korean War: An Annotated Bibliography, Revised Second Edition*.
7 One or more of McFarland's works are contained in pirated online libraries, including LibGen and
8 Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile,
9 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
10 illegally downloaded McFarland's works. McFarland's works were also published in print and were
11 commercially available for purchase in physical form during the period in which Meta conducted
12 Project Vellum, as alleged below.

13 51. Plaintiff Shaka McGlotten is an author who resides in New York. Shaka McGlotten
14 is the author of *A Networked Self and Love*; *Queer Data Studies*; *Queerspaces and Sexpublics*;
15 *Virtual Intimacies: Media, Affect, and Queer Sociality*; and *Zombies and Sexuality: Essays on*
16 *Desire and the Living Dead*. One or more of McGlotten's works are contained in pirated online
17 libraries, including LibGen, Z-Library, and Anna's Archive. 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 McGlotten's works.
20 McGlotten'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 52. Plaintiff John McManamon is an author who resides in Ohio. John McManamon is
23 the author of *Caligula's Barges and the Renaissance Origins of Nautical Archaeology Under Water*;
24 *Funeral Oratory and the Cultural Ideals of Italian Humanism*; *Pierpaolo Vergerio the Elder and*
25 *Saint Jerome*; *Pierpaolo Vergerio the Elder: The Humanist as Orator*; and *Ut Crescat Laudata*
26 *Virtus: Funeral Oratory and the Culture*. One or more of McManamon's works are contained in
27 pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or
28 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,

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

5 53. Plaintiff Peter Milonni is an author who resides in New Mexico. Peter Milonni is the
6 author of *Laser Physics*; *Lasers*; and *The Quantum Vacuum: An Introd. to Quantum*
7 *Electrodynamics*. One or more of Milonni's works are contained in pirated online libraries,
8 including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally
9 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
10 inference that Meta illegally downloaded Milonni's works. Milonni's works were also published in
11 print and were commercially available for purchase in physical form during the period in which
12 Meta conducted Project Vellum, as alleged below.

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

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

27 56. Plaintiff Bobby Owsinski is an author who resides in California. Bobby Owsinski is
28 the author of *Hollywood Institute of Communication Art & Sciences*; *Mixing and Mastering with IK*

1 *Multimedia T-Racks: The Official Guide, 1e; Mixing Engineer's Handbook; Nova Interactive*
 2 *Multimedia Studies Curriculum; The Mastering Engineer's Handbook; The Mixing Engineer's*
 3 *Handbook; The Recording Engineer's Handbook; and The Recording Engineer's Handbook, 3e.*
 4 One or more of Owsinski's works are contained in pirated online libraries, including Books3 (and
 5 thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded
 6 books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is
 7 accordingly a reasonable inference that Meta illegally downloaded Owsinski's works. Owsinski's
 8 works were also published in print and were commercially available for purchase in physical form
 9 during the period in which Meta conducted Project Vellum, as alleged below.

10 57. Plaintiff Iain Pardoe is an author who resides in Canada. Iain Pardoe is the author of
 11 *Applied Regression Modeling, 2nd Edition* and *Applied Regression Modeling: A Business*
 12 *Approach*. One or more of Pardoe's works are contained in pirated online libraries, including
 13 LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The
 14 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
 15 Meta illegally downloaded Pardoe's works. Pardoe's works were also published in print and were
 16 commercially available for purchase in physical form during the period in which Meta conducted
 17 Project Vellum, as alleged below.

18 58. Plaintiff Harry Poulos is an author who resides in Australia. Harry Poulos is the
 19 author of *Pile Foundation Analysis and Design* and *Tall Building Foundation Design*. One or more
 20 of Poulos's works are contained in pirated online libraries, including Books3 (and thus The Pile),
 21 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
 22 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
 23 inference that Meta illegally downloaded Poulos's works. Poulos's works were also published in
 24 print and were commercially available for purchase in physical form during the period in which
 25 Meta conducted Project Vellum, as alleged below.

26 59. Plaintiff Tammy Powley is an author who resides in Florida. Tammy Powley is the
 27 author of *Jewelry & Beading Designs for Dummies; Making Designer Mixed-Media and Memory*
 28 *Jewelry: Fun and Experimental Techniques and Materials for the Home Studio; Making Designer*

1 *Seed Bead, Stone, and Crystal Jewelry; Memory-Craft; Picture Yourself Creating Metal Clay*
2 *Jewelry: Step-by-Step Instruction for Forming, Firing, and Constructing Finished Jewelry, 1st Ed;*
3 *Retelling Jane Austen: Essays on Recent Adaptations and Derivative Works; and Women of Florida*
4 *Fiction: Essays on 12 Sunshine State Writers.* One or more of Powley's works are contained in
5 pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's
6 Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen,
7 Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
8 downloaded Powley's works. Powley's works were also published in print and were commercially
9 available for purchase in physical form during the period in which Meta conducted Project Vellum,
10 as alleged below.

11 60. Plaintiff Douglas Reid is an author who resides in Maryland. Douglas Reid is the
12 author of *Intrepid Sailors: The Legacy of Preble's Boys and the Tripoli Campaign; Lion in the Bay:*
13 *The British Invasion of the Chesapeake, 1813-14; and To the Walls of Derne: William Eaton, the*
14 *Tripoli Coup, and the End of the First Barbary War.* One or more of Reid's works are contained in
15 pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's
16 Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen,
17 Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
18 downloaded Reid's works. Reid'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 61. Plaintiff Zandria Robinson is an author who resides in Washington, D.C.. Zandria
22 Robinson is the author of *Chocolate Cities: The Black Map of American Life; Repositioning Race:*
23 *Prophetic Research in a Postracial Obama Age; and This Ain't Chicago: Race, Class, and Regional*
24 *Identity in the Post-Soul South.* One or more of Robinson's works are contained in pirated online
25 libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has
26 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
27 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
28 Robinson's works. Robinson's works were also published in print and were commercially available

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

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

11 63. Plaintiff Cathy Lisa Schneider is an author who resides in Washington, D.C.. Cathy
12 Lisa Schneider is the author of *Collective Violence, Contentious Politics, and Social Change: A*
13 *Charles Tilly Reader; Police Power and Race Riots: Urban Unrest in Paris and New York;*
14 *Shantytown Protest in Pinochet's Chile;* and *The Mobilization at the Grassroots: Shantytowns and*
15 *Resistance In*. One or more of Schneider's works are contained in pirated online libraries, including
16 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
17 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
18 inference that Meta illegally downloaded Schneider's works. Schneider's works were also published
19 in print and were commercially available for purchase in physical form during the period in which
20 Meta conducted Project Vellum, as alleged below.

21 64. Plaintiff Stan Sherer is an author who resides in Massachusetts. Stan Sherer is the
22 author of *Founding Farms: Portraits of Five Massachusetts Family Farms* and *Long Life to Your*
23 *Children: A Portrait of High Albania*. One or more of Sherer's works are contained in pirated online
24 libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
25 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
26 is accordingly a reasonable inference that Meta illegally downloaded Sherer's works. Sherer's works
27 were also published in print and were commercially available for purchase in physical form during
28 the period in which Meta conducted Project Vellum, as alleged below.

1 65. Plaintiff Thomas Spear is an author who resides in Maine. Thomas Spear is the author
2 of *Being Maasai: Ethnicity & Identity in East Africa*; *East African Expressions of Christianity*;
3 *Kenya's Past: An Introduction to Historical Method in Africa*; *Mountain Farmers: Moral*
4 *Economies of Land & Agricultural Development in Arusha & Meru*; *The Oxford Encyclopedia of*
5 *African Historiography: Methods and Sources (2 Volume Set)*; *The Swahili: Reconstructing the*
6 *History and Language of an African Society, 800-1500*; and *Traditions of Origin and Their*
7 *Interpretation: The Mijikenda of Kenya*. One or more of Spear's works are contained in pirated
8 online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
9 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
10 is accordingly a reasonable inference that Meta illegally downloaded Spear's works. Spear's works
11 were also published in print and were commercially available for purchase in physical form during
12 the period in which Meta conducted Project Vellum, as alleged below.

13 66. Plaintiff Maria Stavrinaki is an author who resides in France. Maria Stavrinaki is the
14 author of *DADA PRESENTISM, an Essay on Art & History*, by Maria Stavrinaki, Translated by
15 Daniela Ginsburg and *Power and Time: Temporalities in Conflict and the Making of History*. One
16 or more of Stavrinaki's works are contained in pirated online libraries, including LibGen, Z-Library,
17 and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The
18 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
19 Meta illegally downloaded Stavrinaki's works. Stavrinaki's works were also published in print and
20 were commercially available for purchase in physical form during the period in which Meta
21 conducted Project Vellum, as alleged below.

22 67. Plaintiff Richard Ulrich is an author who resides in Arkansas. Richard Ulrich is the
23 author of *Advanced Electronic Packaging* and *Integrated Passive Component Technology*. One or
24 more of Ulrich's works are contained in pirated online libraries, including LibGen, Z-Library, and
25 Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile,
26 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
27 illegally downloaded Ulrich's works. Ulrich'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 68. Plaintiff Rob Wassmer is an author who resides in California. Rob Wassmer is the
4 author of *Applied Policy Research: Concepts and Cases, Second Edition*; *Bidding for Business: The*
5 *Efficacy of Local Economic Development Incentives in a Metropolitan Area*; *Taxes, Property Tax*
6 *Abatement, Expenditure, and the Composition Of*; and *The Benefits of Growth: A ULI Working*
7 *Paper on Land Use Policy*. One or more of Wassmer's works are contained in pirated online
8 libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
9 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
10 is accordingly a reasonable inference that Meta illegally downloaded Wassmer's works. Wassmer's
11 works were also published in print and were commercially available for purchase in physical form
12 during the period in which Meta conducted Project Vellum, as alleged below.

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

20 70. Plaintiff Kate Wheeler is an author who resides in Massachusetts. Kate Wheeler is
21 the author of *Not Where I Started From* and *When Mountains Walked*. One or more of Wheeler's
22 works are contained in pirated online libraries, including Anna's Archive. Meta has directly or
23 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
24 and there is accordingly a reasonable inference that Meta illegally downloaded Wheeler's works.
25 Wheeler's works were also published in print and were commercially available for purchase in
26 physical form during the period in which Meta conducted Project Vellum, as alleged below.

27 71. Plaintiff Eberhard Wolff is an author who resides in Germany. Eberhard Wolff is the
28 author of *A Practical Guide to Continuous Delivery* and *Microservices: Flexible Software*

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

7 72. Plaintiff Xiao-Qiang Zhao is an author who resides in Canada. Xiao-Qiang Zhao is
8 the author of *Dynamical Systems in Population Biology* and *Fields Institute Communications:*
9 *Nonlinear Dynamics and Evolution Equations: Vol. 48*. One or more of Zhao's works are contained
10 in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly
11 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
12 is accordingly a reasonable inference that Meta illegally downloaded Zhao's works. Zhao's works
13 were also published in print and were commercially available for purchase in physical form during
14 the period in which Meta conducted Project Vellum, as alleged below.

15 73. Plaintiff Susan Zimmer is an author who resides in Canada. Susan Zimmer is the
16 author of *I Love Coffee!*. Zimmer's work is contained in pirated online libraries, including Books3
17 (and thus The Pile) and Anna's Archive. 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 Zimmer's work. Zimmer's work was also published in
20 print and was commercially available for purchase in physical form during the period in which Meta
21 conducted Project Vellum, as alleged below.

22 74. A non-exhaustive list of registered copyrights owned by Plaintiffs is included as
23 Exhibit A (herein, the "Plaintiffs' Works").³

24
25
26
27 ³ 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 **B. Defendant**

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

9 **III. JURISDICTION AND VENUE**

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

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

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

25 **IV. FACTUAL ALLEGATIONS**

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

27 79. Plaintiffs' Works are published books. Each was conceived, researched, written, and
28 revised by its author over weeks, months, or years, and each was published and offered for sale to

1 the reading public. Whatever their genre or subject matter, the works embody their authors’ original
2 creative choices—of structure, organization, selection, emphasis, tone, and expression—and
3 constitute precisely the kind of structured, long-form, polished prose that, as alleged below, Meta
4 and other LLM developers prize as training data. Collectively, the works reflect substantial original
5 authorship intended to inform, educate, entertain, and engage readers.

6 80. Plaintiffs have either legal or beneficial ownership of the rights to reproduce, adapt,
7 and distribute Plaintiffs’ Works.

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

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

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

20 83. “Generative artificial intelligence” or “generative AI” refers to systems and models
21 that create outputs—such as text or images—that simulate human expression, often in response to
22 user prompts.

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

85. Development of LLMs requires, among other things, “training” the model on data—*i.e.*, the model’s inputs. Training “requires identifying a formal measure or ‘objective’ for how well the model performs, and then repeatedly adjusting the model’s parameters based on that objective as the model is exposed to training data.”⁴ While not all LLM developers use the same terminology, developers commonly reference a “pre-training” process—in which a massive amount of computing power and data is spent to teach the model the broad foundations of language, grammar, and reasoning—and a “post-training” or “fine-tuning” process “where the pre-trained model is further trained on a (relative to pre-training) smaller amount of carefully curated data of specific tasks.”⁵ For purposes of this Complaint, “training” refers to all stages of the LLM training process.

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

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

⁴ U.S. Copyright Office, *Copyright and Artificial Intelligence, Part 3: Generative AI Training Pre-Publication*, at 17 (May 2025), <https://perma.cc/EY5U-EFUY>.

⁵ *Id.*

⁶ James Betker, *The “it” in AI Models is the Dataset*, (June 10, 2023), <https://perma.cc/ZCH9-H53S>.

1 88. In determining the corpus of training data for an LLM, a model’s developer typically
2 considers “the quantity of data, its quality, and the ultimate purpose(s) of the model.”⁷

3 89. In terms of data quantity, the LLMs that have been developed and are being
4 developed by leading technology companies are trained on enormous datasets, typically on the scale
5 of terabytes. This is because AI researchers have found that “increasing the quantity of training data
6 typically increases a model’s ‘performance.’”⁸

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

14 91. That books make for particularly valuable training material for LLMs is evidenced
15 by the fact that several—if not most—of the major LLM developers either have paid for or expressly
16 contemplated paying (large sums) for licenses to use copyrighted works to train their LLMs. Indeed,
17 in an *amicus* brief filed in *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-3417 (N.D. Cal.), the
18 Association of American Publishers cited *dozens* of “AI licensing deals for textual works,” including
19 deals involving Amazon, Microsoft, OpenAI, and Perplexity.¹⁰ And, as the Court found in that
20
21
22

23 _____
24 ⁷ U.S. Copyright Office, *Copyright and Artificial Intelligence, Part 3: Generative AI Training Pre-Publication*, at 9.

25 ⁸ *Id.* at 10.

26 ⁹ *Id.* at 11.

27 ¹⁰ *Kadrey v. Meta Platforms, Inc.*, ECF No. 535 at 12, Case No. 3:23-cv-3417-VC (N.D. Cal.) (Apr.
28 11, 2025).

1 litigation, Meta considered spending up to \$100 million to license copyrighted materials in
2 connection with LLM training.¹¹

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

6 93. As discussed in the introduction, technology companies have treated generative AI
7 as the next foundational layer of the digital economy. Industry leaders publicly describe an “AI arms
8 race,” in which they have redirected their corporate strategies to seize control of what they believe
9 will become a new infrastructure layer for commerce, communication, and knowledge work.¹² For
10 these companies, staying ahead of competitors is “code red.”¹³ Among other things, being too slow
11 out of the blocks could mean ending up in last place; if by the time you publicly released *your* LLM
12 model, the public writ large and private industry had already adopted *other companies’* models,
13 there would be a real risk that your model would be left on the shelf.

14 94. Like other participants in this race, Meta risked falling behind early industry leaders
15 like OpenAI. Indeed, market analysts consistently reported that Meta had, in fact, fallen behind,
16 with many pointing to Meta’s intense focus on investment in the “Metaverse.”¹⁴

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

21 ¹¹ *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1040 (N.D. Cal. 2025).

22 ¹² See Dr. Peter Asaro, *What is an ‘Artificial Intelligence Arms Race’ Anyway?*, 15 I/S: J.L. & Pol’y
23 for Info. Soc’y 45 (2019).

24 ¹³ See Sharon Goldman, *Sam Altman declares ‘Code Red’ as Google’s Gemini surges—three years*
25 *after ChatGPT caused Google CEO Sundar Pichai to do the same*, FORTUNE (Dec. 2, 2025, 11:43
AM), <https://perma.cc/J9MS-UQD2>.

26 ¹⁴ Mike Proulx, *Meta: So Long, Metaverse; Hello, Superintelligence*, FORRESTER (July 30, 2025),
27 <https://perma.cc/BZ57-CR2D>; Karen Hao, et al., *Mark Zuckerberg Was Early in AI. Now Meta Is*
28 *Trying to Catch Up.*, WALL STREET JOURNAL (June 17, 2023, 12:00 AM), <https://perma.cc/G548-B5RD>.

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

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

9 97. These libraries are widely known within both piracy communities and the technology
 10 sector as illegal sources of copyrighted books. Many have been the subject of criminal prosecutions,
 11 civil injunctions, domain seizures, and formal designation as “notorious piracy markets” by United
 12 States trade authorities. As centralized shadow libraries increasingly faced enforcement actions,
 13 including the seizure of domains, third parties responded by creating full mirrored copies of those
 14 repositories for decentralized redistribution.¹⁷

15 98. In 2018, an OpenAI employee downloaded pirated copies of books from Library
 16 Genesis, or “LibGen”—a shadow library repeatedly enjoined by federal courts¹⁸—and used those
 17 books to create two internal datasets OpenAI called “LibGen1” and “LibGen2,” which OpenAI
 18
 19

20 ¹⁵ See Letter from Mary E. Rasenberger, CEO, Authors Guild, & Umair Kazi, Dir. of Pol’y &
 21 Advocacy, Authors Guild, to Daniel Lee, Assistant U.S. Trade Representative for Innovation &
 22 Intellectual Prop., Office of the U.S. Trade Representative (Oct. 7, 2022), <https://perma.cc/XM4R-NDN3>.

23 ¹⁶ See Riddhi Setty, *Rampant ‘Shadow Libraries’ Drive Calls for Anti-Piracy Action*, BLOOMBERG
 24 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>.

25 ¹⁷ Woodcock, *‘Shadow Libraries’ Are Moving Their Pirated Books to the Dark Web After Fed Crackdowns*.
 26

27 ¹⁸ See *Cengage Learning, Inc. et al. v. Does 1-50 d/b/a Library Genesis*, ECF No. 36, Case No. 23-
 28 cv-8136 (S.D.N.Y. Sept. 24, 2024).

publicly referred to as “Books1” and “Books2.”¹⁹ 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.

99. Within weeks of GPT-3’s release, an open research collective, EleutherAI, formed with the express goal of replicating GPT-3’s capabilities and democratizing access to large-scale language modeling. To do so, EleutherAI assembled and publicly released *The Pile*, a large (800+ gigabyte) general-purpose training dataset expressly designed to be downloaded, used, and incorporated into LLMs by academic researchers, startups, and commercial AI developers.²⁰

100. Because OpenAI did not disclose the precise makeup of its training datasets, members of EleutherAI constructed a book corpus of their own: “Books3,” consisting of approximately 196,640 books.²¹ Books3 comprises about 12 percent (just over 100 gigabytes) of *The Pile*.²²

101. EleutherAI and the compiler of Books3, Shawn Presser, have confirmed the genesis of Books3 is the shadow library Bibliotik. Presser has publicly stated that Books3 represents “all of bibliotik,”²³ and an EleutherAI paper likewise confirms that Books3 “is a dataset of books derived from a copy of the contents of the Bibliotik private tracker.”²⁴

102. Bibliotik is a private, invitation-only torrent tracker that has long functioned as a centralized source of pirated e-books. The repository, which hosts and distributes hundreds of

¹⁹ *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>.

²⁰ Leo Gao et al., *The Pile: An 800GB Dataset of Diverse Text for Language Modeling*, arXiv, 1 (2020), <https://perma.cc/NHV6-R8YE>.

²¹ Stella Biderman et al., *Datasheet for the Pile*, arXiv, 8 (2020), <https://perma.cc/7KL2-LTLF>.

²² *Id.*

²³ Shawn Presser, X (Oct. 25, 2020 1:32 AM), <https://perma.cc/7WRD-NHRX>.

²⁴ Biderman et al., *Datasheet for the Pile*, arXiv, 8.

1 thousands of copyrighted books, is only accessible with the use of torrenting protocols, which are
2 explained below.²⁵

3 103. Bibliotik has been widely recognized in piracy communities, academic literature, and
4 AI-research documentation as a shadow library devoted to copyrighted books. Unsurprisingly, then,
5 *The Pile*'s datasheet acknowledges that "Books3 is almost entirely comprised of copyrighted
6 works."²⁶ Bibliotik's illicit nature is no secret—it has been openly discussed for years prior to major
7 tech companies' use of datasets derived directly from it.

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

10 [Presser] created Books3 in the hope that it would allow any developer to create
11 generative-AI tools. "It would be better if it wasn't necessary to have something like
12 Books3," he said. "But the alternative is that, without Books3, only OpenAI can do
13 what they're doing."²⁷

14 105. EleutherAI's compilation and distribution of *The Pile* and Books3 provided "off-the-
15 shelf" access to a corpus of infringing works that any AI developer could download and immediately
16 incorporate into a large-scale training pipeline. As a result, Books3's presence within *The Pile*
17 facilitated wide downstream distribution and adoption of a corpus derived from a pirate book
18 library.²⁸

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

21
22
23 ²⁵ Ruheni Mathenge, *The 12 Best Private Torrent Sites Still Working in 2026*, PRIVACYSAVVY (last
24 accessed March 9, 2026), <https://perma.cc/4V8M-3ALY>.

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

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

27 ²⁸ See Stella Biderman, *The Pile: An 800GB Dataset of Diverse Text for Language Modeling*,
28 ELEUTHERAI (Dec. 31, 2020), <https://perma.cc/JGT9-LLTK>.

107. As noted above, OpenAI’s developers trained its LLMs on a books corpus derived from LibGen, one of the largest and longest-running shadow libraries in the world. LibGen hosts millions of pirated books, academic texts, and scholarly articles.²⁹ It operates as a centralized repository offering direct downloads of full-fidelity e-book files. It also distributes its entire database through bulk archives and torrent files, enabling third parties to download and locally host complete copies of its collection.³⁰ LibGen has been repeatedly enjoined by federal courts for copyright infringement and has been designated a “notorious market” by the United States Trade Representative.³¹ Despite enforcement actions, LibGen has remained accessible through shifting domains, mirrors, and downloadable archives. Its persistence is a function of deliberate decentralization designed to evade shutdown.³²

108. Z-Library (also known as “B-ok”) is another well-known shadow library that emerged as an expanded and user-friendly derivative of LibGen. It incorporated large portions of LibGen’s catalog while adding additional titles, metadata, and interface features.³³ Z-Library offered premium features—including faster downloads and higher volume limits—in exchange for payment, operating in effect as a commercial piracy service.³⁴ In 2022, Z-Library’s domains were

²⁹ Office of the U.S. Trade Representative, REVIEW OF NOTORIOUS MARKETS FOR COUNTERFEITING AND PIRACY, 27 (2024), <https://perma.cc/N4WT-AHJ9> (“Libgen ... hosts a large number of digital copies of books, manuals, journals, and other works, many of which are unauthorized copies of copyright protected content.”).

³⁰ See Letter from Mary E. Rasenberger, CEO, Authors Guild, & Umair Kazi, Dir. of Pol’y & Advocacy, Authors Guild, to Daniel Lee, Assistant U.S. Trade Representative for Innovation & Intellectual Prop., Office of the U.S. Trade Representative at n.5.

³¹ See Office of the U.S. Trade Representative, 2019 REVIEW OF NOTORIOUS MARKETS FOR COUNTERFEITING AND PIRACY, 27, <https://perma.cc/22VN-9VD7>.

³² See Andrew Albanese, *Textbook Publishers Sue Notorious ‘Shadow Library’ Libgen*, PUBLISHERS WEEKLY (Sep. 14, 2023), <https://perma.cc/3NPY-UJCX>.

³³ Jordana Rosenfeld, *Z-Library*, ENCYCLOPEDIA BRITANNICA (last accessed March 9, 2026), <https://perma.cc/4H26-GDCC>.

³⁴ See Masood Farivar, *Two Russian Nationals Charged With Operating E-Book Piracy Site*, VOA (Nov. 16, 2022), <https://perma.cc/6MPD-QNKB>.

1 seized by law-enforcement authorities, and its operators were arrested and later indicted for criminal
 2 copyright infringement.³⁵ These actions confirmed what had long been publicly known: Z-Library
 3 was an illegal piracy operation. The seizure of Z-Library did not eliminate access to its content.
 4 Instead, third parties responded by creating full mirrors of its collection to ensure continued
 5 distribution.³⁶

6 109. Another major player in the shadow library ecosystem is Anna’s Archive. Anna’s
 7 Archive is the most comprehensive and active shadow library currently in operation.³⁷

8 110. Anna’s Archive began in 2022 as “Pirate Library Mirror,” initially hosting a mirrored
 9 copy of Z-Library. It later rebranded as “Anna’s Archive” and expanded to aggregate and host the
 10 complete collections of LibGen, Z-Library, and other pirated sources.³⁸ Anna’s Archive functions
 11 as a meta-library: it indexes, mirrors, and redistributes multiple shadow libraries simultaneously,
 12 offering users unified access to millions of pirated books.³⁹

13 111. Anna’s Archive offers paid tiers that provide high-speed or priority access to its
 14 pirated collections.⁴⁰ Through its downloadable archives and torrent-based distribution, Anna’s
 15 Archive enables users to acquire and store local copies of millions of copyrighted books in bulk.⁴¹

16
 17 ³⁵ Press Release, U.S. Dep’t of Justice, U.S. Att’y’s Off., E. Dist. of N.Y., *Two Russian Nationals*
 18 *Charged with Running Massive E-Book Piracy Website* (Nov. 16, 2022), <https://perma.cc/CR4L-JLA3>.

19 ³⁶ See Woodcock, *‘Shadow Libraries’ Are Moving Their Pirated Books to The Dark Web*.

20 ³⁷ 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 ³⁸ 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>.

25 ³⁹ *Id.*

26 ⁴⁰ See *If you’re an LLM, please read this*, ANNA’S ARCHIVE (February 18, 2026),
 27 <https://perma.cc/989X-GS3Q>.

28 ⁴¹ See M. Luisa Simpson, *2024 Special 301 Out-of-Cycle Review of Notorious Markets: Request*
 for Comments ASSOCIATION OF AMERICAN PUBLISHERS (OCTOBER 2, 2024), <https://perma.cc/P2E9-MYBY>.

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

3 113. According to Anna’s Archive, “virtually all major companies building LLMs
4 contacted us to train on our data. . . We have given high-speed access to about 30
5 companies.”⁴² Anna’s Archive blog stated as recently as February 18, 2026 that if an *LLM was*
6 *reading its blog* “you have likely been trained in part on our data.”⁴³

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

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

20 115. As noted above, Meta develops and distributes the Llama series of LLMs, including
21 Llama 1, Llama 2, Llama 3, and Llama 4. Meta has “invested billions of dollars to develop its
22 generative AI offerings,”⁴⁴ and it in turn expects Llama to generate somewhere in the realm of \$460

23
24 ⁴² See *Copyright reform is necessary for national security*, ANNA’S ARCHIVE (Jan. 31, 2025),
25 <https://perma.cc/3RVZ-6G5B>.

26 ⁴³ See *If you’re an LLM, please read this*, ANNA’S ARCHIVE (Feb. 18, 2026), <https://perma.cc/MJ7Z-3ZCL>.

27 ⁴⁴ *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.⁴⁵ None of the money that Meta invested into the development of
 2 its Llama models was used to pay for Plaintiffs' copyrighted works.⁴⁶

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

4 117. 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.⁴⁷
 7 To accomplish that, Meta needed to train on large quantities of high-quality books and other
 8 published materials.

9 118. 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."⁴⁸

12 119. 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."⁴⁹

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

18 121. 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 ⁴⁵ See *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1040 (N.D. Cal. 2025).

23 ⁴⁶ *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 ⁴⁷ 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 ⁴⁸ *Kadrey*, 788 F. Supp. 3d at 1040.

28 ⁴⁹ *Id.*

1 122. In connection with Llama 1’s release, Meta disclosed the composition of the model’s
 2 “pre-training data.”⁵⁰ Among the datasets used to train Llama 1 were “two book corpora,” including
 3 “the Books3 section of ThePile.”⁵¹ 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.”⁵² This distinction
 6 was deliberate and notable, as Meta *knew* that Books3 was *not* a corpus of books in the “public
 7 domain.”

8 123. 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.⁵³ “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.”⁵⁴ The C4 dataset itself is derived from CommonCrawl.⁵⁵ In short, a massive amount (and
 14 the majority) of data used by Meta to train Llama was “scraped” from the Internet.

15 124. 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.⁵⁶ But
 17 this idea was short-lived. “[I]n spring 2023, after failing to acquire licenses and following escalation
 18

19 ⁵⁰ Hugo Touvron et al., *LLaMA: Open and Efficient Foundation Language Models*, arXiv, 2 (2023),
 20 <https://perma.cc/LB97-ZK2C>.

21 ⁵¹ *Id.*

22 ⁵² *Id.*

23 ⁵³ Touvron et al., *LLaMA: Open and Efficient Foundation Language Models*, arXiv, 2.

24 ⁵⁴ 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 ⁵⁵ *Id.*

27 ⁵⁶ *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.”⁵⁷

3 125. 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.

7 126. “In early 2024, Meta also downloaded Anna’s Archive, a compilation of shadow
8 libraries including LibGen, Z-Library, and others.”⁵⁸

9 127. Meta likewise downloaded copyrighted materials from Z-Library, Sci-Hub, and
10 other shadow libraries.⁵⁹ Indeed, by April 2024, Meta had downloaded over 25 TB of data from Z-
11 Library and 10 TB of data from LibGen.⁶⁰

12 128. To obtain the massive quantities of data from these shadow libraries, Meta used
13 torrenting protocols.⁶¹ 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.”⁶²

16 129. 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 ⁵⁷ *Kadrey*, 788 F. Supp. 3d at 1041.

21 ⁵⁸ *Id.*

22 ⁵⁹ *Cambronne*, ECF No. 109 (Meta Answer), ¶ 83.

23 ⁶⁰ *Kadrey*, ECF No. 472 (Pls.’ Partial Mot. for Summ. J.), at 11.

24 ⁶¹ *Cambronne*, ECF No. 109 at ¶ 83 (admitting that “it used BitTorrent to download certain portions
25 of these publicly available datasets”).

26 ⁶² *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.

3 130. 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.

7 131. 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.

12 132. 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.

19 133. 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.

22 134. 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.

24 135. 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.

27 136. 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.

4 137. As Meta has admitted elsewhere, Meta “created a script to remove repetitive text
5 from the training data for its Llama models.”⁶³ As admitted elsewhere, Meta’s “program was
6 designed to remove lines that contain the word ‘copyright.’”⁶⁴

7 138. 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.

11 139. 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.⁶⁵ 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.

17 140. 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 ⁶³ *Kadrey*, ECF No. 485, ¶ 88 (Meta’s Answer to Third Am. Compl.).

26 ⁶⁴ *Kadrey*, ECF No. 485, ¶ 89.

27 ⁶⁵ *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 4.⁶⁶ 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.”⁶⁷ In other
 4 words, if a book was in LibGen, Meta downloaded it.

5 141. 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.⁶⁸

7 142. 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**

14 143. 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.⁶⁹ Meta
 17 internally referred to this book-acquisition and scanning operation as “Project Vellum,” and it is
 18 referred to as such herein.⁷⁰

19
 20 ⁶⁶ 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 ⁶⁷ *Id.* at 12–14.

24 ⁶⁸ *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1014 (N.D. Cal. 2025) (Anthropic’s amassing of
 “a central library”).

25 ⁶⁹ Meta Interrog. Resps. at 23.

26 ⁷⁰ *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>.

1 144. 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.⁷¹

7 145. 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.⁷² 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.⁷³

14 146. 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.

19 147. 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 ⁷¹ 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 ⁷² Meta Interrog. Resps. at 23, 25.

28 ⁷³ *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).

3 148. 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.⁷⁴ 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.⁷⁵

13 149. 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.”⁷⁶ In other words, Meta scanned each book’s copyright page,
 18 ingested the copyright notice, and trained on the work anyway.

19 150. 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 ⁷⁴ *Id.* at 18.

26 ⁷⁵ *Id.* at 24.

27 ⁷⁶ *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.

3 151. 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.

6 152. 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**

13 153. 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.

19 154. *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.

155. *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.”⁷⁷

156. *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

157. Plaintiffs incorporate the allegations above.

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

159. 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.

160. 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;

⁷⁷ 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.

8 161. 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.

11 162. 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.⁷⁸

18 163. 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.

20 164. 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.

23 165. 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**

26 166. Plaintiffs incorporate the allegations above.

27
 28 ⁷⁸ 17 U.S.C. § 506(a)(1)(A).

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

3 168. 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).

10 169. 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.

15 170. 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.

21 171. 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.

24 172. 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.

1 173. 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.

3 174. 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.

6 175. 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)**

9 176. Plaintiffs incorporate the allegations above.

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

12 178. 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.

16 179. 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.

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

26 181. 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.

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

COUNT IV

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

183. Plaintiffs incorporate the allegations above.

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

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

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

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

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

PRAYER FOR RELIEF

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

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

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

JURY TRIAL DEMANDED

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

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*