

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

PAUL LEVINE, NITTANY VALLEY
PRODUCTIONS, INC., ABRAHAM
ACOSTA, RICHARD ANTARAMIAN,
CARL BAUER, JOSEPH BUENKER,
KATHLEEN CIOFFI, ANGIE CRUZ, KYLE
CUORDILEONE, RENEE DIRESTA, ADAM
DORIN, RIC DRAGON, GABRIELE
GALIMBERTI, CHARLES GIVRE, NICOLE
GONZALEZ VAN CLEVE, MARK
GOODACRE, MARK GOODALE, MITCHE
GRAF, ANDREW HAGERMAN, BARRY
HALLIWELL, PETE HAVEL, KERRI
HAWKINS, ARI HERSTAND, LEON
HIGLEY, LUKE HOHMANN, RAY
JACKENDOFF, JASON ALAN
JANKOVSKY, MARNEY KEENAN, MARK
KOLTKO-RIVERA, RAEF LAWSON, ROB
LONG, JOHN LURZ, THOMAS LYS,
GUADALUPE GARCIA MCCALL,
SHANNON MCELROY, MARSHALL KIRK
MCKUSICK, EMILY MENDENHALL,
MIKE MURPHEY, MIKE MURPHEY
BOOKS, LLC, DAVID NELSON, PETROS
NICOPOLITIDIS, JOHN NOYES, MICHAEL
OBERMAN, HANS OCHS, ROBERT
PASNAK, ERROL REISS, SEAN RICHEY,
RICK ROBINSON, INGRID ROJAS
CONTRERAS, WILLIAM FRANCIS
SHADWICK, JOEL SIMON, G. KEITH
STILL, ROB TANNENBAUM, NICK

Civil Case No.:

COMPLAINT

DEMAND FOR JURY TRIAL

1 THIEBERGER, NEIL VAN NIEKERK,
2 KIRSTEN VERNON, VENDELA VIDA,
3 DAVID WEISS, ERICA WEITZMAN, LIZ
WILSON, and ROMANUS WOLTER,

4 Plaintiffs,

5 v.

6 META PLATFORMS, INC.,

7 Defendant.

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

1 1. Plaintiffs Paul Levine, Nittany Valley Productions, Inc., Abraham Acosta, Richard
 2 Antaramian, Carl Bauer, Joseph Buenker, Kathleen Cioffi, Angie Cruz, Kyle Cuordileone, Renee
 3 DiResta, Adam Dorin, Ric Dragon, Gabriele Galimberti, Charles Givre, Nicole Gonzalez Van
 4 Cleve, Mark Goodacre, Mark Goodale, Mitche Graf, Andrew Hagerman, Barry Halliwell, Pete
 5 Havel, Kerri Hawkins, Ari Herstand, Leon Higley, Luke Hohmann, Ray Jackendoff, Jason Alan
 6 Jankovsky, Marney Keenan, Mark Koltko-Rivera, Raef Lawson, Rob Long, John Lurz, Thomas
 7 Lys, Guadalupe Garcia McCall, Shannon McElroy, Marshall Kirk McKusick, Emily Mendenhall,
 8 Mike Murphey, Mike Murphey Books, LLC, David Nelson, Petros Nicopolitidis, John Noyes,
 9 Michael Oberman, Hans Ochs, Robert Pasnak, Errol Reiss, Sean Richey, Rick Robinson, Ingrid
 10 Rojas Contreras, William Francis Shadwick, Joel Simon, G. Keith Still, Rob Tannenbaum, Nick
 11 Thieberger, Neil van Niekerk, Kirsten Vernon, Vendela Vida, David Weiss, Erica Weitzman, Liz
 12 Wilson, and Romanus Wolter (collectively “Plaintiffs”) bring this action against Meta Platforms,
 13 Inc. (“Meta” or “Defendant”), and allege as follows:

14 **I. INTRODUCTION**

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

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 Paul Levine is an author who resides in California. Plaintiff Nittany Valley
 13 Productions, Inc. is a California corporation owned and/or controlled by Paul Levine. Collectively,
 14 Paul Levine and Nittany Valley Productions, Inc. own *9 Scorpions*; *Bum Deal*; *Bum Luck*; *Bum Rap*;
 15 *Cheater’s Game*; *Early Grave*; *False Dawn*; *Flesh and Bones: A Jake Lassiter Novel*; *Fool Me*
 16 *Twice: A Jake Lassiter Novel*; *Illegal*; *Lassiter: A Novel*; *Last Chance Lassiter*; *Midnight Burning*;
 17 *Mortal Sin*; *Night Vision*; *Paydirt*; *Slashback: A Jake Lassiter Novel*; *Solomon vs. Lord*; *State vs.*
 18 *Lassiter*; *The Road to Hell*; *To Speak for the Dead*; and *What’s Your Verdict?*. One or more of these
 19 works are contained in pirated online libraries, including Z-Library and Anna’s Archive. Meta has
 20 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 these works. These works were also published in print and were commercially available for purchase
 23 in physical form during the period in which Meta conducted Project Vellum, as alleged below.

24 18. Plaintiff Abraham Acosta is an author who resides in Arizona. Abraham Acosta is
 25 the author of *Thresholds of Illiteracy: Theory, Latin America, and the Crisis of Resistance*. Acosta’s
 26 work is contained in pirated online libraries, including LibGen, Z-Library, and Anna’s Archive.

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

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

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

14 20. Plaintiff Carl Bauer is an author who resides in Arizona. Carl Bauer is the author of
15 *Against the Current: Privatization, Water Markets, and the State in Chile; Against the Current?:*
16 *Privatization, Markets, and the State in Water Rights, Chile, 1979-1993; and Siren Song: Chilean*
17 *Water Law as a Model for International Reform*. One or more of Bauer's works are contained in
18 pirated online libraries, including LibGen and Z-Library. 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 Bauer's works. Bauer's works
21 were also published in print and were commercially available for purchase in physical form during
22 the period in which Meta conducted Project Vellum, as alleged below.

23 21. Plaintiff Joseph Buenker is an author who resides in Arizona. Joseph Buenker is the
24 author of *Encyclopedia of the Gilded Age and Progressive Era*. Buenker's work is contained in
25 pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or
26 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
27 and there is accordingly a reasonable inference that Meta illegally downloaded Buenker's work.

28

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

3 22. Plaintiff Kathleen Cioffi is an author who resides in New Jersey. Kathleen Cioffi is
4 the author of *Alternative Theatre in Poland, 1954-1989* and *Theatermachine: Tadeusz Kantor in*
5 *Context*. One or more of Cioffi's works are contained in pirated online libraries, including LibGen
6 and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile,
7 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
8 illegally downloaded Cioffi's works. Cioffi's works were also published in print and were
9 commercially available for purchase in physical form during the period in which Meta conducted
10 Project Vellum, as alleged below.

11 23. Plaintiff Angie Cruz is an author who resides in New York. Angie Cruz is the author
12 of *Angelica and La Guira*; *Daring to Write: Contemporary Narratives by Dominican Women*;
13 *Dominicana*; *How Not to Drown in a Glass of Water*; *Let It Rain Coffee*; *Love Rehearsals*; and
14 *Soledad*. One or more of Cruz's works are contained in pirated online libraries, including LibGen,
15 Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
16 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
17 inference that Meta illegally downloaded Cruz's works. Cruz's works were also published in print
18 and were commercially available for purchase in physical form during the period in which Meta
19 conducted Project Vellum, as alleged below.

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

27 25. Plaintiff Renee DiResta is an author who resides in Virginia. Renee DiResta is the
28 author of *Invisible Rulers* and *The Hardware Startup*. One or more of DiResta's works are contained

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

6 26. Plaintiff Adam Dorin is an author who resides in California. Adam Dorin is the
7 author of *Jihad and American Medicine: Thinking Like a Terrorist to Anticipate Attacks via Our*
8 *Health System*. Dorin's work is contained in pirated online libraries, including LibGen and Z-
9 Library. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen,
10 Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
11 downloaded Dorin's work. Dorin's work was also published in print and was commercially
12 available for purchase in physical form during the period in which Meta conducted Project Vellum,
13 as alleged below.

14 27. Plaintiff Ric Dragon is an author who resides in Colombia. Ric Dragon is the author
15 of *Social Marketology: Improve Your Social Media Processes and Get Customers to Stay Forever*.
16 Dragon's work is contained in pirated online libraries, including LibGen and Z-Library. Meta has
17 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
18 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
19 Dragon's work. Dragon's work was also published in print and was commercially available for
20 purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
21 below.

22 28. Plaintiff Gabriele Galimberti is an author who resides in Italy. Gabriele Galimberti
23 is the author of *In Her Kitchen: Stories and Recipes from Grandmas Around the World*; *My Couch*
24 *is Your Couch: Exploring How People Live Around the World*; and *Toy Stories*. One or more of
25 Galimberti's works are contained in pirated online libraries, including Books3 (and thus The Pile),
26 LibGen, and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The
27 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
28 Meta illegally downloaded Galimberti's works. Galimberti's works were also published in print and

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

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

10 30. Plaintiff Nicole Gonzalez Van Cleve is an author who resides in Rhode Island. Nicole
11 Gonzalez Van Cleve is the author of *Crook County: Racism and Injustice in America's Largest*
12 *Criminal Court* and *The Waiting Room*. One or more of Gonzalez Van Cleve's works are contained
13 in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or
14 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
15 and there is accordingly a reasonable inference that Meta illegally downloaded Gonzalez Van
16 Cleve's works. Gonzalez Van Cleve's works were also published in print and were commercially
17 available for purchase in physical form during the period in which Meta conducted Project Vellum,
18 as alleged below.

19 31. Plaintiff Mark Goodacre is an author who resides in North Carolina. Mark Goodacre
20 is the author of *The Case Against Q: Studies in Markan Priority and the Synoptic Problem*; *The*
21 *Fourth Synoptic Gospel*; *Thomas and the Gospels*; and *The Routledge Encyclopedia of Ancient*
22 *Mediterranean Religions*. One or more of Goodacre's works are contained in pirated online libraries,
23 including LibGen and Z-Library. 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 Goodacre's works. Goodacre's works were also published
26 in print and were commercially available for purchase in physical form during the period in which
27 Meta conducted Project Vellum, as alleged below.

32. Plaintiff Mark Goodale is an author who resides in Switzerland. Mark Goodale is the author of *A Revolution in Fragments: Traversing Scales of Justice, Ideology, and Practice in Bolivia*; *Anthropology and Law: A Critical Introduction*; *Dilemmas of Modernity, Bolivian Encounters with Law and Liberalism*³; *Human Rights at the Crossroads*; *Human Rights Encounter*; *Legal Pluralism: Normative and Empirical Approaches*; *Letters to the Contrary: A Curated History of the UNESCO Human Rights Survey*; *Neoliberalism, Interrupted: Social Change and Contested Governance in Contemporary Latin America*; *Practicing Ethnography in Law: New Dialogues, Enduring Methods*; *Reinventing Human Rights*; *Surrendering to Utopia: An Anthropology of Human Rights*; and *The Bolivia Reader*. One or more of Goodale's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Goodale's works. Goodale's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

33. Plaintiff Mitche Graf is an author who resides in Oregon. Mitche Graf is the author of *Power Marketing for Wedding and Portrait Photographers* and *Power Marketing, Selling, and Pricing: A Business Guide for Wedding and Portrait Photographers, 2nd Edition*. One or more of Graf's works are contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Graf's works. Graf's works were also published in print and were commercially available for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged below.

34. Plaintiff Andrew Hagerman is an author who resides in Japan. Andrew Hagerman is the author of *Pro Tools 10 Ignite!: The Visual Guide for New Users, 1e*; *Pro Tools 11 Ignite! The*

³ This title reflects an error included in the U.S. Copyright Office registration data; the correct title is *Dilemmas of Modernity, Bolivian Encounters with Law and Liberalism*.

1 *Visual Guide for New Users, 1e; Pro Tools 9 Ignite!, 1e; Pro Tools LE 7 Ignite!: The Visual Guide*
2 *for New Users, 2nd Ed; Pro Tools LE 8 Ignite!: The Visual Guide for New Users, 1e; Pro Tools LE7*
3 *Ignite!: The Visual Guide for New Users; and Working with Beats in Pro Tools: Skill Pack.* One or
4 more of Hagerman's works are contained in pirated online libraries, including LibGen, Z-Library,
5 and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The
6 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
7 Meta illegally downloaded Hagerman's works. Hagerman's works were also published in print and
8 were commercially available for purchase in physical form during the period in which Meta
9 conducted Project Vellum, as alleged below.

10 35. Plaintiff Barry Halliwell is an author who resides in Singapore. Barry Halliwell is
11 the author of *DNA and Free Radicals* and *Herbal and Traditional Medicine: Molecular Aspects of*
12 *Health*. One or more of Halliwell's works are contained in pirated online libraries, including
13 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
14 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
15 inference that Meta illegally downloaded Halliwell's works. Halliwell's works were also published
16 in print and were commercially available for purchase in physical form during the period in which
17 Meta conducted Project Vellum, as alleged below.

18 36. Plaintiff Pete Havel is an author who resides in Texas. Pete Havel is the author of
19 *The Arsonist in the Office: Fireproofing Your Life Against Toxic Coworkers, Bosses, Employees,*
20 *and Cultures*. Havel's work is contained in pirated online libraries, including LibGen, Z-Library,
21 and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The
22 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
23 Meta illegally downloaded Havel's work. Havel's work was also published in print and was
24 commercially available for purchase in physical form during the period in which Meta conducted
25 Project Vellum, as alleged below.

1 37. Plaintiff Kerri Hawkins is an author who resides in California. Kerri Hawkins is the
2 author of *Blood Legacy*⁴ and *The House of Alexander*. One or more of Hawkins's works are
3 contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has
4 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
5 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
6 Hawkins's works. Hawkins's works were also published in print and were commercially available
7 for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
8 below.

9 38. Plaintiff Ari Herstand is an author who resides in California. Ari Herstand is the
10 author of *How to Make It in the New Music Business: Practical Tips on Building a Loyal Following*
11 *and Making a Living as a Musician*. Herstand's work is contained in pirated online libraries,
12 including Books3 (and thus The Pile), LibGen, and 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 Herstand's work. Herstand's
15 work was also published in print and was commercially available for purchase in physical form
16 during the period in which Meta conducted Project Vellum, as alleged below.

17 39. Plaintiff Leon Higley is an author who resides in Nebraska. Leon Higley is the author
18 of *Biotic Stress and Yield Loss; Economic Thresholds for Integrated Pest Management; Manual of*
19 *Entomology and Pest Management; and The Ecology of Place: Contributions of Place-Based*
20 *Research to Ecological Understanding*. One or more of Higley's works are contained in pirated
21 online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
22 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
23 is accordingly a reasonable inference that Meta illegally downloaded Higley's works. Higley's
24 works were also published in print and were commercially available for purchase in physical form
25 during the period in which Meta conducted Project Vellum, as alleged below.

26
27 _____
28 ⁴ This title reflects an error included in the U.S. Copyright Office registration data; the correct title
is *Blood Legacy: The Story of Ryan*.

1 40. Plaintiff Luke Hohmann is an author who resides in California. Luke Hohmann is
2 the author of *Beyond Software Architecture: Creating and Sustaining Winning Solutions; Innovation*
3 *Games: Creating Breakthrough Products Through Collaborative Play*; and *Journey of the Software*
4 *Professional: A Sociology of Software Development*. One or more of Hohmann's works are
5 contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has
6 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 Hohmann's works. Hohmann's works were also published in print and were commercially available
9 for purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
10 below.

11 41. Plaintiff Ray Jackendoff is an author who resides in Massachusetts. Ray Jackendoff
12 is the author of *A Generative Theory of Tonal Music; Consciousness and the Computational Mind;*
13 *Language, Consciousness, Culture: Essays on Mental Structure; Language, Logic, and Concepts:*
14 *Essays in Memory of John Macnamara; Languages of the Mind: Essays on Mental Representation;*
15 *Patterns in the Mind: Language and Human Nature; Semantic Structures; Semantics and Cognition;*
16 and *The Architecture of the Language Faculty*. One or more of Jackendoff's works are contained in
17 pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or
18 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
19 and there is accordingly a reasonable inference that Meta illegally downloaded Jackendoff's works.
20 Jackendoff'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 42. Plaintiff Jason Alan Jankovsky is an author who resides in Florida. Jason Alan
23 Jankovsky is the author of *The Art of the Trade: What I Learned (and Lost) Trading the Chicago*
24 *Futures Markets; Time Compression Trading; and Trading Rules That Work: The 28 Essential*
25 *Lessons Every Trader Must Master*. One or more of Jankovsky's works are contained in pirated
26 online 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 Jankovsky's works.

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

3 43. Plaintiff Marney Keenan is an author who resides in Michigan. Marney Keenan is
4 the author of *The Snow Killings: Inside the Oakland County Child Killer Investigation*. Keenan's
5 work is contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive.
6 Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-
7 Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
8 downloaded Keenan's work. Keenan'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 44. Plaintiff Mark Koltko-Rivera is an author who resides in Virginia. Mark Koltko-
12 Rivera is the author of *Cracking Codes & Cryptograms for Dummies* and *The Worldview*
13 *Assessment Instrument (WAI): The Development and Preliminary Validation of an Instrument to*
14 *Assess Worldview Components Relevant to Counseling and Psychotherapy*. One or more of Koltko-
15 Rivera's works are contained in pirated online libraries, including LibGen and Anna's Archive.
16 Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-
17 Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
18 downloaded Koltko-Rivera's works. Koltko-Rivera's works were also published in print and were
19 commercially available for purchase in physical form during the period in which Meta conducted
20 Project Vellum, as alleged below.

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

1 46. Plaintiff Rob Long is an author who resides in New Jersey. Rob Long is the author
2 of *Bigly: Donald Trump in Verse* and *Conversations with My Agent*. One or more of Long's works
3 are contained in pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library,
4 and Anna's Archive. Meta has directly or indirectly downloaded books illegally contained in The
5 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
6 Meta illegally downloaded Long's works. Long'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 47. Plaintiff John Lurz is an author who resides in Massachusetts. John Lurz is the author
10 of *The Death of the Book: Modernist Novels and the Time of Reading*. Lurz's work is contained in
11 pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or
12 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
13 and there is accordingly a reasonable inference that Meta illegally downloaded Lurz's work. Lurz's
14 work was also published in print and was commercially available for purchase in physical form
15 during the period in which Meta conducted Project Vellum, as alleged below.

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

23 49. Plaintiff Guadalupe Garcia McCall is an author who resides in Texas. Guadalupe
24 Garcia McCall is the author of "*Ofrendas*"; *All the Stars Denied*; *Echoes of Grace*; *Secret of the*
25 *Moon Conch*; *Shame the Stars*; *Summer of the Mariposas*; and *Under the Mesquite*. One or more of
26 McCall's works are contained in pirated online libraries, including Books3 (and thus The Pile),
27 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
28 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable

1 inference that Meta illegally downloaded McCall's works. McCall'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 50. Plaintiff Shannon McElroy is an author who resides in Alabama. Shannon McElroy
5 is the author of *Ramblings from a Captured Heart: 30 Day Devotional*. McElroy's work is contained
6 in pirated online libraries, including Z-Library. Meta has directly or indirectly downloaded books
7 illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a
8 reasonable inference that Meta illegally downloaded McElroy's work. McElroy's work was also
9 published in print and was commercially available for purchase in physical form during the period
10 in which Meta conducted Project Vellum, as alleged below.

11 51. Plaintiff Marshall Kirk McKusick is an author who resides in California. Marshall
12 Kirk McKusick is the author of *Register Allocation and Data Conversion in Machine; The Design*
13 *and Implementation of the 4.3BSD UNIX Operating System; The Design and Implementation of the*
14 *4.3BSD Unix Operating System: Answer Book; The Design and Implementation of the 4.4BSD*
15 *Operating System; The Design and Implementation of the FreeBSD Operating System; and The*
16 *Design and Implementation of the FreeBSD(R) Operating System, 2/E*. McKusick's works are
17 contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has
18 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
19 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
20 McKusick's works. McKusick's works were also published in print and were commercially
21 available for purchase in physical form during the period in which Meta conducted Project Vellum,
22 as alleged below.

23 52. Plaintiff Emily Mendenhall is an author who resides in Maryland. Emily Mendenhall
24 is the author of *Environmental Health Narratives: A Reader for Youth; Global Health Narratives:*
25 *A Reader for Youth; Rethinking Diabetes: Entanglements with Trauma, Poverty, and HIV; Syndemic*
26
27
28

1 *Suffering*⁵; *The VIDDA Syndemic: Distress and Diabetes in Social and Cultural Context*; and
2 *Unmasked: COVID, Community, and the Case of Okoboji*. One or more of Mendenhall's works are
3 contained in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has
4 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
5 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
6 Mendenhall's works. Mendenhall's works were also published in print and were commercially
7 available for purchase in physical form during the period in which Meta conducted Project Vellum,
8 as alleged below.

9 53. Plaintiff Mike Murphey is an author who resides in Washington. Plaintiff Mike
10 Murphey Books, LLC is a Washington entity owned and/or controlled by Mike Murphey.
11 Collectively, Mike Murphey and Mike Murphey Books, LLC own *Killing Time: A Tale of Physics*,
12 *Lust and Greed Book 3; Taking Time... a Tale of Physics, Lust and Greed; The Conman*; and *The*
13 *Outlaw Gillis Kerg ... Physics, Lust and Greed Series, Book 4*. One or more of these works are
14 contained in pirated online libraries, including Z-Library. Meta has directly or indirectly
15 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
16 is accordingly a reasonable inference that Meta illegally downloaded these works. These works were
17 also published in print and were commercially available for purchase in physical form during the
18 period in which Meta conducted Project Vellum, as alleged below.

19 54. Plaintiff David Nelson is an author who resides in Massachusetts. David Nelson is
20 the author of *Solkattu Manual: An Introduction to the Rhythmic Language of South Indian Music*.
21 Nelson's work is contained in pirated online libraries, including LibGen and Z-Library. Meta has
22 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
23 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
24 Nelson's work. Nelson's work was also published in print and was commercially available for
25

26
27 ⁵ This title reflects an error included in the U.S. Copyright Office registration data; the correct title
28 is *Syndemic Suffering: Social Distress, Depression, and Diabetes among Mexican Immigrant Women*.

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

3 55. Plaintiff Petros Nicopolitidis is an author who resides in Greece. Petros Nicopolitidis
4 is the author of *Modeling and Simulation of Computer Networks and Systems: Methodologies and*
5 *Applications, First Edition*. Nicopolitidis's work is contained in pirated online libraries, including
6 LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The
7 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
8 Meta illegally downloaded Nicopolitidis's work. Nicopolitidis's work was also published in print
9 and was commercially available for purchase in physical form during the period in which Meta
10 conducted Project Vellum, as alleged below.

11 56. Plaintiff John Noyes is an author who resides in Canada. John Noyes is the author of
12 *Literature and Cartography: Theories, Histories, Genres* and *The Mastery of Submission:*
13 *Inventions of Masochism*. One or more of Noyes's works are contained in pirated online libraries,
14 including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly
15 or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's
16 Archive, and there is accordingly a reasonable inference that Meta illegally downloaded Noyes's
17 works. Noyes's works were also published in print and were commercially available for purchase
18 in physical form during the period in which Meta conducted Project Vellum, as alleged below.

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

26 58. Plaintiff Hans Ochs is an author who resides in Washington. Hans Ochs is the author
27 of *Primary Immunodeficiency Diseases: A Molecular & Cellular Approach, 2nd Ed*; *Primary*
28 *Immunodeficiency Diseases: A Molecular and Genetic Approach*; and *Primary Immunodeficiency*

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

8 59. Plaintiff Robert Pasnak is an author who resides in Virginia. Robert Pasnak is the
9 author of *Emerging Themes in Cognitive Development: Vol. 1, Foundations*; *Emerging Themes in*
10 *Cognitive Development: Vol. 2, Competencies*; *Piacceleration Concept Manual*; and *Simplified*
11 *Statistics*. One or more of Pasnak's works are contained in pirated online libraries, including LibGen
12 and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The Pile,
13 LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta
14 illegally downloaded Pasnak's works. Pasnak'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 60. Plaintiff Errol Reiss is an author who resides in Georgia. Errol Reiss is the author of
18 *Fundamental Medical Mycology* and *Molecular Immunology of Mycotic and Actinomycotic*
19 *Infections*. One or more of Reiss's works are contained in pirated online libraries, including LibGen,
20 Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
21 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
22 inference that Meta illegally downloaded Reiss's works. Reiss's works were also published in print
23 and were commercially available for purchase in physical form during the period in which Meta
24 conducted Project Vellum, as alleged below.

25 61. Plaintiff Sean Richey is an author who resides in Georgia. Sean Richey is the author
26 of *Collateral Damage: The Influence of Political Rhetoric on the Incorporation of Second-*
27 *Generation Americans*; *Google and Democracy: Politics and the Power of the Internet*; *Social*
28 *Networks and Japanese Democracy: The Beneficial Impact of Interpersonal Communication in East*

1 *Asia; The Social Basis of the Rational Citizen: How Political Communication in Social Networks*
2 *Improves Civic Competence; and Voting by Mail: Institutional Reform and Political Participation*
3 *in Oregon*. One or more of Richey's works are contained in pirated online libraries, including
4 LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded books illegally
5 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
6 inference that Meta illegally downloaded Richey's works. Richey's works were also published in
7 print and were commercially available for purchase in physical form during the period in which
8 Meta conducted Project Vellum, as alleged below.

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

16 63. Plaintiff Ingrid Rojas Contreras is an author who resides in California. Ingrid Rojas
17 Contreras is the author of *Fruit of the Drunken Tree: A Novel* and *The Man Who Could Move Clouds*.
18 One or more of Rojas Contreras's works are contained in pirated online libraries, including Books3
19 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly
20 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
21 is accordingly a reasonable inference that Meta illegally downloaded Rojas Contreras's works.
22 Rojas Contreras's works were also published in print and were commercially available for purchase
23 in physical form during the period in which Meta conducted Project Vellum, as alleged below.

24 64. Plaintiff William Francis Shadwick is an author who resides in Brazil. William
25 Francis Shadwick is the author of *Baecklund Transformations and Their Applications; Local Jet*
26 *Bundle Formulation of Baecklund Transformations: With Applications to Non-linear Evolution*
27 *Equations; and Mechanics Day*. One or more of Shadwick's works are contained in pirated online
28 libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly

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

5 65. Plaintiff Joel Simon is an author who resides in New York. Joel Simon is the author
6 of *Solution Focused Practice in End-of-Life & Grief Counseling* and *Solution-Focused Case*
7 *Management*. One or more of Simon's works are contained in pirated online libraries, including
8 LibGen, Z-Library, and Anna's Archive. 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 Simon's works. Simon'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 66. Plaintiff G. Keith Still is an author who resides in the United Kingdom. G. Keith Still
14 is the author of *Introduction to Crowd Science*. Still's work is contained in pirated online libraries,
15 including LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally
16 contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable
17 inference that Meta illegally downloaded Still's work. Still's work was also published in print and
18 was commercially available for purchase in physical form during the period in which Meta
19 conducted Project Vellum, as alleged below.

20 67. Plaintiff Rob Tannenbaum is an author who resides in New York. Rob Tannenbaum
21 is the author of *I WANT MY MTV: The Uncensored Story of the Music Video Revolution*.
22 Tannenbaum's work is contained in pirated online libraries, including Books3 (and thus The Pile),
23 LibGen, and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The
24 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
25 Meta illegally downloaded Tannenbaum's work. Tannenbaum's work was also published in print
26 and was commercially available for purchase in physical form during the period in which Meta
27 conducted Project Vellum, as alleged below.

28

1 68. Plaintiff Nick Thieberger is an author who resides in Australia. Nick Thieberger is
2 the author of *A Dictionary of Nafsan, South Efate, Vanuatu: Mpet Nafsan Ni Erakor* and *A Grammar*
3 *of South Efate: An Oceanic Language of Vanuatu*. One or more of Thieberger's works are contained
4 in pirated online libraries, including LibGen and Z-Library. Meta has directly or indirectly
5 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
6 is accordingly a reasonable inference that Meta illegally downloaded Thieberger's works.
7 Thieberger's works were also published in print and were commercially available for purchase in
8 physical form during the period in which Meta conducted Project Vellum, as alleged below.

9 69. Plaintiff Neil van Niekerk is an author who resides in New Jersey. Neil van Niekerk
10 is the author of *Direction & Quality of Light: Your Key to Better Portrait Photography Anywhere;*
11 *Off-Camera Flash Techniques for Digital Photographers;* and *On-Camera Flash Techniques for*
12 *Digital Wedding and Portrait Photography*. One or more of Van Niekerk's works are contained in
13 pirated online libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's
14 Archive. Meta has directly or indirectly downloaded books illegally contained in The Pile, LibGen,
15 Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that Meta illegally
16 downloaded Van Niekerk's works. Van Niekerk's works were also published in print and were
17 commercially available for purchase in physical form during the period in which Meta conducted
18 Project Vellum, as alleged below.

19 70. Plaintiff Kirsten Vernon is an author who resides in New Hampshire. Kirsten Vernon
20 is the author of *Career Distinction: Stand Out by Building Your Brand*. Vernon's work is contained
21 in pirated online libraries, including Books3 (and thus The Pile), LibGen, and Z-Library. Meta has
22 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
23 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
24 Vernon's work. Vernon's work was also published in print and was commercially available for
25 purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
26 below.

27 71. Plaintiff Vendela Vida is an author who resides in California. Vendela Vida is the
28 author of *And Now You Can Go: A Novel; Girls on the Verge: Debutante Dips, Gang Drive-Bys,*

1 *and Other Initiations; Let the Northern Lights Erase Your Name; The Diver's Clothes Lie Empty;*
2 *The Lovers; and We Run the Tides.* One or more of Vida's works are contained in pirated online
3 libraries, including Books3 (and thus The Pile), LibGen, Z-Library, and Anna's Archive. Meta has
4 directly or indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or
5 Anna's Archive, and there is accordingly a reasonable inference that Meta illegally downloaded
6 Vida's works. Vida's works were also published in print and were commercially available for
7 purchase in physical form during the period in which Meta conducted Project Vellum, as alleged
8 below.

9 72. Plaintiff David Weiss is an author who resides in California. David Weiss is the
10 author of *A Science of Decision Making: The Legacy of Ward Edwards; Analysis of Variance and*
11 *Functional Measurement: A Practical Guide; and Analysis of Variance and Functional*
12 *Measurement: A Practical Guide, Includes CD-ROM.* One or more of Weiss's works are contained
13 in pirated online libraries, including LibGen, Z-Library, and Anna's Archive. Meta has directly or
14 indirectly downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive,
15 and there is accordingly a reasonable inference that Meta illegally downloaded Weiss's works.
16 Weiss's works were also published in print and were commercially available for purchase in
17 physical form during the period in which Meta conducted Project Vellum, as alleged below.

18 73. Plaintiff Erica Weitzman is an author who resides in Illinois. Erica Weitzman is the
19 author of *At the Limit of the Obscene: German Realism and the Disgrace of Matter; Irony's Antics:*
20 *Walser, Kafka, Roth, and the German Comic Tradition; and Returns of Laughter: Comic Irony in*
21 *Walser, Kafka, and Roth.* One or more of Weitzman's works are contained in pirated online libraries,
22 including LibGen, Z-Library, and Anna's Archive. Meta has directly or indirectly downloaded
23 books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there is
24 accordingly a reasonable inference that Meta illegally downloaded Weitzman's works. Weitzman's
25 works were also published in print and were commercially available for purchase in physical form
26 during the period in which Meta conducted Project Vellum, as alleged below.

27 74. Plaintiff Liz Wilson is an author who resides in Ohio. Liz Wilson is the author of
28 *Charming Cadavers: Horrific Figurations of the Feminine in Indian Buddhist Hagiographic*

1 *Literature; Family in Buddhism; and The Living and the Dead: Social Dimensions of Death in South*
 2 *Asian Religions*. One or more of Wilson's works are contained in pirated online libraries, including
 3 LibGen and Z-Library. Meta has directly or indirectly downloaded books illegally contained in The
 4 Pile, LibGen, Z-Library, or Anna's Archive, and there is accordingly a reasonable inference that
 5 Meta illegally downloaded Wilson's works. Wilson's works were also published in print and were
 6 commercially available for purchase in physical form during the period in which Meta conducted
 7 Project Vellum, as alleged below.

8 75. Plaintiff Romanus Wolter is an author who resides in Florida. Romanus Wolter is the
 9 author of *Kick Start Your Dream Business: Getting It Started and Keeping You Going; Kick Start*
 10 *Your Success; and Trunkus the Traveling Mousyphant: In Washington, D. C.* Wolter's works are
 11 contained in pirated online libraries, including Z-Library. Meta has directly or indirectly
 12 downloaded books illegally contained in The Pile, LibGen, Z-Library, or Anna's Archive, and there
 13 is accordingly a reasonable inference that Meta illegally downloaded Wolter's works. Wolter's
 14 works were also published in print and were commercially available for purchase in physical form
 15 during the period in which Meta conducted Project Vellum, as alleged below.

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

18 **B. Defendant**

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

26
 27

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

1 **III. JURISDICTION AND VENUE**

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

5 79. This Court has personal jurisdiction over the Defendant. The Defendant has
6 purposefully availed itself of the privilege of conducting business in this District and the State of
7 California. The Defendant committed acts of copyright infringement in this District, directed
8 conduct toward this District, or knowingly caused harm that was suffered in this District. The
9 Defendant maintains substantial, continuous, and systematic contacts with this District.

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

17 **IV. FACTUAL ALLEGATIONS**

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

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

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

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

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

B. LLMs and the Generative AI "Arms Race"

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

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

87. Development of LLMs requires, among other things, "training" the model on data—*i.e.*, the model's inputs. Training "requires identifying a formal measure or 'objective' for how well the model performs, and then repeatedly adjusting the model's parameters based on that objective as the model is exposed to training data."⁷ While not all LLM developers use the same terminology,

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

1 developers commonly reference a “pre-training” process—“in which a massive amount of
2 computing power and data is spent to teach the model the broad foundations of language, grammar,
3 and reasoning”—and a “post-training” or “fine-tuning” process “where the pre-trained model is
4 further trained on a (relative to pre-training) smaller amount of carefully curated data of specific
5 tasks.”⁸ For purposes of this Complaint, “training” refers to all stages of the LLM training process.

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

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

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

22 91. In terms of data quantity, the LLMs that have been developed and are being
23 developed by leading technology companies are trained on enormous datasets, typically on the scale
24

25 ⁸ *Id.*

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

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

1 of terabytes. This is because AI researchers have found that “increasing the quantity of training data
2 typically increases a model’s ‘performance.’”¹¹

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

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

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

21 95. As discussed in the introduction, technology companies have treated generative AI
22 as the next foundational layer of the digital economy. Industry leaders publicly describe an “AI arms
23

24 ¹¹ *Id.* at 10.

25 ¹² *Id.* at 11.

26 ¹³ *Kadrey v. Meta Platforms, Inc.*, ECF No. 535 at 12, Case No. 3:23-cv-3417-VC (N.D. Cal.) (Apr.
27 11, 2025).

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

1 race,” in which they have redirected their corporate strategies to seize control of what they believe
 2 will become a new infrastructure layer for commerce, communication, and knowledge work.¹⁵ For
 3 these companies, staying ahead of competitors is “code red.”¹⁶ Among other things, being too slow
 4 out of the blocks could mean ending up in last place; if by the time you publicly released *your* LLM
 5 model, the public writ large and private industry had already adopted *other companies’* models,
 6 there would be a real risk that your model would be left on the shelf.

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

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

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

14 98. For years now, there have been illicit, online marketplaces for digital copies of books.
 15 These “shadow libraries,” as they are commonly known, systematically acquire, store, index, and
 16 disseminate full-fidelity digital copies of copyrighted books. These online repositories maintain
 17 searchable catalogs, metadata, mirrors, bulk-download mechanisms, and—critically—complete
 18 downloadable archives designed to allow third parties to replicate the entire collection.¹⁸ And they

19
 20
 21 ¹⁵ See Dr. Peter Asaro, *What is an ‘Artificial Intelligence Arms Race’ Anyway?*, 15 I/S: J.L. & Pol’y
 for Info. Soc’y 45 (2019).

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

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

27 ¹⁸ See Letter from Mary E. Rasenberger, CEO, Authors Guild, & Umair Kazi, Dir. of Pol’y &
 28 Advocacy, Authors Guild, to Daniel Lee, Assistant U.S. Trade Representative for Innovation &

1 do all of this without authorization from authors or publishers.¹⁹ In short, shadow libraries facilitate
 2 the reproduction and distribution of unauthorized copies of copyrighted works at industrial scale.

3 99. These libraries are widely known within both piracy communities and the technology
 4 sector as illegal sources of copyrighted books. Many have been the subject of criminal prosecutions,
 5 civil injunctions, domain seizures, and formal designation as “notorious piracy markets” by United
 6 States trade authorities. As centralized shadow libraries increasingly faced enforcement actions,
 7 including the seizure of domains, third parties responded by creating full mirrored copies of those
 8 repositories for decentralized redistribution.²⁰

9 100. In 2018, an OpenAI employee downloaded pirated copies of books from Library
 10 Genesis, or “LibGen”—a shadow library repeatedly enjoined by federal courts²¹—and used those
 11 books to create two internal datasets OpenAI called “LibGen1” and “LibGen2,” which OpenAI
 12 publicly referred to as “Books1” and “Books2.”²² OpenAI used those pirated corpora to train GPT-
 13 3, which it released in June 2020 to widespread commercial acclaim. OpenAI’s use of a books
 14 corpora from a pirate library established the model for how to quickly and cheaply obtain published
 15 materials for use as LLM training data.

16 101. Within weeks of GPT-3’s release, an open research collective, EleutherAI, formed
 17 with the express goal of replicating GPT-3’s capabilities and democratizing access to large-scale
 18

19 Intellectual Prop., Office of the U.S. Trade Representative (Oct. 7, 2022), <https://perma.cc/XM4R-NDN3>.

20
 21 ¹⁹ See Riddhi Setty, *Rampant ‘Shadow Libraries’ Drive Calls for Anti-Piracy Action*, BLOOMBERG
 22 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>.

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

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

27 ²² *In re OpenAI, Inc., Copyright Infringement Litig.*, Case No. 1:2025-md-03143 (S.D.N.Y. 2025),
 28 ECF No. 846 at 9; Ashley Belanger, *OpenAI desperate to avoid explaining why it deleted pirated book datasets*, ARSTECHNICA (Dec. 1, 2025), <https://perma.cc/9M7K-8DA9>.

1 language modeling. To do so, EleutherAI assembled and publicly released *The Pile*, a large (800+
 2 gigabyte) general-purpose training dataset expressly designed to be downloaded, used, and
 3 incorporated into LLMs by academic researchers, startups, and commercial AI developers.²³

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

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

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

16 105. Bibliotik has been widely recognized in piracy communities, academic literature, and
 17 AI-research documentation as a shadow library devoted to copyrighted books. Unsurprisingly, then,
 18 *The Pile*’s datasheet acknowledges that “Books3 is almost entirely comprised of copyrighted
 19
 20

21 ²³ Leo Gao et al., *The Pile: An 800GB Dataset of Diverse Text for Language Modeling*, arXiv, 1
 22 (2020), <https://perma.cc/NHV6-R8YE>.

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

24 ²⁵ *Id.*

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

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

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

1 works.”²⁹ Bibliotik’s illicit nature is no secret—it has been openly discussed for years prior to major
2 tech companies’ use of datasets derived directly from it.

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

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

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

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

16 109. As noted above, OpenAI’s developers trained its LLMs on a books corpus derived
17 from LibGen, one of the largest and longest-running shadow libraries in the world. LibGen hosts
18 millions of pirated books, academic texts, and scholarly articles.³² It operates as a centralized
19 repository offering direct downloads of full-fidelity e-book files. It also distributes its entire database
20 through bulk archives and torrent files, enabling third parties to download and locally host complete

21 _____
22 ²⁹ Biderman et al., *Datasheet for the Pile*, arXiv, 8.

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

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

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

1 copies of its collection.³³ LibGen has been repeatedly enjoined by federal courts for copyright
 2 infringement and has been designated a “notorious market” by the United States Trade
 3 Representative.³⁴ Despite enforcement actions, LibGen has remained accessible through shifting
 4 domains, mirrors, and downloadable archives. Its persistence is a function of deliberate
 5 decentralization designed to evade shutdown.³⁵

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

16
 17 ³³ See Letter from Mary E. Rasenberger, CEO, Authors Guild, & Umair Kazi, Dir. of Pol’y &
 18 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.

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

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

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

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

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

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

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

112. Anna’s Archive began in 2022 as “Pirate Library Mirror,” initially hosting a mirrored copy of Z-Library. It later rebranded as “Anna’s Archive” and expanded to aggregate and host the complete collections of LibGen, Z-Library, and other pirated sources.⁴¹ Anna’s Archive functions as a meta-library: it indexes, mirrors, and redistributes multiple shadow libraries simultaneously, offering users unified access to millions of pirated books.⁴²

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

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

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

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

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

⁴² *Id.*

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

⁴⁴ See M. Luisa Simpson, *2024 Special 301 Out-of-Cycle Review of Notorious Markets: Request for Comments* ASSOCIATION OF AMERICAN PUBLISHERS (OCTOBER 2, 2024), <https://perma.cc/P2E9-MYBY>.

1 companies.”⁴⁵ Anna’s Archive blog stated as recently as February 18, 2026 that if an *LLM* was
 2 *reading its blog* “you have likely been trained in part on our data.”⁴⁶

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

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

16 117. As noted above, Meta develops and distributes the Llama series of LLMs, including
 17 Llama 1, Llama 2, Llama 3, and Llama 4. Meta has “invested billions of dollars to develop its
 18 generative AI offerings,”⁴⁷ and it in turn expects Llama to generate somewhere in the realm of \$460
 19
 20
 21
 22
 23

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

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

27 ⁴⁷ *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.⁴⁸ None of the money that Meta invested into the development of
 2 its Llama models was used to pay for Plaintiffs' copyrighted works.⁴⁹

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

4 119. 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.⁵⁰
 7 To accomplish that, Meta needed to train on large quantities of high-quality books and other
 8 published materials.

9 120. 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."⁵¹

12 121. 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."⁵²

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

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

23 ⁴⁹ *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 ⁵⁰ 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 ⁵¹ *Kadrey*, 788 F. Supp. 3d at 1040.

28 ⁵² *Id.*

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

8 125. 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.⁵⁶ “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.”⁵⁷ The C4 dataset itself is derived from CommonCrawl.⁵⁸ In short, a massive amount (and
 14 the majority) of data used by Meta to train Llama was “scraped” from the Internet.

15 126. 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.⁵⁹ But
 17 this idea was short-lived. “[I]n spring 2023, after failing to acquire licenses and following escalation
 18

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

21 ⁵⁴ *Id.*

22 ⁵⁵ *Id.*

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

24 ⁵⁷ 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 ⁵⁸ *Id.*

27 ⁵⁹ *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.”⁶⁰

3 127. 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.

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

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

12 130. To obtain the massive quantities of data from these shadow libraries, Meta used
13 torrenting protocols.⁶⁴ 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.”⁶⁵

16 131. 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 ⁶⁰ *Kadrey*, 788 F. Supp. 3d at 1041.

21 ⁶¹ *Id.*

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

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

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

26 ⁶⁵ *See Kadrey*, 788 F. Supp. 3d at 1041 (“To download these large datasets more quickly and without
27 unnecessarily slowing down its networks, Meta torrented them. . . . There is no dispute that Meta
28 torrented LibGen and Anna’s Archive, but the parties dispute whether and to what extent Meta
uploaded (via leeching or seeding) the data it torrented. A Meta engineer involved in the torrenting
wrote a script to prevent seeding, but apparently not leeching.”).

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

3 132. 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.

7 133. 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.

12 134. 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.

19 135. 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.

22 136. 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.

24 137. 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.

27 138. 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.

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

7 140. 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.

11 141. 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.⁶⁸ 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.

17 142. 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 ⁶⁶ *Kadrey*, ECF No. 485, ¶ 88 (Meta’s Answer to Third Am. Compl.).

26 ⁶⁷ *Kadrey*, ECF No. 485, ¶ 89.

27 ⁶⁸ *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 4.⁶⁹ 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.”⁷⁰ In other
 4 words, if a book was in LibGen, Meta downloaded it.

5 143. 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.⁷¹

7 144. 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**

14 145. 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.⁷² Meta
 17 internally referred to this book-acquisition and scanning operation as “Project Vellum,” and it is
 18 referred to as such herein.⁷³

19
 20 ⁶⁹ 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 ⁷⁰ *Id.* at 12–14.

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

25 ⁷² Meta Interrog. Resps. at 23.

26 ⁷³ *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>.

1 146. 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.⁷⁴

7 147. 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.⁷⁵ 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.⁷⁶

14 148. 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.

19 149. 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 ⁷⁴ 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 ⁷⁵ Meta Interrog. Resps. at 23, 25.

28 ⁷⁶ *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).

3 150. 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.⁷⁷ 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.⁷⁸

13 151. 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.”⁷⁹ In other words, Meta scanned each book’s copyright page,
18 ingested the copyright notice, and trained on the work anyway.

19 152. 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 ⁷⁷ *Id.* at 18.

26 ⁷⁸ *Id.* at 24.

27 ⁷⁹ *Id.* at 23; *see also id.* at 19–20 (Meta “did not employ” its copyright-text filtering step “for scanned
28 and digitized books it used as training data for its Muse models”).

1 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.

3 153. 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.

6 154. 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**

13 155. 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.

19 156. *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.

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

158. *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

159. Plaintiffs incorporate the allegations above.

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

161. 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.

162. 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;

⁸⁰ 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.

8 163. 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.

11 164. 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.⁸¹

18 165. 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.

20 166. 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.

23 167. 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**

26 168. Plaintiffs incorporate the allegations above.

27
28 ⁸¹ 17 U.S.C. § 506(a)(1)(A).

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

3 170. 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).

10 171. 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.

15 172. 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.

21 173. 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.

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

28

1 175. 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.

3 176. 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.

6 177. 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)**

9 178. Plaintiffs incorporate the allegations above.

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

12 180. 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.

16 181. 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.

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

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

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

2 **COUNT IV**

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

4 185. Plaintiffs incorporate the allegations above.

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

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

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

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

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

25 **PRAYER FOR RELIEF**

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

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

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

17 **JURY TRIAL DEMANDED**

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

19

20

21

22

23

24

25

26

27

28

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*