

1 HUESTON HENNIGAN LLP
Moez M. Kaba (#257456)
2 mkaba@houston.com
Michael M. Purpura (#338216)
3 mpurpura@houston.com
Lee Linderman (#280743)
4 llinderman@houston.com
Brittnee Bui (#333372)
5 bbui@houston.com
523 West 6th Street, Suite 400
6 Los Angeles, CA 90014
Telephone: (213) 788-4340
7 Facsimile: (888) 866-4825

NEWMEYER & DILLION LLP
Michael B. McClellan (#241570)
Michael.McClellan@ndlf.com
Benjamin P. Pugh (#202025)
benjamin.pugh@ndlf.com
Jason L. Morris (#295271)
Jason.Morris@ndlf.com
Harlye S. Carlton (#261641)
Harlye.Carlton@ndlf.com
895 Dove Street, Second Floor
Newport Beach, California 92660
(949) 854-7000; (949) 854-7099 (Fax)

8 Attorneys for Plaintiff
Entrepreneur Media, LLC
9

10 UNITED STATES DISTRICT COURT
11 NORTHERN DISTRICT OF CALIFORNIA
12 SAN FRANCISCO DIVISION
13

14 Entrepreneur Media, LLC,
15 Plaintiff,
16 v.
17 Meta Platforms, Inc.,
18 Defendant.

Case No.:

**COMPLAINT FOR COPYRIGHT
INFRINGEMENT AND VIOLATION
OF THE DIGITAL MILLENNIUM
COPYRIGHT ACT**

DEMAND FOR JURY TRIAL

19
20 Plaintiff Entrepreneur Media, LLC (“Entrepreneur”), for its complaint against Defendant
21 Meta Platforms, Inc. (“Meta”), alleges as follows:

22 **NATURE OF THE ACTION**

23 1. Meta seeks to build a multibillion-dollar artificial intelligence empire on a
24 foundation of systematic and widespread copyright theft. While other companies invested
25 millions in licensing content to train their large language models (“LLMs”), Meta—at the
26 direction of senior executives, including CEO Mark Zuckerberg—chose industrial-scale piracy.
27
28

1 Meta deliberately and willfully stole hundreds of terabytes of copyrighted works to train its
2 family of LLMs, known collectively as “Llama.”

3 2. Entrepreneur is a publisher with a storied history. For more than fifty years,
4 Entrepreneur has published books and magazines, mostly on topics such as startup advice,
5 business growth, franchising, technology and innovation, personal finance, leadership and
6 management, social entrepreneurship, and marketing strategies. Entrepreneur owns the copyrights
7 for dozens of books and hundreds of magazine issues and articles that span decades of
8 publication.

9 3. Entrepreneur embraces legitimate technological advancement and recognizes AI as
10 a potentially valuable tool when developed through lawful means. Entrepreneur would have
11 welcomed good-faith negotiations to license its copyrighted works for AI training purposes.
12 Instead, Meta chose to steal Entrepreneur’s works to train Llama to generate competing outputs
13 that displace the market for Entrepreneur’s copyrighted works.

14 4. Meta pirated at least hundreds of copyrighted works owned by Entrepreneur,
15 including professionally published books, issues of Entrepreneur Magazine, and individual
16 articles published in those magazines. These works, which represent years of creative effort and
17 substantial investment by their authors, were taken without permission, without payment, and
18 without regard for the law.

19 5. While Meta touts the billions it is paying AI engineers, it has offered nothing to
20 the creators, like Entrepreneur, of the works that are the foundation for Llama. When training its
21 LLMs, Meta made a calculated business decision to prioritize speed and cost savings over
22 copyright compliance.

23 6. Meta systematically downloaded and copied millions of copyrighted works from
24 sources it knew or should have known contained pirated content, including notorious repositories
25 that have been the subject of copyright-enforcement actions. While downloading millions of
26 copies of copyrighted materials from some of these shadow libraries, Meta also acted as a
27 distributor of the pirated works, facilitating further copyright infringement.

28

1 **DIVISIONAL ASSIGNMENT**

2 12. Under Civil Local Rule 3.2(d), assignment of this case to the San Francisco
3 Division is proper because Meta is headquartered in San Mateo County, where a substantial part
4 of the events giving rise to the claim occurred.

5 **PARTIES**

6 13. Plaintiff Entrepreneur Media, LLC is a California Limited Liability Company with
7 headquarters located at 1651 East Fourth Street, Suite 125, Santa Ana, CA 92701. Entrepreneur is
8 a leading publisher of business, entrepreneurship, and professional development content
9 distributed nationwide and owns or exclusively controls copyrights in hundreds of books and
10 magazines and thousands of magazine articles.

11 14. Defendant Meta is a Delaware corporation with its principal place of business at
12 1601 Willow Road, Menlo Park, CA 94025. Meta operates one of the world’s largest technology
13 companies.

14 **FACTUAL ALLEGATIONS**

15 **A. Meta’s Theft of Copyrighted Works**

16 15. Meta has invested billions of dollars into developing its Llama family of LLMs.
17 These AI systems, designed to generate naturalistic text in response to user prompts, were built on
18 a foundation of systematic copyright infringement.

19 16. While competitors in the AI industry have paid for content licenses, Meta adopted
20 what it calls the “gap approach”—pirate as much copyrighted content as possible and use
21 licensing only to fill in the gaps of content that cannot be pirated, a strategy adopted at the
22 direction of its most senior executives.

23 17. Meta obtained Entrepreneur’s copyrighted works through deliberate acts of piracy
24 from notorious “shadow libraries”—websites that host troves of stolen copyrighted materials.

25 **B. How LLMs Depend on Copyrighted Content**

26 18. An LLM, including Llama, is AI software designed to process and generate
27 humanlike text. Their capabilities depend on the quality and quantity of text used to train them.

1 As one AI researcher put it, an LLM’s behavior “is not determined by architecture,
2 hyperparameters, or optimizer choices. It’s determined by your dataset, nothing else.”¹

3 19. The training process involves copying vast quantities of text and feeding it into
4 neural networks that learn to predict linguistic patterns. Through millions or billions of
5 adjustments to internal parameters, the model learns not only grammar and vocabulary, but also
6 style, tone, structure, and the complex relationships between ideas. This is why copyrighted
7 materials are so valuable to an LLM: they embody professionally edited high-quality writing that
8 reflects current usage, credible information, and coherent expression. Under Article I, Section 8
9 of the Constitution, copyrights may be secured only “for limited times.” Copyrighted works are
10 therefore the best examples of authentic, up-to-date human expression—the very qualities LLMs
11 must emulate to produce realistic, humanlike text.

12 20. Books are uniquely valuable training material. Unlike fragmented internet text,
13 books provide sustained, professionally edited narratives that model how to maintain coherence
14 over thousands of words. They demonstrate the sophisticated ways humans organize and express
15 complex ideas. Industry research and Meta’s own engineers have acknowledged that books are
16 essential for developing long-range context modeling and coherent narratives. Without substantial
17 quantities of book-length texts, LLMs, including Llama, cannot achieve the sophistication
18 required for commercial success.

19 21. Periodicals provide distinct but equally valuable training material. Professional
20 magazines deliver concise, authoritative content on specialized topics, offering models of clarity,
21 structure, and practical focus. Their targeted length teaches LLMs such as Llama to generate
22 actionable, domain-specific content while maintaining subject-matter expertise. Articles in
23 *Entrepreneur Magazine*, for example, distill complex topics into accessible guidance and timely
24 analysis—precisely the attributes that LLMs trained on professional periodicals now replicate.

25
26 ¹ “The ‘it’ in AI models is the data set,” James Betker, <https://nonint.com/2023/06/10/the-it-in-ai-models-is-te-dataset/> (June 10, 2023), available at
27 <https://web.archive.org/web/20250902175704/https://nonint.com/2023/06/10/the-it-in-ai-models-is-the-dataset/>
28 (last visited November 6, 2025).

1 22. By training on copyrighted works like Entrepreneur’s books and magazines, LLMs
2 including Llama learn to reproduce the same practical, instructional content that forms the core of
3 Entrepreneur’s business.

4 **C. The Licensing Market Meta Deliberately Bypassed**

5 23. A market for licensing copyrighted training data has developed in recent years. AI
6 companies including Anthropic, OpenAI, and Google have paid billions of dollars to obtain
7 proactive and retroactive licenses for LLM training—and according to public records, Meta itself
8 has AI-licensing deals with Reuters, Shutterstock, and Universal Music Group.²

9 24. Meta recognized the critical importance of published works for training
10 competitive LLMs and understood that such content required licensing. Meta explored legitimate
11 acquisition of training content, i.e., paying for licenses to copyrighted materials. Yet despite these
12 licensing opportunities, Meta determined it was faster and cheaper to just take what it wanted
13 without regard for the rights of copyright owners or the creators of the works being infringed.

14 **D. Meta’s Piracy Through Shadow Libraries and Books3**

15 25. In February 2023, Meta introduced Llama in a paper (“Llama 1 Paper”) claiming it
16 was “possible to train state-of-the-art models using publicly available datasets exclusively,
17 without resorting to proprietary and inaccessible datasets” and further claimed its training data
18 was “compatible with open-sourcing.”³

19 26. These claims were at best misleading. “Publicly available” does not mean “public
20 domain.” A work in the public domain is not protected by copyright, whereas a work that is
21 publicly available may be protected by copyright laws. Open sourcing requires copyright holders
22 to explicitly place material under an open-source license—copyrighted materials are not
23 “compatible with open sourcing” unless the copyright holder grants such permission.

24
25
26 ² For a comprehensive list of copyright licensing deals, including those involving the listed AI
companies, see <https://copyrightalliance.org/artificial-intelligence-copyright/licensing/> (last
visited November 5, 2025).

27 ³ “LLaMA: Open and Efficient Foundation Language Models,” Touvron et al., accessible at
28 arxiv.org/pdf/2302.13971, at 1 (last visited November 6, 2025).

1 27. The Llama 1 Paper revealed that Meta’s training dataset included “the Books3
2 section of The Pile,” described as “a publicly available dataset for training large language
3 models.”⁴ In truth, Books3 is a collection of pirated books.

4 28. The copyrighted works in Books3 became “publicly available” only because
5 Shawn Presser and his organization, Eleuther AI, illegally pirated these works. Presser has
6 confirmed in public statements that Books3 represents “all of Bibliotik” and contains 196,640
7 books.

8 29. Bibliotik is one of several notorious “shadow library” websites. Eleuther AI
9 revealed in a December 2020 paper that Books3 is comprised of “books derived from a copy of
10 the contents of the Bibliotik private tracker,” adding: “We included Bibliotik because books are
11 invaluable for long-range context modeling research and coherent storytelling.”⁵

12 30. Eleuther AI’s paper concedes that “there is little acknowledgement of the fact that
13 the processing and distribution of data owned by others may also be a violation of copyright
14 law.”⁶ The Pile’s datasheet explicitly states that “Books3 is almost entirely comprised of
15 copyrighted works.”⁷

16 **E. Meta’s Torrenting and Distribution of Pirated Works**

17 31. Shadow libraries like Bibliotik, Library Genesis (“LibGen”), Z-Library, and others
18 make copyrighted works available for download and distribution using various methods,
19 including decentralized peer-to-peer file-sharing systems called torrenting. Torrenting is a method
20 in which “the *downloaders* of a file *barter* for chunks of [the file] by uploading and downloading
21 them in a tit-for-tat manner to prevent parasitic behavior.”⁸ Users share data both while
22 downloading (called “leeching”) and after downloading (called “seeding”).

23
24 ⁴ *Id.* at 2.

25 ⁵ “The Pile: An 800GB Dataset of Diverse Text for Language Modeling,” Gao et al.,
<https://arxiv.org/pdf/2101.00027> (Dec. 31, 2020), at 3-4 (last visited November 6, 2025).

26 ⁶ *Id.* at 14.

27 ⁷ “Datasheet for the Pile,” Gao et al., <https://arxiv.org/pdf/2201.07311> (Jan. 20, 2022), at 15 (last
visited November 6, 2025).

28 ⁸ Johan Pouwelse et al., *The BitTorrent P2P File-Sharing System: Measurements and Analysis*,
IPTPS 2005, at 206 (2005) (emphasis in original).

1 32. Upon information and belief, Meta utilized torrenting sites such as BitTorrent and
2 LibTorrent to download and redistribute Entrepreneur’s protected works from shadow libraries,
3 including LibGen. By using these protocols, Meta not only reproduced Entrepreneur’s works but
4 also automatically distributed them to other users in the torrenting network, compounding the
5 infringement and facilitating global piracy.

6 33. Upon information and belief, Meta developed and deployed software designed to
7 process pirated books for training, including functionality to remove copyright management
8 information (“CMI”) such as copyright notices, metadata, and author identifiers. Meta did so to
9 conceal its infringement and to facilitate further unauthorized use.

10 **F. Meta’s Pivot to Commercial Exploitation**

11 34. Despite relying on pirated datasets, Meta initially claimed that Llama 1 would be
12 limited to a “noncommercial license focused on research use cases” to “maintain integrity and
13 prevent misuse.”⁹

14 35. This non-commercial positioning was temporary. Starting with Llama 2, Meta
15 pivoted to full commercial exploitation. The Llama FAQ page now states that “Llama models are
16 licensed under a bespoke commercial license” allowing “broad commercial use.”¹⁰

17 36. Meta has deliberately concealed the training datasets used for models after Llama
18 1. This concealment serves to hide the extent of its infringement and avoid litigation from
19 copyright holders whose works Meta continues to exploit. On information and belief, each
20 successive Llama model was and is trained using prior Llama checkpoints and overlapping
21 training datasets, so that once Entrepreneur’s works were ingested, their protected expression
22 persisted in and was carried forward into later—and future—Llama iterations unless affirmatively
23 identified and purged.

24
25
26 ⁹ “Introducing LLaMA: A foundational, 65-billion-parameter large language model,”
27 <https://ai.meta.com/blog/large-language-model-llama-meta-ai/> (Feb. 24, 2023) (last visited
November 6, 2025).

28 ¹⁰ “Troubleshooting & FAQ,” <https://www.llama.com/faq/> (last visited November 6, 2025).

1 **G. Entrepreneur’s Stolen Works and Resulting Harm**

2 37. Entrepreneur owns valid registered copyrights in books, magazines and magazine
3 articles that Meta misappropriated to train one or more models of Llama. These works span
4 business strategy, entrepreneurship, professional development, and practical guidance.

5 38. Public filings in *Kadrey* describe internal Meta communications reflecting CEO
6 Mark Zuckerberg’s authorization to download and use the LibGen corpus to train Llama. And
7 publicly available LibGen metadata shows that numerous Entrepreneur titles and registered
8 magazine issues appear in LibGen, including the exemplars listed below.¹¹

9 39. Representative book titles that appear in the LibGen dataset include:

- 10 • “Start Your Own Coaching Business: Your Step-by-Step Guide to
11 Success”
- 12 • “Start Your Own Import/Export Business”
- 13 • “Ultimate Guide to Pinterest for Business”
- 14 • “Breakthrough: How to Harness the AHA! Moments that Spark Success”
- 15 • “Start Your Own Business, 6th Edition”

16 40. Representative magazine articles from 2010 issues of *Entrepreneur Magazine* that
17 appear in the LibGen dataset include:

- 18 • “The Successful Optimist”
- 19 • “The Red Pen Rule for Marketing Copy”
- 20 • “The Four Keys to Raising Capital”
- 21 • “Weird Companies That Work”
- 22 • “A Whale of an Idea”
- 23 • “Myth of the Business Plan”

24
25
26
27 ¹¹ The LibGen database was created and published by *The Atlantic*, and is available at
28 <https://www.theatlantic.com/technology/archive/2025/03/search-libgen-data-set/682094/> (last
accessed on November 5, 2025).

1 41. A list of Entrepreneur’s registered copyrighted works stolen by Meta is attached as
2 **Exhibit A** (the “Infringed Works”).¹² The Infringed Works were registered with the U.S.
3 Copyright Office prior to the acts of infringement alleged herein. For each of the Infringed
4 Works, Entrepreneur owns or exclusively controls the copyrights in those works. These works
5 represent substantial creative and financial investment, having undergone development, editing,
6 design, and marketing. They generate revenue through direct sales, licensing agreements, and
7 derivative works.

8 42. For any asserted magazine articles, Entrepreneur owns the copyrights in the
9 constituent contributions by written agreements and/or as works made for hire; Entrepreneur’s
10 collective-work registrations for the issues identified in Exhibit A encompass those contributions.

11 43. Upon information and belief, the Infringed Works were included in the datasets
12 Meta used to train Llama. Meta made multiple unauthorized reproductions throughout Llama’s
13 development: initial downloading, server storage, processing for training, and incorporation into
14 model parameters. Each reproduction violated Entrepreneur’s exclusive rights in the Infringed
15 Works.

16 44. At no time did Meta seek or obtain permission from Entrepreneur to copy,
17 distribute, or use the Infringed Works.

18 **H. Market Dilution and Ongoing Harm**

19 45. Meta’s decision to steal Entrepreneur’s Works rather than lawfully license them
20 inflicted immediate economic harm, even before any outputs displaced sales. By bypassing the
21 existing market for licensing training data, Meta deprived Entrepreneur of licensing revenue that
22 would ordinarily be paid for reproductions made during acquisition, preprocessing, and the
23 training of its LLMs. Entrepreneur is entitled to the fair-market value of licenses for Meta’s
24 copying and use of the Infringed Works in its LLM development, in addition to any damages for
25 downstream substitution.

26
27 _____
28 ¹² Entrepreneur’s investigation is ongoing. Entrepreneur reserves the right to amend this
Complaint to identify additional copyrighted works infringed by Meta.

1 46. Meta’s infringement has inflicted severe and ongoing damage to Entrepreneur’s
2 business. Since the introduction of LLMs to the market, Entrepreneur’s digital book sales have
3 dropped by approximately 50%.

4 47. Llama’s output competes directly with the Infringed Works. As the United States
5 District Court for the Northern District of California observed, “the market for certain nonfiction
6 works—for example, books about how to take care of your garden—could be greatly diminished
7 by the ability of LLMs to produce books on that topic.”¹³ The market for professional periodicals
8 faces similar displacement, as users can now generate competing content on demand instead of
9 visiting Entrepreneur’s website, purchasing *Entrepreneur Magazine*, or licensing articles from
10 Entrepreneur’s archives. *See Kadrey*, 788 F. Supp. 3d at 1035 (“If a company uses copyrighted
11 magazine articles to train a model capable of generating similar articles, it’s easy to imagine the
12 market for the copied articles diminishing substantially. Especially if the AI-generated articles are
13 made available for free.”).

14 48. Llama, trained on Entrepreneur’s protected works, can generate variations of
15 Entrepreneur’s business advice, startup guides, and instructional content. For example, users who
16 might have purchased “Start Your Own Import/Export Business” can instead prompt Llama to
17 generate similar guidance instantly and at no cost.

18 49. Had Meta chosen licensing—as other developers have done, and as Meta itself
19 does when it needs to fill “gaps” in its training data—Entrepreneur would have received
20 substantial compensation. Instead, Meta’s willful infringement has excluded Entrepreneur from
21 the licensing market while forcing it to compete against AI trained on its own works.

22 50. The U.S. Copyright Office has warned that “the speed and scale at which AI
23 systems generate content pose a serious risk of diluting markets for works of the same kind as in
24 their training data.”¹⁴ This risk has materialized for Entrepreneur. Every sale lost to AI-generated
25 alternatives represents both immediate revenue loss and long-term market erosion.

26 ¹³ *Kadrey v. Meta Platforms, Inc.*, 3:23-cv-03417-VC, Dkt. No. 598 at 29 (N.D. Cal. June 25,
27 2025).

28 ¹⁴ U.S. Copyright Office, *Copyright and Artificial Intelligence, Part 3: Generative AI Training*
(*prepublication version*), (May 2025) (“Copyright Office Draft”) at 65.

1 51. Empirical research shows that piracy depresses legitimate sales. For example, a
2 2016 study found that pirated eBooks reduced legitimate sales by as much as 14%¹⁵—and that
3 was before LLMs could replicate the content on which they were trained.

4 52. Treating largescale, systematic copying of copyrighted materials as “fair use”
5 would undermine the market incentive for original reporting and analysis by signaling to
6 publishers and authors alike that their creative efforts may be used to develop commercial
7 products without compensation. This use is not “transformative”: Meta used Entrepreneur’s
8 works to create a commercial product that serves the same or highly similar purpose and is likely
9 to supplant the originals. Because LLMs like Llama depend on the continuous ingestion of fresh,
10 human-created content to remain competitive, permitting such copying at scale would depress
11 future creation of nonfiction, journalism, and even fiction.

12 **CLAIMS FOR RELIEF**

13 **Count I: Direct Copyright Infringement**

14 **17 U.S.C. § 501**

15 53. Entrepreneur incorporates by reference the preceding factual allegations as though
16 set forth fully herein.

17 54. As owner of the registered copyrights in the Infringed Works, Entrepreneur
18 possesses exclusive rights to those works under 17 U.S.C. § 106.

19 55. Meta infringed on the exclusive rights of Entrepreneur by, among other things,
20 downloading and copying pirated works, including the Infringed Works, from various shadow
21 libraries such as LibGen and Z-Lib. Meta made multiple unauthorized copies of these pirated
22 datasets, storing them on local and external servers.

23 56. Entrepreneur never authorized Meta to make copies or derivatives of its Infringed
24 Works. All rights under copyright law belong exclusively to Entrepreneur.

25 57. Meta made copies of the Infringed Works during the training of Llama (including
26 all versions of Llama) without Entrepreneur’s permission.

27 ¹⁵ Imke Reimers, Can Private Copyright Protection Be Effective? Evidence from Book Publishing
28 (2016), 59 J. L. & ECON. 411, 414 (2016) <https://doi.org/10.1086/687521>.

1 58. During the download process of LibGen and other shadow libraries, Meta also
2 operated as a distributor of the pirated works.

3 59. Entrepreneur has been injured by Meta’s willful acts of direct copyright
4 infringement. Entrepreneur is entitled to statutory damages, actual damages, restitution of profits,
5 disgorgement, attorneys’ fees and costs, and other remedies available by law.

6 60. Entrepreneur has been and continues to be irreparably injured due to Meta’s
7 conduct, for which there is no adequate remedy at law. Meta will continue to infringe the
8 exclusive rights of Entrepreneur unless Meta’s infringing activity is enjoined by the Court.
9 Entrepreneur is therefore entitled to permanent injunctive relief barring Meta’s ongoing
10 infringement.

11 **Count II: Contributory Copyright Infringement**

12 **17 U.S.C. § 501 and federal common law**

13 61. Entrepreneur incorporates by reference the preceding factual allegations as though
14 fully set forth herein.

15 62. Meta used torrenting protocols to download the Infringed Works from pirated
16 sources. The inherent architecture of torrenting protocols require users to simultaneously upload
17 (“seed”) content to other users while downloading, automatically making Meta a distributor of
18 pirated works.

19 63. Meta’s willful infringement materially contributed to and induced further
20 infringement of Entrepreneur’s copyrights by making the Infringed Works available through peer-
21 to-peer file-sharing networks such as BitTorrent and LibTorrent.

22 64. Torrenting protocols operate on reciprocity—download speeds increase based on
23 upload contributions. To achieve the rapid, large-scale acquisition of millions of works, Meta
24 used the default BitTorrent configuration to maximize bandwidth, necessarily seeding substantial
25 portions of the Infringed Works to third parties without authorization.

1 65. Meta’s actions provided the means and encouragement for widespread
2 infringement by others, including redistribution of Entrepreneur’s works across global piracy
3 networks.

4 66. Meta knew or had reason to know that its conduct constituted infringement of
5 Entrepreneur’s rights.

6 67. By knowingly inducing and materially contributing to infringement by third
7 parties, Meta is liable for contributory copyright infringement.

8 68. As a direct and proximate result of Meta’s conduct, Entrepreneur has suffered
9 damages in an amount to be proven at trial.

10 **Count III: Violation of the Digital Millennium Copyright Act**

11 **17 U.S.C. § 1202**

12 69. Entrepreneur incorporates by reference the preceding factual allegations as though
13 fully set forth herein.

14 70. Entrepreneur’s works contain CMI, including front-matter copyright notices
15 identifying Entrepreneur Media, LLC (or Entrepreneur Press), author names, publication year,
16 and publisher, and embedded EPUB/PDF metadata fields. Copies of these works, available in
17 shadow-library repositories during Meta’s training windows, carried such CMI.

18 71. On information and belief, Meta’s preprocessing pipeline extracted raw text from
19 EPUB and PDF containers and programmatically discarded front-matter pages and embedded
20 metadata, thereby intentionally removing or altering CMI from the Infringed Works before
21 writing the text to training datasets and shards.

22 72. Public filings in *Kadrey* describe a Meta engineer “filtering copyright lines” and
23 other identifiers from LibGen files to prepare a CMI-stripped corpus for Llama training. On
24 information and belief, Meta applied the same or similar filters to the Infringed Works.

25 73. Meta knew or had reason to know that removing CMI would facilitate
26 infringement because CMI removal makes it harder to trace which copyrighted works were used
27 in training; the absence of CMI in training data reduces the likelihood that LLMs will output
28

1 identifying information that would reveal their training sources; and industry participants
2 regularly discuss CMI removal as a method to avoid detection of copyrighted materials in training
3 sets.

4 74. Meta’s removal or alteration of CMI was intentional and designed to prevent
5 Llama from revealing that it was trained on copyrighted works; reduce the “memorization” of
6 identifying information that could expose infringement; and create plausible deniability about
7 which specific works were used in training.

8 75. Following removal, on information and belief, Meta distributed and made
9 available copies of the Infringed Works without CMI, including by seeding CMI-less files during
10 BitTorrent acquisition and by sharing CMI-less training datasets and derivatives within Meta and
11 with third-party service providers.

12 76. The removal of CMI has harmed Entrepreneur by making it more difficult to
13 detect the full extent of Meta’s use of the Infringed works, prove which specific works were
14 incorporated into the training of Llama, and track how the Infringed works are being processed
15 and exploited.

16 77. Meta’s actions violate 17 U.S.C. § 1202(b), which prohibits the intentional
17 removal or alteration of CMI knowing that it will facilitate infringement.

18 78. For each violation, Entrepreneur is entitled to statutory damages or alternatively
19 actual damages and profits of the violator under 17 U.S.C. § 1203(c). Entrepreneur is also entitled
20 to attorneys’ fees under 17 U.S.C. § 1203(b)(5).

21 **DEMAND FOR JUDGMENT**

22 Wherefore, Entrepreneur requests that the Court enter judgment on its behalf by ordering:

- 23 a. Judgment in favor of Entrepreneur and against Meta.
24 b. A declaration that Meta has infringed Entrepreneur’s exclusive copyrights in
25 the Infringed Works under the Copyright Act.
26 c. A declaration that such infringement was and is willful.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

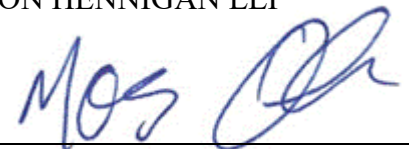
- d. An award of actual and statutory damages to the maximum extent as allowed by law or as elected by Entrepreneur.
- e. Permanent injunctive relief enjoining Meta from engaging in the infringing conduct alleged herein.
- f. An injunction prohibiting the removal or alteration of CMI from Entrepreneur’s works and requiring preservation/restoration of CMI and maintenance of provenance logs in any ingestion, preprocessing, or retraining.
- g. Awarding Entrepreneur pre-judgment and post-judgment interest pursuant to the claim(s) for relief.
- h. Awarding Entrepreneur costs, expenses, and attorneys’ fees as permitted by law under 17 U.S.C. § 505 and/or 17 U.S.C. § 1203(b).
- i. Awarding Entrepreneur further relief as the Court may deem just and proper under the circumstances.

JURY TRIAL DEMANDED

Under Federal Rule of Civil Procedure 38(b), Entrepreneur demands a trial by jury on all claims and issues so triable.

Dated: November 6, 2025

HUESTON HENNIGAN LLP

By: 

Moez M. Kaba
Michael M. Purpura
Lee Linderman
Brittnee Bui

NEWMAYER & DILLION LLP
Michael B. McClellan
Benjamin P. Pugh
Jason L. Morris
Harlye S. Carlton

Attorneys for Plaintiff
ENTREPRENEUR MEDIA, LLC