Elizabeth Brannen (SBN 226234)
ebrannen@stris.com
John Stokes (SBN 310847)
jstokes@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

Christopher M. Rigali
(*pro hac vice* forthcoming)
crigali@stris.com
Jacqueline Sahlberg
(*pro hac vice* forthcoming)
jsahlberg@stris.com
STRIS & MAHER LLP
1717 K St NW Ste 900
Washington, DC 20006
T: (202) 800-5749

Kyle Roche (*pro hac vice* forthcoming)
kroche@fnf.law
Devin (Velvel) Freedman (*pro hac vice* forthcoming)
vel@fnf.law
Alex Potter (*pro hac vice* forthcoming)
apotter@fnf.law
FREEDMAN NORMAND FRIEDLAND LLP
155 E. 44th Street, Ste 915
New York, NY 10017
T: (646) 494-2900

Counsel for Plaintiff

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

CHICKEN SOUP FOR THE SOUL, LLC, Plaintiff, v. ANTHROPIC PBC, Defendant.	Civil Case No.: **COMPLAINT** **DEMAND FOR JURY TRIAL**

1. Plaintiff Chicken Soup for the Soul, LLC ("Chicken Soup for the Soul" or, simply, "Chicken Soup") brings this action against Anthropic PBC ("Anthropic" or "Defendant"), and alleges as follows:

I. INTRODUCTION

2. This case concerns Defendant Anthropic's exploitation of Plaintiff Chicken Soup for the Soul's copyrighted works to build its "Claude" family of large language models ("LLMs"). Defendant is in open and active competition with its tech company rivals to win what many have deemed the "generative AI arms race." Participants in this race are under immense pressure to build more, better, and faster, with an eye towards ensuring that *its* generative AI models are widely adopted. Broad adoption of a company's generative AI models will yield great profits, the thinking goes; on the converse, failure to move quickly means missing the boat on this trillion-dollar industry. To remain competitive in this race, Anthropic turned to notorious online "shadow libraries" and similar pirated datasets to obtain vast quantities of copyrighted materials—books, articles, training materials, and the like—which it desperately wanted to train and optimize its LLM models. Anthropic could have—*and should have*—paid copyright owners, including Plaintiff, for licenses to use their copyrighted works in connection with training its Claude models. Instead, Anthropic fed from, and then back into, the illegal market for digitally available copyrighted materials. Ultimately, Anthropic downloaded, reproduced, distributed, and defaced these copyrighted works, including Plaintiff's works, to get ahead in the generative AI arms race and improve its bottom line. Plaintiff brings this action to hold Anthropic accountable for its brazen acts of copyright infringement.[1]

3. LLMs are a form of "generative artificial intelligence" or "generative AI." These models are designed to process and emit natural language. Over the last several years, many of the world's biggest and wealthiest technology companies have entered the market to develop, distribute, and commercialize generative AI models, believing these and other forms of artificial intelligence have massive potential for growth in revenue and profits. Because artificial intelligence, including

[1] Unless otherwise indicated, references to Anthropic's "Claude" refers to all versions of Claude in any stage of their development or deployment.

generative AI, is viewed by industry leaders as the next foundational layer of the digital economy, the pressure to build more, better, and faster has led to the aforementioned AI arms race.

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

5. For its Claude training datasets, Anthropic needed gold standard training data. To get its hands on published content, it bypassed the licensing market and downloaded copyrighted works from these shadow libraries. Anthropic often torrented them. "Torrenting" works by breaking a file into many small pieces and distributing those pieces across a network of participating computers. A user downloads portions of a file from numerous other computers that already possess the file, and software reassembles those pieces into a completed whole on the user's machine.

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

7. Anthropic not only torrented copyrighted materials from online shadow libraries, it also purchased and scanned millions of physical books—reproducing their contents—without authorization. Its downloading, torrenting, and scanning activities were done both to build and train

its Claude models, but also to build and maintain a central library that it intended to retain indefinitely. In connection with its training of the Claude models, Anthropic also embedded near-verbatim copies of copyrighted works, including Plaintiff's works, in Claude's model weights.

8. In addition to ripping pirated copies, reproducing, and distributing without authorization Plaintiff's copyrighted materials, Anthropic also stripped these materials of copyright management information, enabling, facilitating, and concealing the infringement of these works. Anthropic did this to optimize its models' performance.

9. Adding insult to all of this injury, Anthropic's use of Plaintiff's copyrighted materials facilitated Anthropic's creation of AI models capable of generating content that directly competes with and will compete directly with Plaintiff's content. "Learning from" the creativity and expression embodied in thousands upon thousands of copyrighted works, including Plaintiff's works, Claude has the capability to flood the market with free and paid content that vies with Plaintiff for consumer attention.

10. The result of Anthropic's use of copyrighted materials to train Claude? An estimated $1 trillion valuation as of April 2026.[2] The Copyright Act doesn't allow Anthropic to get a free ride on the backs of Plaintiff and the other creators and publishers whose copyrighted works it exploited to build its trillion-dollar generative AI enterprise. Plaintiff brings this action to hold Anthropic accountable for the infringement that enabled its rise in the generative-AI marketplace, and to enforce the fundamental principle that creative expression cannot be taken, copied, or exploited without permission or compensation.

11. To redress Anthropic's repeated, unlawful, and *en masse* infringement of its works, Plaintiff seeks (1) damages, (2) permanent injunctive relief barring Anthropic's ongoing infringement, and (3) any additional remedies the law provides.

12. Plaintiff elects not to bring this case as a class action because the Copyright Act entitles it to recover individualized statutory damages, determined by a jury, for Anthropic's

[2] Ben Bergman, *Anthropic has surged to a trillion-dollar valuation on secondary markets, overtaking OpenAI*, Business Insider (Apr. 22, 2026, 5:29 PM), https://perma.cc/Q5GH-HSKZ.

infringement and related conduct. Plaintiff desires to retain full control of its case and avoid having its rights diluted by being swept into sprawling class-action settlements structured to resolve claims for pennies on the dollar. Recent history has shown that certain class actions and proposed settlement(s) seem to serve the tech conglomerate-infringers, not creators and publishers. LLM companies should not be able to so easily extinguish thousands upon thousands of high-value claims at bargain-basement rates, eliding what should be the true cost of their massive willful infringement.

13. That is not how Plaintiff plans to proceed. Under established Supreme Court precedent, "the amount of statutory damages is a question for the jury."[3] The Copyright Act thus vests authors with the right to have a jury evaluate the willfulness of infringement and assign a damages amount tailored to Anthropic's conduct.

14. In sum, the Copyright Act's statutory-damages and attorneys' fee regime empowers individual authors and publishers to hold infringers accountable without the need for class action treatment. That is what Plaintiff has chosen to do.

II. PARTIES

A. Plaintiff

15. Plaintiff Chicken Soup for the Soul, LLC, is a Connecticut limited liability company and the owner or exclusive licensee of numerous copyrights in books published under the *Chicken Soup for the Soul* brand.

16. A non-exhaustive list of registered copyrights owned by Plaintiff is included as Exhibit A (herein, the "Chicken Soup Works").[4]

B. Defendant

17. Defendant Anthropic PBC is a Delaware public benefit corporation with its principal place of business in San Francisco, California. Anthropic develops and commercializes large

[3] *Feltner v. Columbia Pictures Television, Inc.*, 523 U.S. 340, 353 (1998).

[4] Even where other individuals or entities are listed as copyright claimants on the relevant copyright registrations, Plaintiff is the owner of all copyrights listed in Exhibit A.

language models (including the Claude series), which were trained using datasets sourced in part from shadow libraries and other datasets containing pirated books.

III. JURISDICTION AND VENUE

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

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

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

IV. FACTUAL ALLEGATIONS

A. Chicken Soup and its Protected Works

21. Plaintiff Chicken Soup for the Soul, LLC is one of the most widely recognized and commercially successful nonfiction publishing franchises in modern history, publishing more than 500 million copies worldwide. Since the publication of the first volume in 1993, the series has expanded into hundreds of titles and has sold more than 500 million copies worldwide. The books consist of curated collections of original narrative nonfiction, typically written in the first person and edited to convey authentic human experiences, emotional storytelling, and clear narrative structure.

22. For decades, books in the *Chicken Soup for the Soul* series have been curated, edited, and refined to present clear, compelling narrative structures and expressive prose that resonate with readers across cultures and generations. The extraordinary commercial success and editorial consistency of the *Chicken Soup for the Soul* series made it among the most recognizable and valuable collections of narrative nonfiction in the publishing industry—and, as alleged below, an especially attractive target for companies seeking high-quality textual data to train large language models.

23. As alleged below, Defendant targeted Plaintiff's works because they were of exceptional value as training data for its LLMs. Books teach models how narrative flows, how human expression is structured, how syntax and rhythm operate, and how ideas are communicated through creative choices.

24. The *Chicken Soup for the Soul* series is particularly valuable in this regard. Each volume contains dozens of tightly edited, first-person narratives written in natural, conversational language that conveys emotion, moral reflection, and coherent storytelling in concise form. These characteristics make the series uniquely well suited for training large language models to replicate authentic human voice, narrative pacing, emotional tone, and story structure. Instead of paying for that value or licensing access to these works, Defendant pilfered illegal copies and used those copies to build systems now worth many hundreds of billions of dollars.

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

B. LLMs and the Generative AI "Arms Race"

26. "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.

27. 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 they have learned and been trained to recognize.

28. 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."[5] While not all LLM developers use the same terminology, developers commonly reference a "pre-training" process—"in which a massive amount of computing power and data is spent to teach the model the broad foundations of language, grammar, and reasoning"—and a "post-training" or "fine-tuning" process "where the pre-trained model is further trained on a (relative to pre-training) smaller amount of carefully curated data of specific tasks."[6] For purposes of this Complaint, "training" refers to all stages of the LLM training process.

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

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

[6] *Id.*

short, with respect to training data, the LLM development pipeline typically involves numerous reproductions of any given piece of the dataset.

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

31. In determining the corpus of training data for an LLM, a model's developer typically considers "the quantity of data, its quality, and the ultimate purpose(s) of the model."[8]

32. In terms of data quantity, the LLMs that have been developed and are being developed by leading technology companies are trained on enormous datasets, typically on the scale of terabytes. This is because AI researchers have found that "increasing the quantity of training data typically increases a model's 'performance.'"[9]

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

34. That books make for particularly valuable training material for LLMs is evidenced by the fact that several—if not most—of the major LLM developers either have paid for or expressly contemplated paying (large sums) for licenses to use copyrighted works to train their LLMs. Indeed,

---

[7] James Betker, *The "it" in AI Models is the Dataset*, (June 10, 2023), https://perma.cc/ZCH9-H53S.

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

[9] *Id.* at 10.

[10] *Id.* at 11.

in an *amicus* brief filed in *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-3417 (N.D. Cal.), the Association of American Publishers cited *dozens* of "AI licensing deals for textual works," including deals involving Amazon, Microsoft, OpenAI, and Perplexity.[11] And, as the Court found in that litigation, Meta considered spending up to $100 million to license copyrighted materials in connection with LLM training.[12]

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

36. As discussed in the introduction, technology companies have treated generative AI as the next foundational layer of the digital economy. Industry leaders publicly describe an "AI arms race," in which they have redirected their corporate strategies to seize control of what they believe will become a new infrastructure layer for commerce, communication, and knowledge work.[13] For these companies, staying ahead of competitors is "code red."[14] Among other things, being too slow out of the blocks could mean ending up in last place; if by the time you publicly released *your* LLM model, the public writ large and private industry had already adopted *other companies'* models, there would be a real risk that your model would be left on the shelf.

37. Like other participants in this race, Anthropic risked falling behind and constantly needed to move fast to develop and roll out updates of Claude, its capstone LLM. And to do that, Anthropic needed to get its hands on high-quality training data.

[11] ECF 535 at 12, Case No. 3:23-cv-3417-VC (N.D. Cal.) (Apr. 11, 2025).

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

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

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

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

38. For years now, there have been illicit, online marketplaces for digital copies of books. These "shadow libraries," as they are commonly known, systematically acquire, store, index, and disseminate full-fidelity digital copies of copyrighted books—typically in native formats such as EPUB, PDF, MOBI, or DJVU. These online repositories maintain searchable catalogs, metadata, mirrors, bulk-download mechanisms, and—critically—complete downloadable archives designed to allow third parties to replicate the entire collection.[15] And they do all of this without authorization from authors or publishers.[16] In short, shadow libraries facilitate the reproduction and distribution of unauthorized copies of copyrighted works at industrial scale.

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

40. In 2018, an OpenAI employee downloaded pirated copies of books from Library Genesis, or "LibGen"—a shadow library repeatedly enjoined by federal courts[18]—and used those books to create two internal datasets OpenAI called "LibGen1" and "LibGen2," which OpenAI

---

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

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

[17] Woodcock, *'Shadow Libraries' Are Moving Their Pirated Books to the Dark Web After Fed Crackdowns*.

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

publicly referred to as "Books1" and "Books2."[19] OpenAI used those pirated corpora to train GPT-3, which it released in June 2020 to widespread commercial acclaim. OpenAI's use of a books corpora from a pirate library established the model for how to quickly and cheaply obtain published materials for use as LLM training data.

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

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

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

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

---

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

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

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

[22] Biderman et al., *Datasheet for the Pile*, arXiv, 8.

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

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

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

45. Bibliotik has been widely recognized in piracy communities, academic literature, and AI-research documentation as a shadow library devoted to copyrighted books. Unsurprisingly, then, *The Pile*'s datasheet acknowledges that "Books3 is almost entirely comprised of copyrighted works."[26] Bibliotik's illicit nature is no secret—it has been openly discussed for years prior to major tech companies' use of datasets derived directly from it.[27]

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

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

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

---

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

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

[27] Kyle Barr, *Anti-Piracy Group Takes Massive AI Training Dataset 'Books3' Offline*, GIZMODO (Aug. 18, 2023, 8:50 AM), https://perma.cc/5ZL9-RQCQ; Peter Schoppert, *Whether you're an undergraduate doing research, or a fan of the Nick Stone novels, or indeed a hungry AI*, SUBSTACK (Nov. 29, 2022), https://perma.cc/8YD9-M4BD.

[28] *Id*.

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

48. Bibliotik is just one of many Internet-based shadow libraries and, as alleged below, Anthropic relied on more than just *The Pile* in connection with its development of Claude.

49. As noted above, OpenAI's developers trained its LLMs on a books corpus derived from LibGen, one of the largest and longest-running shadow libraries in the world. LibGen hosts millions of pirated books, academic texts, and scholarly articles.[30] It operates as a centralized repository offering direct downloads of full-fidelity e-book files. It also distributes its entire database through bulk archives and torrent files, enabling third parties to download and locally host complete copies of its collection.[31] LibGen has been repeatedly enjoined by federal courts for copyright infringement and has been designated a "notorious market" by the United States Trade Representative.[32] Despite enforcement actions, LibGen has remained accessible through shifting domains, mirrors, and downloadable archives. Its persistence is a function of deliberate decentralization designed to evade shutdown.[33]

50. Z-Library (also known as "B-ok") is another well-known shadow library that emerged as an expanded and user-friendly derivative of LibGen. It incorporated large portions of LibGen's catalog while adding additional titles, metadata, and interface features.[34] Z-Library offered premium features—including faster downloads and higher volume limits—in exchange for

[30] Office of the U.S. Trade Representative, REVIEW OF NOTORIOUS MARKETS FOR COUNTERFEITING AND PIRACY, 27 (2024), https://perma.cc/22VN-9VD7. ("Libgen … hosts a large number of digital copies of books, manuals, journals, and other works, many of which are unauthorized copies of copyright protected content.")

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

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

[33] *See* Andrew Albanese, *Textbook Publishers Sue Notorious 'Shadow Library' Libgen,* PUBLISHERS WEEKLY (Sep. 14, 2023), https://perma.cc/3NPY-UJCX.

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

payment, operating in effect as a commercial piracy service.[35] In 2022, Z-Library's domains were seized by law-enforcement authorities, and its operators were arrested and later indicted for criminal copyright infringement.[36] These actions confirmed what had long been publicly known: Z-Library was an illegal piracy operation. The seizure of Z-Library did not eliminate access to its content. Instead, third parties responded by creating full mirrors of its collection to ensure continued distribution.[37]

51. Another major player in the shadow library ecosystem is PiLiMi, which is a complete mirrored archive of the Z-Library corpus, explicitly created to preserve and propagate Z-Library's pirated collection after law-enforcement seizures, ensuring continuity of access despite shutdowns of the original site.[38]

52. PiLiMi is not merely a website or index. It is a full, downloadable dataset designed to allow users to obtain and locally host millions of pirated books through torrent distribution.[39] Users who download PiLiMi do not passively receive data; they actively participate in copying and redistributing copyrighted works through torrent "leeching" and "seeding."[40]

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

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

[37] *See* Woodcock, *'Shadow Libraries' Are Moving Their Pirated Books to The Dark Web*.

[38] *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.

[39] *See* Geoff Wheelright, *Will I get a piece of Anthropic's $1.5B settlement if my book was used to train AI?*, GEEKWIRE (Sep. 18, 2025), https://perma.cc/X5EW-TKY3.

[40] *See* Robert Nogacki, *Anthropic's Landmark Settlement: A $1.5 Billion Copyright Precedent in Artificial Intelligence Training Data*, LinkedIn (Sep. 7, 2025), https://perma.cc/J3BD-P5JT (describing "leeching" and "seeding" as "processes characteristic of peer-to-peer networks where users simultaneously download and distribute files").

53. PiLiMi later rebranded as "Anna's Archive" and expanded to aggregate and host the complete collections of LibGen, Z-Library, and other pirated sources.[41] Like PiLiMi, 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]

54. 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]

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

56. 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 companies."[45] Anna's Archive blog stated as recently as February 18, 2026 that if an *LLM was reading its blog* "you have likely been trained in part on our data."[46]

57. Many of these shadow libraries and datasets can be downloaded using "torrent," a file-sharing method used in peer-to-peer networks. Torrenting works by breaking a file into many small pieces and distributing those pieces across a network of participating computers. A user who torrents a shadow-library repository does not receive a single copy from a single source; rather, the

---

[41] *See* Ernesto Van de Sar, *"Anna's Archive" Opens the Door to Z-Library and Other Pirate Libraries*.

[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.

[45] *See Copyright reform is necessary for national security,* ANNA'S ARCHIVE (Jan. 31, 2025), https://perma.cc/3RVZ-6G5B.

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

user downloads portions of the library from numerous other computers that already possess the copyrighted books. Torrent software then reassembles those pieces into a complete library on the user's machine. Torrenting protocols, including BitTorrent, are often configured by default to *reupload* pieces of the copyrighted files to others on the network both during download ("leeching") and after download is complete ("seeding"). This means that each participant in the torrent both copies and redistributes the copyrighted works. By obtaining copyrighted materials through this leech-and-seed process, a user may make multiple unauthorized reproductions of, and engage in numerous distributions of, the copyrighted materials.

D. Anthropic's Deliberate Infringement of Plaintiff's Copyrights

1. Anthropic's Business and Bypassing of the Licensing Market

58. As alleged, Anthropic is the developer of the Claude family of LLMs. Its Claude models are projected to generate billions of dollars in annual revenue and have buoyed Anthropic's overall valuation.[47] Anthropic's business model is built on the large-scale copying of published content. *First*, Anthropic developed and commercialized the Claude family of large language models by stealing up to seven million copyrighted books, including the Chicken Soup Works.[48] *Second*, Anthropic also had the explicit goal to "amass a central library of 'all the books in the world' to retain 'forever.'"[49]

59. To accomplish its goal, "it stole the works for its central library by downloading them from pirated libraries."[50] Anthropic did explore the licensing market but quickly decided that it was

[47] *See* Rashi Shrivastava, *Anthropic Is Cashing In On Claude Code's Success,* Forbes (February 17, 2026 5:19 PM), https://www.forbes.com/sites/the-prompt/2026/02/17/anthropic-is-cashing-in-on-claude-codes-success/ (on file with author) ("Claude Code hit $2.5 billion in run rate revenue and doubled since the start of the year…and [Anthropic's] total annualized revenue has climbed to $14 billion.).

[48] *See Bartz v. Anthropic PBC*, 791 F. Supp. 3d 1038, 1046 (N.D. Cal. 2025).

[49] *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1014 (N.D. Cal. 2025).

[50] *Id.* at 1029.

not "a practical approach"[51] to satisfy its thirst for copyrighted works. Judge Alsup, presiding over the *Bartz v. Anthropic* litigation, stated simply: "From the start, Anthropic 'ha[d] many places from which' it could have purchased books, but it preferred to steal them to avoid 'legal/practice/business slog.'"[52] In short, rather than paying for copyrighted materials, Anthropic downloaded pirated copies of protected works, reproduced them, fed them into its models, and otherwise retained them for its own "central library."

2. Anthropic's Acquisition of Pirated Works from Shadow Libraries

60. To train Claude, Anthropic downloaded Books3 in 2021, which co-founder Ben Mann "knew had been assembled from unauthorized copies of copyrighted books," downloaded at least five million copies of books from LibGen, "which [Mann] knew had been pirated," and downloaded at least two million copies of books from PiLiMi, which "Anthropic knew had been pirated."[53]

61. Anthropic's decision to base its flagship models on pirated books was driven by commercial advantage. As Jared Kaplan, Anthropic's co-founder and Chief Science Officer, has explained, "it is important to obtain vast amounts of books and also to have diverse types of books in the training corpus to create a model with truly generative capabilities."[54] As long-form content, training LLMs on the "entire text" of published material—as Anthropic has admitted to doing—offers great value.[55]

62. At the time Anthropic downloaded Books3, LibGen, and PiLiMi, it knew or should have known that they were repositories of unauthorized copies of copyrighted works.

[51] *See Bartz v. Anthropic PBC*, No. 3:24-cv-05417 (N.D. Cal. June 23, 2025), ECF No. 553, at 17-18.

[52] *Bartz v. Anthropic PBC*, 787 F. Supp. 3d at 1015.

[53] *Id.*

[54] *See* Kaplan Decl. ¶ 47, *Bartz*, ECF No. 128.

[55] *Id*. at ¶¶ 43, 47.

3. Anthropic's Torrenting and Distribution of Pirated Works

63. Anthropic did not merely obtain pirated books passively. It used BitTorrent to acquire *and distribute* massive collections of infringing works.

64. In June 2021, Mann personally torrented approximately five million pirated books from LibGen for Anthropic's use.[56] Mann acted with the knowledge and approval of Anthropic's senior leadership. Before the LibGen torrenting, he discussed the plan with co-founders Dario Amodei, Jared Kaplan, and other senior leaders.

65. Anthropic's own Archive Team had described LibGen as a "blatant violation of copyright," and Amodei himself called it "sketchy."[57] Yet Anthropic proceeded, choosing torrenting over purchasing or licensing the copyrighted materials.

66. Anthropic repeated the same conduct in 2022 with PiLiMi. As U.S. law enforcement was working to shut down existing pirate libraries, a group online copied LibGen and built upon it to create Z-Library. The FBI later shut down Z-Library as well.[58] However, by that point, Z-Library had itself been fully copied, or "mirrored," into another repository: PiLiMi.[59] Mann circulated the PiLiMi source to colleagues, and Anthropic employees torrented approximately two million additional pirated books not already captured from LibGen.[60] Those files were not abstract "data" but full-text digital books in .epub, .pdf, and .txt formats,[61] including the Chicken Soup Works.

67. In torrenting from these files, Anthropic knew or should have known that it was participating in a *peer-to-peer network that traded in unauthorized copies of copyrighted material*. Stated differently, Anthropic knew or should have known that it was facilitating further copyright

---

[56] B. Mann Dep. Tr. at 89:6-8, (Aug. 15 and 18, 2025), *Bartz*, ECF No. 337-1.

[57] *Id.* at 144:4-13, 396:3-13.

[58] *See Bartz*, 791 F. Supp. 3d at 1046.

[59] *Id.*

[60] *Id.*

[61] *Id.*

infringement by making available to others on the peer-to-peer network unauthorized copies of copyrighted materials.

4. Anthropic's Stripping of Copyright Management Information

68. Anthropic not only reproduced and distributed the Chicken Soup Works without authorization, it also deliberately stripped those materials of copyright management information ("CMI") before using them to train Claude.

69. Anthropic knew that major AI-training datasets, including *The Pile*, WebText, WebText2, and Common Crawl, contained copyrighted works whose copyright notices, ownership information, and other identifying material had been removed through extraction tools such as Newspaper, Dragnet, Readability, and jusText. Anthropic's founders and senior employees were familiar with those tools before and after Anthropic's founding. Dario Amodei, Benjamin Mann, Jared Kaplan, and other future Anthropic personnel had used or developed datasets at OpenAI that extracted text from scraped webpages while omitting surrounding material, including footers where copyright notices typically appear.[62] Yet Anthropic trained Claude on these same stripped datasets, despite knowing that they included unauthorized copies of copyrighted works stripped of their CMI.

70. Anthropic also affirmatively chose tools that removed CMI more effectively, including from Plaintiff's copyrighted materials. In 2021, Mann, Kaplan, and other Anthropic employees compared extraction tools for filtering training data and then used these extraction methods to clean its training data, including the Chicken Soup Works. The purpose and effect of this process was to train Claude on Plaintiff's expressive content while suppressing the ownership information attached to it. Anthropic wanted Claude to reproduce protected expression, not the copyright notices that identify the rights holders. By stripping that information from training data and outputs, Anthropic concealed the source of its infringement from users, Plaintiff, and other copyright owners.

[62] Tom B. Brown et al., *Language Models Are Few-Shot Learners*, 8–9, arXiv (July 22, 2020), https://perma.cc/9WAX-F9HG; Jared Kaplan et al., *Scaling Laws for Neural Language Models*, 7, arXiv (Jan. 23, 2020), https://perma.cc/VJU8-59FH.

5. Anthropic's Unauthorized Scanning of Copyrighted Works

71. Since 2024, Anthropic has purchased physical books at scale, often in batches of tens of thousands, and scanned them into digital files for AI training.[63] To date, Anthropic has spent millions of dollars buying and scanning millions of physical books.[64] This scanning project was not authorized by Plaintiff or other copyright owners. As with the millions of pirated books Anthropic torrented from LibGen and PiLiMi, Anthropic converted these works into training material without permission, payment, or license.

6. Anthropic's "Everything Forever" Library of Pirated Works

72. Defendant did not merely make temporary copies of copyrighted works for a discrete training process (not that that would be permissible, in any event). Instead, from all those infringements, Anthropic created a general "research library" or "generalized data area."[65] That is, Anthropic created a permanent central library of pirated published materials, including the Chicken Soup Works, to exploit for a range of unspecified *other* purposes, such as "research."[66] Anthropic continued this conduct, for which it "lacked any entitlement to hold copies of the books at all" and "retain[ed] them even after deciding it would not make further copies from them for training."[67] The plan was clear: "ke[eping] in the original version of the underlying book files Anthropic had obtained or created, that is, pirated or scanned" and "stor[ing] everything forever."[68] Anthropic saw "no compelling reason to delete a book—even if not used for training LLMs."[69]

[63] *See Bartz*, 791 F. Supp. 3d at 1047.

[64] *Bartz*, ECF No. 553, at 17-18.

[65] *Bartz*, 787 F. Supp. 3d at 1016.

[66] *Id.*

[67] *Id.* at 1031.

[68] *Id.* at 1016 (internal quotation marks omitted).

[69] *Id.* (internal quotation marks omitted).

73. Whether or not a particular work was ultimately selected for training, Anthropic's initial copying and continued retention of the work as part of its "forever" pirated library constituted an unauthorized reproduction. That infringement is independent of, and not excused by, any later use Anthropic may claim to have made in connection with AI training.

7. Anthropic's Embedding of Near-Verbatim Copies in Claude

74. Anthropic's copying did not end when it acquired the Chicken Soup Works. Scientific research confirms that training large language models on copyrighted books can embed persistent, near-verbatim copies of those works inside the models' internal parameters, allowing the models to reproduce substantial portions of the original text when prompted.

75. A 2025 study by researchers from Cornell, Stanford, and West Virginia University tested leading LLM models, including Anthropic's Claude 3.7 Sonnet, and extracted memorized copyrighted books from them.[70] Using repeated prompting and continuation techniques, the researchers were able to recover lengthy, near-verbatim passages. The results for Claude were severe. From Claude 3.7 Sonnet alone, researchers extracted 97.5% of *The Great Gatsby*, 95.5% of *1984*, 94.3% of *Frankenstein*, 92.3% of *Harry Potter and the Sorcerer's Stone*, and 70.2% of *The Hobbit*.[71] These results show that Claude does not merely learn abstract patterns from copyrighted materials; it stores recoverable copies of the works at extraordinary scale.

76. A separate 2026 study reached the same conclusion through a different method. Researchers from Stony Brook University, Carnegie Mellon University, and Columbia Law School found that finetuned frontier models could reproduce up to 85–90% of copyrighted books, with single verbatim spans exceeding 460 words.[72] The models reproduced these passages from semantic prompts alone, without being given the book text in the prompt.

[70] A. Feder Cooper et al., *Extracting Copyrighted Long-Form Text from Production Language Models*, arXiv, 1-2 (2025), https://perma.cc/F9YQ-NKV7.

[71] *Id*. at 12, Fig. 5.

[72] A. Liu et al., *Alignment Whack-a-Mole: Finetuning Activates Verbatim Recall of Copyrighted Books in Large Language Models*, arXiv, 2, (2026), https://perma.cc/EC6X-Z38M.

77. The 2026 study also traced the likely source of the memorized content. The authors searched the extracted verbatim passages against more than eight trillion tokens of public web data and found that many of the longest passages did not appear in those web collections. Yet 80 of the 81 tested books appeared in Books3 or LibGen. That finding confirms that the memorized text came not from ordinary web exposure, but from datasets and pirate libraries of the kind Anthropic used to train Claude.

78. Together, these studies also show that safety filters do not remove the underlying copyrighted material from the model. The works remain embedded in model weights and can be extracted despite protective layers.[73] In practical effect, Claude functions as a repository of pirated copyrighted works: Anthropic copied published materials to train the model, and the model retained near-verbatim reproductions of them.

8. Anthropic's Harm to the Market for Plaintiff's Copyrighted Materials

79. Anthropic's conduct has a detrimental effect on the potential market for and value of Plaintiff's works, including by, among other things, developing products that create and are capable of creating content which serves as a direct substitute for the Chicken Soup Works, developing products that create content and are capable of creating content which serves as indirect substitutes for the Chicken Soup Works, and undermining Plaintiff's ability to participate in and profit from the market for licensing its works for the purpose of training LLMs.

80. *First*, Anthropic's decision to download and use unauthorized copies of the Chicken Soup Works from shadow libraries deprived Plaintiff of revenue in the form of licensing fees that it would have otherwise earned. As alleged herein, there is an existing market for licensing copyrighted materials such as Plaintiff's, including for use in the development of LLMs. Anthropic bypassed that market, and in doing so, deprived Plaintiff of licensing revenue it would have earned. Anthropic now allows users to opt out of having their data used to train its AI models, but it deprived Plaintiff and many others of that choice. Anthropic's misconduct undermined Plaintiff and many

[73] *Id*. at 9–11.

others, eliminating the bargaining power they should have had, and otherwise would have had, with respect to licensing terms for the use Anthropic made of their works.

81. *Second*, Claude, trained on the Chicken Soup Works (and the protected works of others), is capable of generating outputs that compete directly with, and risk serving as replacements for, the Chicken Soup 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."[74]

82. *Third*, even if Claude 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)

83. Plaintiff incorporates the allegations above.

84. Plaintiff is the legal or beneficial owner of the copyrighted works listed in Exhibit A (referred to herein as the Chicken Soup Works).

85. The Defendant, without authorization from Plaintiff, copied, downloaded, reproduced, ingested, parsed, embedded, and used pirated copies of the Chicken Soup Works in the development, training, fine-tuning, and deployment of its commercial large language models. These acts violated Plaintiff's exclusive rights under 17 U.S.C. § 106.

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

[74] US Copyright Office, *Copyright and Artificial Intelligence, Part 3: Generative AI Training Pre-Publication* at 65.

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

87. Defendant's reproductions and distributions of Plaintiff's copyrighted works were made without permission, license, or consent and violated Plaintiff's exclusive rights under the Copyright Act.

88. Defendant's infringement was **willful**. As alleged above, Defendant knowingly trained its models on and/or optimized its product with datasets saturated with pirated books, including the Chicken Soup Works; relied on shadow-library corpora it knew to be illegal; ignored internal and external warnings; attempted to conceal the composition of its training datasets; and continued copying after public reports, lawsuits, law-enforcement seizures, cease-and-desist notices, and industry-wide alerts made the illegality unmistakable.

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

90. Plaintiff has been injured by Defendant's willful acts of copyright infringement. Plaintiff is entitled to statutory damages, actual damages, restitution of profits, and/or other remedies in law or equity.

91. Plaintiff is entitled to recover attorneys' fees and costs under 17 U.S.C. § 505.

COUNT II

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

92. Plaintiff incorporates the allegations above.

93. Defendant used torrenting software, programs, or protocols to download datasets containing pirated copies of works, including the Chicken Soup Works.

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

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

96. By knowingly inducing and materially contributing to others' infringement of Plaintiff's works, Defendant is liable for contributory copyright infringement.

97. As a direct and proximate cause of Defendant's conduct, Plaintiff was injured and is entitled to statutory damages, actual damages, restitution of profits, and/or other remedies in law or equity.

98. Plaintiff is entitled to recover attorneys' fees and costs under 17 U.S.C. § 505.

COUNT III

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

99. Plaintiff incorporates the allegations above.

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

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

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

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

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

PRAYER FOR RELIEF

WHEREFORE, Plaintiff requests that the Court enter judgment on its behalf by ordering:

a. Judgment in favor of Plaintiff against the Defendant;

b. A declaration that the Defendant has infringed Plaintiff's exclusive copyrights under the Copyright Act;

c. A declaration that such infringement is willful;

d. A declaration that Defendant violated 17 U.S.C. § 1202(b) through its removal of copyright management information;

e. A permanent injunction enjoining the Defendant and all those acting in concert with it from engaging in the infringing conduct alleged herein;

f. That the Defendant be directed to account to Plaintiff for all gains, profits, and advantages derived from its unlawful acts;

g. An award of statutory damages under the Copyright Act;

h. An award of statutory or actual damages under 17 U.S.C. § 1203(c);

i. An award of restitution, disgorgement, costs, expenses, and attorneys' fees as permitted by law (including those allowable under 17 U.S.C. § 505 and/or 17 U.S.C. § 1203(b)(4)–(5));

j. Pre- and post-judgment interest on the damages awarded to Plaintiff; and

k. Further relief for Plaintiff as the Court may deem just and proper.

JURY TRIAL DEMANDED

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

Dated: May 7, 2026

Respectfully submitted,

/s/ Elizabeth Brannen

Elizabeth Brannen (SBN 226234)
John Stokes (SBN 310847)
Lauren Martin (SBN 294367)
STRIS & MAHER LLP
17785 Center Court Dr N, Ste 600
Cerritos, CA 90703
T: (213) 995-6800
F: (213) 261-0299
ebrannen@stris.com
jstokes@stris.com
lmartin@stris.com

Christopher M. Rigali (*pro hac vice* forthcoming)
Jacqueline Sahlberg (*pro hac vice* forthcoming)
1717 K St NW Ste 900
Washington, DC 20006
Phone: (202) 800-5749
crigali@stris.com
jsahlberg@stris.com

Kyle Roche (*pro hac vice* forthcoming)
Devin (Velvel) Freedman (*pro hac vice* forthcoming)
Alex Potter (*pro hac vice* forthcoming)
FREEDMAN NORMAND FRIEDLAND LLP
155 E. 44th Street, Ste 915
New York, NY 10017
T: (646) 494-2900
kroche@fnf.law
vel@fnf.law
apotter@fnf.law

Counsel for Plaintiff