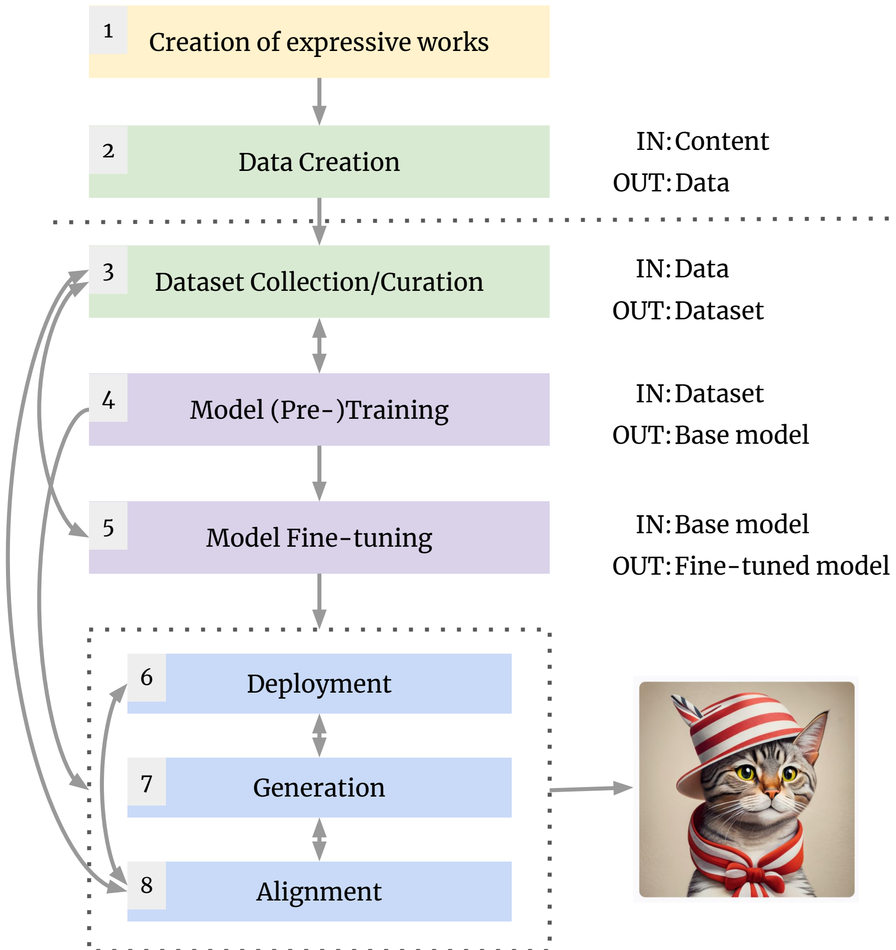

Figure 4: The generative-AI supply chain. We map out eight different stages: 1) The creation of expressive works, (*see infra* Part I.C.1), 2) data creation (*see infra* Part I.C.2), 3) dataset collection/curation (*see infra* Part I.C.3), 4) model (pre-)training (*see infra* Part I.C.4), 5) model fine-tuning (*see infra* Part I.C.5), 6) system deployment (*see infra* Part I.C.6), 7) generation (*see infra* Part I.C.7), and 8) model alignment (*see infra* Part I.C.8), Different stages are connected to each other, handing off outputs from one stage as inputs to another. The creation of expressive works and data creation pre-date the advent of today's generative-AI systems (indicated by a dotted line). There are many possible ways to connect the other six stages. System deployment, model alignment, and generation tend to happen in concert (indicated by the dotted box). Generations can in turn be used as training data (*see infra* Part I.C.7). We indicate this in the figure with the arrow from generation (7) to dataset collection/curation (3). In this case, generation serves simultaneously as the creation of expressive works (1) and data creation (2).

niques, like input-prompt and output-generation filtering,[255] in generative-AI systems.

II. Tracing Copyright Through the Supply Chain

The hornbook statement of United States copyright doctrine is that original works of authorship are protected by copyright when they are fixed in a tangible medium of expression.[256] A defendant directly infringes when they engage in conduct implicating one of several enumerated exclusive rights (reproducing, publicly distributing, etc.),[257] with a work of their own that is substantially similar to a copyrighted work[258] because it was copied from that work.[259] Other parties may be held secondarily liable for conduct that bears a sufficiently close nexus to the infringement under one of several theories.[260] Otherwise infringing conduct is legal when it is protected by one of several defenses, including the DMCA Section 512 safe harbors,[261] fair use,[262] or an express[263] or implied[264] license. In addition, we consider conditions for which different remedies may be granted when courts find infringement:[265] damages and profits, statutory damages, attorney's fees, injunctions, and destruction of generative-AI models.[266]

This Part applies this orthodox, uncontested statement of copyright law to the generative-AI supply chain.[267] It takes up these issues in the above order — the same logical order that they typically arise in an copyright lawsuit — to analyze the copyright implications of each link in the supply chain. Our goal is to be careful and systematic, not to say anything dramatically new.

255. *See supra* Part I.C.7.

256. *See infra* Part II.A.

257. *See infra* Part II.B.

258. *See infra* Part II.C.

259. *See infra* Part II.D.

260. *See infra* Part II.E (direct infringement); *infra* Part II.F (indirect infringement).

261. *See infra* Part II.G.

262. *See infra* Part II.H.

263. *See infra* Part II.I.

264. *See infra* Part II.J.

265. *See infra* Part II.K.

266. We do not consider paracopyright liability, which attaches to the intentional removal, alteration, or forgery of copyright management information with the intent to facilitate infringement. Nevertheless, this, too, likely has potential ramifications for the generative-AI supply chain.

267. *See supra* Part I.C.

A. Authorship

Copyright protects "(1) original works of authorship (2) fixed in any tangible medium of expression."[268] "Original, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it possesses at least some minimal degree of creativity."[269] Fixation is satisfied when the work is embodied in a tangible object in a way that is "sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration."[270]

We start with fixation. Unfixed works have no interaction with the generative-AI supply chain. A work must be fixed to be used as training data. Truly ephemeral creations, like unobserved dances and songs that are never recorded, will never be captured in a way that can be used as an input to a training algorithm. Datasets, models, applications, prompts, and generations are all fixed in computers and storage devices.

Once it is fixed, however, any kind of original expression can be used as inputs for generative AI. Copyrightable subject matter explicitly includes "literary works" (e.g. poems, novels, FAQs, and fanfic),[271] "musical works" (e.g., sheet music and MIDI files)[272] "pictorial . . . works" (e.g. photographs),[273] "audiovisual works" (e.g., Hollywood movies and home-recorded TikToks),[274] "sound recordings" (e.g., pop songs and live comedy recordings),[275] and more. But this list is nonexclusive. Any kind of creative expression that appeals to the eye or the ear is copyrightable.[276] And copyright law does not discriminate among works based on their quality, their morality, or their importance.[277]

Instead, the originality requirement distinguishes material that was created by a human author from facts that "do not owe their origin to an act of authorship."[278] In addition, some types of material are never copyrightable, including any "idea, procedure, process, system, method of operation, con-

268. 17 U.S.C. § 102(a) (numbering added).
269. Feist Publ'ns v. Rural Tel. Serv. Co., 499 U.S. 340, 345 (1991).
270. 17 U.S.C. § 101 (definition of "fixed").
271. 17 U.S.C. § 102(a)(1).
272. *Id.* § 102(a)(2).
273. *Id.* § 102(a)(5).
274. *Id.* § 102(a)(6).
275. *Id.* § 102(a)(7).
276. Christopher Buccafusco, *Making Sense of Intellectual Property Law*, 97 CORNELL L. REV. 501 (2012).
277. Bleistein v. Donaldson Lithographing Co., 188 U.S. 239, 251 (1903).
278. Feist Publ'ns v. Rural Tel. Serv. Co., 499 U.S. 340, 347 (1991).

cept, [or] principle."[279] In practice, this means that the copyright in some works (e.g., product photographs) will be "thinner" and protect fewer aspects of the works than the "thicker" copyrights in others (e.g, abstract art), because the "range of creative choices that can be made in producing the works is narrow."[280] In particular, any copyright in computer software — which is treated as a "literary work" for copyright purposes — typically excludes a great deal of functional material, such as efficient algorithms or coding conventions required by the choice of programming language.[281]

Data

As a result, some of the individual examples that serve as training data[282] are uncopyrightable. (For example, birdsong-recognition AIs are trained on recordings of birds.[283]) But other items are copyrightable, and those copyrights will be held by a variety of authors: photographers, writers, illustrators, musicians, programmers, and other creators of all stripes.

Training Datasets

Moving forward along the supply chain, then, different datasets[284] will include different amounts and proportions of copyrighted material. A dataset of birdsong recordings will be entirely, or almost entirely, copyright-free. A dataset of illustrations, on the other hand, will contain numerous copyrighted works.

Datasets *themselves* may be copyrightable as **compilations**,[285] "formed by the collection and assembling of preexisting materials or of data."[286] A compilation is copyrightable (separately from any copyright in the works it is assembled from) when the compilation itself features a sufficiently original "selection or arrangement."[287] Originality in selection is choosing *what to*

---

279. 17 U.S.C. § 102(b).

280. Rentmeester v. Nike, Inc., 883 F.3d 1111, 1120 (9th Cir. 2018).

281. Pamela Samuelson, *Functionality and Expression in Computer Programs: Refining the Tests for Software Copyright Infringement*, 31 Berkeley Tech. L.J. 1215 (2016).

282. *See supra* Part I.C.2.

283. *See* Stefan Kahl, Connor M. Wood author & Holger Klinck, *BirdNET: A Deep Learning Solution for Avian Diversity Monitoring*, 61 Ecological Informatics 101236 (2021). Animals are not recognized as "authors" for copyright purposes. *See* Naruto v. Slater, 888 F.3d 418 (9th Cir. 2018).

284. *See supra* Part I.C.3.

285. 17 U.S.C. § 103(a).

286. § 101 (definition of "compilation").

287. Feist Publ'ns v. Rural Tel. Serv. Co., 499 U.S. 340, 348 (1991).

include in the dataset; originality in arrangement is choosing *how to organize* the dataset. Every dataset is based on extensive curation,[288] but in some cases it is easier to identify the specific choices that went into intentionally creating a dataset with particular desired attributes. The LAION-Aesthetics dataset, for example, was created by training a discriminative model[289] to predict the ratings that humans gave images, and then using the model to select "high visual quality" images from a much larger dataset.[290]

Pre-Trained/Base Models

Attributing authorship for models is trickier to classify for two reasons.[291] First, there is the question of whether a model possesses the necessary "modicum of creativity" to be a work of authorship at all.[292] In some cases, the answer is probably "no": applying an existing algorithm and well-known architecture to an existing dataset[293] does not involve sufficient creative choices. Any expression in such a model merges into the idea and is uncopyrightable.[294] But it is possible that other models are works of authorship. For one thing, when a training dataset is curated specifically for training a base model, the model may supplant the dataset as the relevant 'work' from the data curation process, just as a finished film is regarded as the 'work' rather than the (much larger) dataset of raw footage.[295] In such a case, the model would inherit the creative choices that went into curating the dataset. For another, base models are often the results of extensive design processes that involve novel architectures and algorithms. While these processes are not themselves copyrightable,[296] and originality in a process is not a guarantee that the out-

288. *See supra* Part I.C.3. *See generally* Katherine Lee, Daphne Ippolito & A. Feder Cooper, The Devil is in the Training Data (2023) (unpublished manuscript), *in* Lee, Cooper, Grimmelmann & Ippolito, *supra* note 59, at 5.

289. *See supra* Part I.A.2a.

290. Christoph Schuhmann, *LAION-Aesthetics*, LAION (Aug. 16, 2022), https://laion.ai/blog/laion-aesthetics/.

291. It is worth noting that many model trainers creators certainly believe that models are copyrightable, and have released those models under licenses that are only intelligible if there is something copyrightable to license in the first place.

292. *Feist Publ'ns*, 499 U.S. at 346.

293. With standard choices of hyperparameters, on standard hardware, etc.

294. *See generally* Pamela Samuelson, *Reconceptualizing Copyright's Merger Doctrine*, 63 J. Copyright Soc'y USA 417 (2016) (describing merger doctrine).

295. *See generally* Margot E. Kaminski & Guy A. Rub, *Copyright's Framing Problem*, 64 UCLA L. Rev. 1102 (2017) (discussing problem of identifying the 'work' in copyright cases).

296. *See* 17 U.S.C. § 102(b).

puts are copyrightable,[297] in some cases, a model's creators[298] will have made creative choices that imbue the model with copyrightable expression.

The second way in which the copyrightability of models is tricky is that they could be described in several different ways under copyright doctrine. One view is that a model is a compilation of its training data — the model is simply a different and complicated arrangement of training examples. Another view is that a model is a **derivative work** of its training data — "a work based upon one or more preexisting works . . . in which [those works are] recast, transformed, or adapted."[299] A derivative work (think of a translation of a novel, a recording of a song, or an action figure based on a character from a movie) combines the authorship in an existing (or "underlying") work with new authorship. The substantive difference between the two is that in a compilation, the underlying works are present in substantially unmodified form, whereas in a derivative work the underlying work is "recast, transformed, or adapted." The line dividing the two characterizations is somewhat metaphysical, but it has consequences in some corners of copyright doctrine, which could in turn have consequences for pre-trained models.[300]

Fine-Tuned Models and Aligned Models

Both of the authorship considerations that we raise above for pre-trained models also apply to fine-tuned and aligned models. We start with the second point, which is simpler: like pre-trained models, both fine-tuned and aligned models will face similar issues of categorization for copyright law. A fine-tuned and/or aligned model will typically be a derivative work of the base model it was trained from.

The first point — that training choices can imbue models with creative attributes — leads to different observations for fine-tuning and model alignment. There is an argument to be made that fine-tuning is, by definition, a creative process. The model trainer is typically optimizing the model's behavior in generating specific desired outputs — the kind of nexus between human choices and resulting material that characterizes copyrightable au-

---

297. *See* James Grimmelmann, *Three Theories of Copyright in Ratings*, 14 Vand. J. Ent. & Tech. L. 851, 878–79 (2011) (criticizing theory that outputs "resulting from a minimally creative process" are thereby copyrightable).

298. In this case, this includes the parties that designed the architectures and algorithms.

299. 17 U.S.C. § 101 (definition of "derivative work).

300. *See, e.g.,* § 203(b)(1) (allowing the creator of an authorized derivative work to continue using it after the author terminates the license in accordance with a statutory procedure).

thorship.[301] The same is true for model alignment. Further, if, for example, the prompt is incorporated as part of the input to RLHF,[302] then the prompt serves as training data that could update the model. In this case, said training data itself is created in a process that includes human choices and has been crafted with specific creative goals in mind.

The prompt, though considered an input to generation, raises additional authorship considerations for both fine-tuning and alignment. As discussed above, when the user of the service supplies a prompt to a generative-AI system, the service host may save that prompt for later use. The service host may use the prompt as additional training data for fine-tuning or aligning the existing model, or for training another model altogether.[303] As a result, fine-tuning and alignment are stages in the supply chain during which copyrighted data can find its way into a generative-AI system — where either the user of the service is the copyright holder, or they have prompted with content for which another entity is the copyright holder. For example, it is currently technologically feasible to prompt a text-to-text system with an entire book.[304] It may be possible to implement content filters to catch known copyrighted material and remove it from training and alignment data, but such implementation considerations typically fall within other aspects of the generative-AI system, rather than the model.[305] Additionally, there could be an express[306] or implied license[307] for user-inputted data, for the cases in which the user of the service is the copyright holder. There are also separate considerations for infringement and safe harbors, which we address below.[308]

301. *See generally* Dan L. Burk, *Thirty-Six Views of Copyright Authorship, by Jackson Pollock*, 58 Hous. L. Rev. 263 (2020) (discussing causal elements of authorship); Shyamkrishna Balganesh, *Causing Copyright*, 117 Colum. L. Rev. 1 (2017) (same).

302. *See supra* Part I.C.8.

303. *See supra* Part I.B.7; *supra* Part I.B.8

304. Anthropic, *supra* note 97.

305. *See infra* note 408 and accompanying text (for a discussion of the challenges of identifying copyrighted data); *infra* note 607 and accompanying text (for a discussion of Copilot's output filters).

306. *See infra* Part II.I.

307. *See infra* Part II.J.

308. *See infra* Part II.E; *infra* Part II.F; *infra* Part II.G

Deployed Services

It is well-established that software is copyrightable.[309] The non-model parts of a user-facing application or developer API will be protected by copyright (subject to the functionality screen noted above). Also, as noted above, it is also possible for content filters to be implemented within the overarching generative-AI system that is hosted in the service. It is at this stage of the supply chain where such filters could, for example, choose not to store user-inputted prompts.

Generations

Generations raise a doctrinal question that has been debated for decades: who, if anyone, owns the copyright in the output of a computer program?[310] Although some commentators have argued that the program itself should be regarded as the author, computer authorship is squarely foreclosed by U.S. copyright law.[311] Computers are not capable of playing the social roles that society and the legal system expect and require of authors.[312] So far, the courts have held firm to this line for AI generations. In *Thaler v. Perlmutter*, the court upheld the Copyright Office's refusal to register copyright in an image allegedly "autonomously created by a computer algorithm running on a machine."[313] The Copyright Office had held that the image lacked human authorship, and the court agreed: computer programs, like animals, are not "authors" within the meaning of the Copyright Act.[314]

Instead, the author (and thus copyright owner) of a generation — if anyone — is some human connected to the generation. The four immediately relevant possibilities are (1) an author or authors whose works the model was trained on, (2) some entity in the generative-AI supply chain (e.g., the model trainer, model fine-tuner, or application developer), (3) the user who

309. *See generally* Comput. Assocs. Intern., Inc. v. Altai, 982 F.2d 693 (2d Cir. 1992) (standard case on software copyright); Pamela Samuelson, Randall Davis, Mitchell D. Kapor & Jerome H. Reichman, *A Manifesto Concerning the Legal Protection of Computer Programs*, 94 Colum. L. Rev. 2308 (1994) (lucid and time-honored analysis of software copyright).

310. Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 47 U. Pitt. L. Rev. 1185 (1985).

311. James Grimmelmann, *There's No Such Thing as a Computer-Authored Work – And It's a Good Thing, Too*, 39 Colum. J.L. & Arts 403 (2016).

312. Carys Craig & Ian Kerr, *The Death of the AI author*, 52 Ottawa L. Rev. 31 (2020).

313. Thaler v. Perlmutter, No. 22-1564 (D.D.C date).

314. *Id.*

prompted the application or API for the specific generation, or (4) no one. As between these four possibilities, there is no one-size-fits-all answer.

As framing for our analysis for these different possibilities, we first note that a generation is a compilation in the trivial sense in the same way that other works are all compilations. It also may seem intuitively attractive to consider generations to be analogous to collages. However, while this may seem like a useful metaphor, it can be misleading in several ways. For one, an artist may make a collage by taking several works and splicing them together to form another work. In this sense, a generation is not a collage: a generative-AI system does not take several works and splice them together. Instead, as we have described above, generative-AI systems are built with models trained on many data examples.[315] Moreover, those data examples are not explicitly referred back to during the generation process. Instead, the extent that a generation resembles specific data examples is dependent on the model encoding in its parameters what the specific data examples look like, and then effectively recreating them.[316] Ultimately, it is nevertheless possible for a generation to look like a collage of several different data examples;[317] however, it is debatable whether the the process that produced this appearance meets the definition for a collage. There is no author "select[ing], coordinat[ing], or arrang[ing]"[318] training examples to produce the resulting generation.

With this in mind, we assess the four relevant authorship possibilities for generations. We start with a generation that closely resembles a work in the training set. If the generation is actually identical to the training example — if it contains no original expression beyond what was present in the input work — then it is simply a copy of that underlying work and not a new copyrightable work at all,[319] Of course the copyright owner remains the original author, possibility (1). If the generation is, however, a derivative work of the underlying work that incorporates new authorship, a new copyright may subsist in it.[320] If the generation infringes, then it is uncopyrightable and the answer is (4): there is no separate copyright in the generation, even though it contains original authorship.[321] In such a case, the underlying copyright

315. *See supra* Part I.B; *supra* Part I.C.4.

316. *See infra* Part II.C.

317. *See infra* Part II.H.

318. 17 U.S.C. § 101 (definition of "compilation").

319. *See infra* Part II.C (concerning memorized training data and substantial similarity)

320. *See* 17 U.S.C. § 103(b) ("The copyright in such [a derivative] work is independent of . . . any copyright in the preexisting material.").

321. 17 U.S.C. § 103(a) ("[Copyright] protection for a [derivative] work . . . does not extend to any part of the work in which such material has been used unlawfully."). The courts have also held, illogically, that even if the underlying work was used with the copyright effectively also gives control over the generation; the user has in effect performed uncompensated creative labor for the benefit of the underlying copyright owner.[322]

Assuming, however, that the generation is sufficiently distinct from training data not to be "used unlawfully," a copyright owned by one of its creators may arise.[323] Some models and applications will produce original generations with minimal user input, which is possibility (2) above. The Draw Things iOS app, for example, suggests the prompt `"8k resolution, beautiful, cozy, inviting, bloomcore, decopunk, opulent, hobbit-house, luxurious, enchanted library in giverny flower garden, lily pond, detailed painting, romanticism, warm colors, digital illustration, polished, psychadelic, matte painting trending on artstation."` The user who taps "Generate" on the app user interface has contributed no authorship to the resulting image. This Person Does Not Exist is a website that creates a new (and uncannily realistic) deepfake photograph of a nonexistent person each time it is reloaded. The user who visits the site and clicks "reload" is not an author. If anyone can claim authorship credit here, it is the creators of these apps.

In other cases, the user will make substantial creative inputs through their choice of prompt. In addition to the authorship inhering in the prompt itself, two additional factors push towards making the user the copyright owner rather than the developer — i.e., possibility (3) from above. First, there is their causal responsibility for making the generation exist;[324] here, as in infringement, copyright law may care who "pushes the button."[325] Second, the providers of many generation applications have decided that as a practical matter they are uninterested in asserting copyright over the outputs. This is a business choice first and a copyright matter second, but widespread business practices often affect courts' decisions about how to allocate copyright ownership.[326]

But it is too hasty to say that the user is necessarily the owner of copyright in a generation, even once the training-data authors and model developers

owner's permission, it is uncopyrightable unless the owner also consents to a derivative copyright. *See, e.g.,* Gracen v. Bradford Exch., 698 F.2d 300 (7th Cir. 1983).

322. *See, e.g.,* Anderson v. Stallone, 11 U.S.P.Q.2d 1161 (C.D. Cal. 1989).

323. For derivative copyright purposes, lawful use includes fair use. *See, e.g.,* Keeling v. Hars, 809 F.3d 43 (2d Cir. 2015).

324. Balganesh, *supra* note 301.

325. Fox Broad. Co. v. Dish Network LLC, 160 F. Supp. 3d 1139, 1169 (C.D. Cal. 2015).

326. *E.g.,* Aalmuhammed v. Lee, 202 F.3d 1227, 1233 (9th Cir. 2000) (deferring to Hollywood practice of treating *auteur* directors as the "master mind[s]" behind films); Thomson v. Larson, 147 F.3d 195 (2d Cir. 1998) (deferring to theatrical crediting practices in holding that a dramaturg was not a co-author of a musical).

are out of the picture. It is also possible that *no one at all* owns a copyright in the generation (possibility (4)). The problem is that the generation may not be the product of sufficient human authorship. Consider the prompt.[327] "Scary lighthouse" is too short to contain sufficient originality to support a copyright;[328] short phrases are uncopyrightable.[329] If this phrase does not have the necessary modicum of creativity by itself, it seems unlikely that the additional choice to use it as a prompt is enough to put it over the threshold.[330] Another way of looking at the problem is that prompts like "Scary lighthouse" do not sufficiently constrain the output to make it the product of human authorship. As the Copyright Office put it when rejecting copyright in images created with Midjourney,

> Because of the significant distance between what a user may direct Midjourney to create and the visual material Midjourney actually produces, Midjourney users lack sufficient control over generated images to be treated as the "master mind" behind them. . . . [T]here is no guarantee that a particular prompt will generate any particular visual output. Instead, prompts function closer to suggestions than orders, similar to the situation of a client who hires an artist to create an image with general directions as to its contents.[331]

This is not the only possible view. A counter might be that for pragmatic reasons the copyright system will or should assign authorship to the user and overlook their minimal contributions.[332] While many current generative-AI systems have primarily text-based interfaces where short prompts might not

---

327. Mark Lemley argues that in fact the prompt is the relevant unit of originality and is in effect the work itself. Mark A. Lemley, How Generative AI Turns Copyright Law on its Head (2023) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4517702.

328. *Cf.* Magic Mktg. v. Mailing Servs. of Pittsburgh, 634 F.Supp. 769 (W.D. Pa. 1986) (holding the phrase "CONTENTS REQUIRE IMMEDIATE ATTENTION!" uncopyrightable).

329. 37 CFR § 202.1(a).

330. *See* Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, 34 Berkeley Tech. L.J. 343 (2019) (advancing this argument); *see also* Burk, *supra* note 301 (exploring variations).

331. Letter from Robert J. Kasunic to Van Lindburg, *Re: Zarya of the Dawn (Registration # VAu001480196)* 9–10 (Feb. 21, 2023), https://www.copyright.gov/docs/zarya-of-the-dawn.pdf.

332. *See, e.g.,* Grimmelmann, *supra* note 311, at 413–14 (discussing this possibility, and its difficulties). As one canonical case puts it, "Having hit upon such a variation unintentionally, the 'author' may adopt it as his and copyright it." Alfred Bell & Co. v. Catalda Fine Arts, 191 F.2d 99 court, 105 (1951).

amount to much creativity, future generative AI systems will likely have different interfaces that introduce other ways of controlling outputs.[333] But for now, it is the law that some generations are uncopyrightable despite containing material that would easily qualify for copyright if they had been produced manually by a human.[334]

This conclusion, however, is not categorical; "some" is not "all." Not every prompt is too short to be copyrightable, and not every user is a spectator to AI generation. Instead, some generations are the product of careful prompt engineering, in which users craft elaborate prompts to cause AI models to achieve specific aesthetic effects. These generations answer both of the objections above. These prompts are often long and intricate, running to dozens or hundreds of words, well above the short-phrase threshold. And these prompts are the result of an iterative creative process, in which the users have acquired a degree of mastery over the (putatively unpredictable) models they use, at least for specific types of outputs.[335] If an artist who flings a sponge against the wall in frustration is entitled to claim copyright in the resulting accidental spatter of paint, why not a user who deliberately crafts the perfect prompt?[336]

B. The Exclusive Rights

It is helpful to break down the *prima facie* case of infringement by the relevant exclusive right, rather than by the stage of the generative-AI supply chain. There are five relevant exclusive rights:

- The right to "reproduce the copyrighted work in copies" (the **reproduction** right).[337]
- The right to "prepare derivative works based upon the copyrighted work" (the **adaptation** right).[338]

333. For example, Ideogram has style tags that can be added to the prompt to modify the output (*Ideogram.AI*, IDEOGRAM.AI (2023), https://ideogram.ai/).

334. *See* James Grimmelmann, *Copyright for Literate Robots*, 101 IOWA L. REV. 657, 657 (2016) ("Almost by accident, copyright law has concluded that it is for humans only . . . ").

335. For a particularly disquieting example, see Emanuel Maiberg, *Inside the AI Porn Marketplace Where Everything and Everyone Is for Sale*, 404 MEDIA (Aug. 22, 2023), https://www.404media.co/inside-the-ai-porn-marketplace-where-everything-and-everyone-is-for-sale/.

336. *Alfred Bell*, 191 F.2d at 105 n.23.

337. 17 U.S.C. § 106(1).

338. 17 U.S.C. § 106(2).

- The right to "distribute copies . . . of the copyrighted work to the public" (the **distribution** right).[339]
- The right to "perform the copyrighted work publicly" (the **performance** right).[340]
- The right to "display the copyrighted work publicly" (the **display** right).[341]

To summarize briefly, every stage in the generative-AI supply chain requires a potentially-infringing reproduction and thus implicates copyright. We examine the other exclusive rights, which raise interesting edge cases.

The Reproduction Right

As relevant here, the reproduction right is triggered when a work is reproduced in "copies," which are defined as "material objects . . . in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device."[342] To be pedantic, a training dataset is not a "copy" because the dataset is not a "material object." Instead, the *computer* or *storage device* on which a dataset is stored is the copy.

The same is true for models and generations.[343] All of them trigger the reproduction right when they are created, because they are stored in material objects. Thus, the assembly of a dataset, the training of a model, the production of a generation, or a generative-AI system's use of a user-inputted prompt is a "reproduction" within the meaning of copyright law. All of these activities can infringe: the question is whether the resulting dataset, model, prompt, or generation is substantially similar[344] to the plaintiff's[345] copyrighted work.

One complication has to do with *how long* a work is fixed. Under the "RAM copy" doctrine, which dates to the 1990s, loading a copyrighted work

339. *Id.* § 106(3).

340. *Id.* § 106(4), (6).

341. *Id.* § 106(5).

342. 17 U.S.C. § 101 (definition of "copies").

343. The same could also be said for individual data examples within the dataset, which is one of the reasons we distinguish between expressive works and their datafied counterparts. *See supra* Part I.A.1; *supra* Part I.C.1; *supra* Part I.C.2.

344. *See infra* Part II.C.

345. Of course, there are different types of actors that can be responsible for each of these reproductions. For example, an application user could supply a reproduction of a copyrighted prompt (for which they do not hold the copyright), and the generative-AI system could in turn store that reproduction in memory. This could happen even for a generative-AI system that only trained its models on public domain data (i.e., did not violate the reproduction right with respect to training).

into a computer's working memory can infringe.[346] (Doing so is often necessary to run a program or to perform a computation on data.) On the other hand, more recent caselaw has held that transient copies do not count for the reproduction right.[347] The leading case, *Cartoon Network LP, LLLP v. CSC Holdings*, held that a buffer that was overwritten every 2.4 seconds was not an infringing reproduction of works that passed through the buffer.

The temporal threshold is not generally an issue for the outputs of stages in the generative-AI supply chain. Datasets, models, applications, prompts, and generations are all typically stored for far longer than the 2.4 seconds in *Cartoon Network*. Instead, the threshold may be more important for the inputs to the different stages. For example, a training example needs to be loaded into working memory to train a model on it. But the details of *how long* the example remains in memory, and *how much it is modified* while it is there, will depend on the training algorithm and architectural details of the environment (e.g., how fast the processors are). Similar considerations apply to the generation process — with similar uncertainties. Some generations run in a fraction of a second; others take minutes or hours.

There is also the problem of purely *internal* reproductions: ones that occur only in the middle of the training or generation process. These algorithms compute numerous new values, and often overwrite them repeatedly to conserve memory. Consider, for example, one of the middle stages of the archaeologist generation in Figure 3. One of these stages might resemble a copyrighted work more closely than the final output. Again, whether these fall underneath the *Cartoon Network* threshold depends on the details of the algorithm and environment.[348]

The Adaptation Right

While the reproduction right is about new copies of an existing work, the adaptation right is about new works based on an existing work. It is best understood as making clear that copyright in a work extends beyond literal similarity to incorporate changes of form, genre, and content such as translations, sequels, and film adaptations.[349] A training dataset is probably not a derivative work of any of the works in the dataset; it is more appropriately classified

346. MAI Sys. Corp. v. Peak Comput., 991 F.2d 511 (9th Cir. 1993).

347. Cartoon Network LP, LLLP v. CSC Holdings, 536 F.3d 121, 128–30 (2d Cir. 2008).

348. Alternatively, there is a strong fair-use case these transient internal copies. *See* Grimmelmann, *supra* note 334 (summarizing caselaw).

349. *See generally* Daniel Gervais, *The Derivative Right, or Why Copyright Law Protects Foxes Better than Hedgehogs*, 15 Vand. J. Ent. & Tech. L. 785 (2013); Pamela Samuelson, *The Quest for a Sound Conception of Copyright's Derivative Work Right*, 101 Geo. L.J. 1505 as a compilation "formed by the collection and assembling of preexisting materials."[350] A model is a good example of material that might or might not be an exact reproduction of the works it was trained on, but is more clearly a derivative work because it is "based on" its training data. Prompts might or might not be exact reproductions of existing works,[351] or they may be derivative works based on, for example, existing text or images. And generations are frequently derivative works of works in the training data, although whether and when a generation is a derivative of any particular work depends on similarity, discussed below.[352] Because the remedies for infringement of a work are the same, regardless of whether the defendant violated one exclusive right or several, it is an almost entirely scholastic exercise to try to identify the exact dividing lines at which the reproduction right leaves off and the adaptation right begins.[353]

More troublingly, it might be that the adaptation right can be infringed by derivative works that do not by themselves incorporate substantial expression from the plaintiff's work. In *Micro Star v. Formgen Inc.*, the defendant distributed fan-made levels for *Duke Nukem 3D*.[354] The level file format consisted entirely of geometry describing where the *Duke Nukem 3D* game engine should place walls and objects; the engine would then perform rendering using copyrighted art assets, but "[t]he MAP file . . . does not actually contain any of the copyrighted art itself; everything that appears on the screen actually comes from the art library."[355] Nonetheless, the court held that these files were infringing derivative works because "the stories told in the N/I MAP files are surely sequels, telling new (though somewhat repetitive) tales of Duke's fabulous adventures."[356]

(2013); Daniel Gervais, *AI Derivatives: The Application to the Derivative Work Right to Literary and Artistic Productions of AI Machines*, 52 Seton Hall L. Rev. 1111 (2022).

350. 17 U.S.C. § 101.

351. Anthropic, *supra* note 97.

352. *See infra* Part II.C.

353. The boundaries of the adaptation right are of greater importance in cases involving *unfixed* derivatives, where the reproduction right does not apply. *See* Lewis Galoob Toys, Inc. v. Nintendo of Am., Inc., 964 F. 2d 96, 967–69 (9th Cir. 1992) (erroneously holding that unfixed modifications of video games produced by altering bytes as they are read from a game cartridge are not derivative works). The boundaries also matter in cases involving the physical transfer of a copy from one substrate to another; here, there is a fixed copy, but there is no reproduction of it. *See, e.g.,* Lee v. ART Co., 125 F.3d 580 (7th Cir. 1997) (holding that the mounting of a page cut from a book on a ceramic tile does not create a derivative work).

354. Micro Star v. Formgen Inc., 154 F.3d 1107 (9th Cir. 1998).

355. *Id.* at 1110.

356. *Id.* at 1112.

A broad way to read *Micro Star* is to reason that models implicate the adaptation right when they "reference" the works they were trained on.[357] This test might be satisfied as long as any identifiable portion of a model was causally derived from a training example. However, reliable attribution of training examples in resulting generations remains an open research question.[358] A narrower reading would be that the model must also be capable of generating a substantially similar output — just as the audiovisual experience of playing a user-made *Duke Nukem 3D* level is substantially similar to the audiovisual experience of playing a canonical level created by 3D Realms.[359]

The Distribution Right

The distribution right applies when the defendant "distribute[s] copies . . . to the public by sale or other transfer of ownership."[360] Internet-era caselaw confirms that downloads and peer-to-peer transfers infringe the distribution right, so that the essence of the right is giving a stranger a copy, whether or not the copy previously existed.[361] Technically, the distribution right is not triggered by merely making a work available for download, but only when someone actually downloads it.[362] That said, in most interesting cases involving generative AI, making an artifact available is followed by an actual distribution.

When there is only a single entity involved in hosting a service, it is arguably not a distribution to assemble a dataset, train a model, program an application, input a prompt, or produce a generation. All of these activities involve only internal copying performed by the single hosting entity. They may result in reproductions and derivative works (as discussed above), but not distributions. The same is true when one party carries out multiple stages — for example, when a model trainer collects its own training data, or when a model owner creates test generations for its own use). Internal copying is not public distribution.

Instead, the distribution right is implicated when parties interact. In our model of the supply chain, there are at least five such kinds of interactions:

---

357. *Id.*

358. *see supra* note 122 and accompanying text (regarding the challenges of assigning "attribution" or "influence").

359. *See generally* MDY Indus., LLC v. Blizzard Ent., 629 F.3d 928 (9th Cir. 2010) (discussing "dynamic" aspects of copyrightable expression in video games).

360. 17 U.S.C. § 106(3).

361. Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1162–63 (9th Cir. 2007); London-Sire Recs., Inc. v. Doe 1, 542 F. Supp. 2d 153, 172 (D. Mass. 2008).

362. *London-Sire Recs.*, 542 F. Supp. 2d at 172.

- When a dataset creator or curator makes the dataset available to model trainers.[363]
- When a model trainer makes the model available for download (rather than for interactive use through a web interface or API).[364]
- When a service produces generations for users on demand.
- When the user of a service sends a potentially copyrighted prompt to the service host.[365]
- When a generation-time plugin retrieves content from an external source, which it then may use to produce a generation.[366]

In addition, when someone who has a dataset, model, prompt, or generation shares it, as is, with others, this is also a distribution. This last case is particularly relevant for open-source models, like those in the Llama family, which are often widely downloaded, shared, and re-uploaded.

The Display and Performance Rights

The display and performance rights characteristically involve human perception of a work. (The difference is that a display is static in time, while a performance is dynamic.) Models are not human-perceptible in any meaningful way, so it is hard to see how a model as such could infringe the display or performance rights. Similarly, while the individual works *within* a dataset can be perceptible, the dataset as a whole is not. Thus, for most practical purposes, only generations implicate these two rights.[367]

Like the distribution right, the display and performance rights are qualified by the word "public," so they apply only when the defendant makes the work perceptible *to others*. When a service produces a generation for a user, it will typically be a public display (for text and images) or a public performance (for audio and video). But in such a case, the generation will usually

363. This can happen in a variety of ways: e.g., open-sourcing a dataset, licensing a dataset, or some other contract between a dataset compiler/owner and a model trainer. For an example of the third case, consider how MosaicML is a platform for training and fine-tuning models for its clients.

364. *See supra* Part I.C.4.

365. *See supra* Part I.C.7.

366. *See supra* Part I.C.7.

367. Some services display user-supplied prompts as examples for other users, as suggestions for how to use the service. These are also public displays. A service, however, can easily protect itself from copyright liability for these prompts. It can require users to provide a license allowing their prompts in this way. As long as the number of such prompts displayed is small, the provider could potentially screen them manually for signs of infringement.

also be a reproduction and/or an adaptation, so the display and performance rights add relatively little. (In addition, if the user can download the generation, that will be a public distribution.)

One exceptional case when the display and performance rights may matter is for transient generations. Midjourney, for example, displays intermediate stages of the denoising process to users, as seen above in Figure 3. If one of those stages — but *not* the final result — infringes, then there might be a display without a reproduction or distribution.[368] Similarly, if an audio generation is played live for a user as it is created, but is not stored or made available for download, then this would be a performance without a reproduction or distribution.[369]

C. Substantial Similarity

Substantial similarity is a qualitative, factual, and frustrating question. Two works are substantially similar to "the ordinary observer, unless he set out to detect the disparities, would be disposed to overlook them, and regard their aesthetic appeal as the same."[370] A common test is a "holistic, subjective comparison of the works to determine whether they are substantially similar in total concept and feel."[371] This is not a standard that can be reduced to a simple formula that can easily be applied across different works and genres.[372]

In addition, except in clear cases, substantial similarity is typically a jury question.[373] Juries, unlike judges, are not required to provide reasoned elaboration justifying their verdicts. A typical case in which substantial similarity is genuinely contested, therefore, will provide little guidance for future cases. As a result, it is simply impossible to provide clear, accurate, and actionable predictions of substantial similarity in the mine-run of close cases.

---

368. *See* Cartoon Network LP, LLLP v. CSC Holdings, 536 F.3d 121 (2d Cir. 2008) (discussing transience exception to reproduction result).

369. *See* United States v. Am. Soc. of Composers, 627 F.3d 64 (2d Cir. 2010) (discussing reverse situation, a download without a performance).

370. Peter Pan Fabrics, Inc. v. Martin Weiner Corp., 274 F.2d 487, 489 (2d Cir. 1960) (Hand, J.).

371. Rentmeester v. Nike, Inc., 883 F.3d 1111, 1118 (9th Cir. 2018) (internal quotation omitted).

372. *But see* Scheffler, Sarah, Eran Tromer & Mayank Varia, *Formalizing Human Ingenuity: A Quantitative Framework for Copyright Law's Substantial Similarity*, *in* 2022 Proc. Symposium on Comput. Sci. & L. 37 (2022) (describing a principled computational basis for comparing works).

373. Tanksley v. Daniels, 902 F.3d 165, 171 (3d Cir. 2018).

Data

Substantial similarity of data poses no new issues distinctive to generative AI. Individual works included in training datasets can be compared to the plaintiff's work using the traditional substantial similarity test.

Training Datasets

Training datasets contain complete literal copies of millions of digitized copyrighted works. Complete literal copying is the paradigm case where substantial similarity is present as a matter of law.

Some datasets may represent works in specialized file formats, or may compress or transform them in ways that remove some of the information present in the work.[374] In these cases, the substantial similarity inquiry may involve returning these modified works to human-perceptible form (i.e., rendering them), followed by a traditional comparison. However, even when scaled down or partially noised,[375] as long as the original is recognizable, that will often be enough to support a finding of substantial similarity.[376]

Pre-Trained/Base Models

A model, as a collection of parameters, is different in kind from the copyrightable works it was trained on. Models are not themselves human-intelligible.[377] No viewer would say that the model has the same "total concept and feel" as a painting; no reader would say that it is substantially similar to a blog post; and so on.

That said, the Copyright Act does not require that copies be directly human-intelligible to infringe. A Blu-Ray is not directly intelligible by humans, either, but it counts as a "copy" of the movie on it. Indeed, all digital copies are unintelligible. Instead, they are objects "from which the work can be perceived, reproduced, or otherwise communicated . . . *with the aid of a machine or device*."[378] Thus, even if a model is uninterpretable, it might still be possible to "perceive[]" or "reproduce[]" a copyrighted work embedded in its parameters through suitable prompting. The resulting generation will render the work perceptible.

374. For an interesting attempt to quantify the information present in a work and what it means to remove some of it, see Scheffler, Tromer & Varia, *supra* note 372.

375. E.g., as in the case of diffusion. *See supra* Part I.B.3b

376. *See* Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir. 2007).

377. *See supra* Part I.A.2 (describing model parameters as vectors of numbers).

378. 17 U.S.C. § 101 (emphasis added).

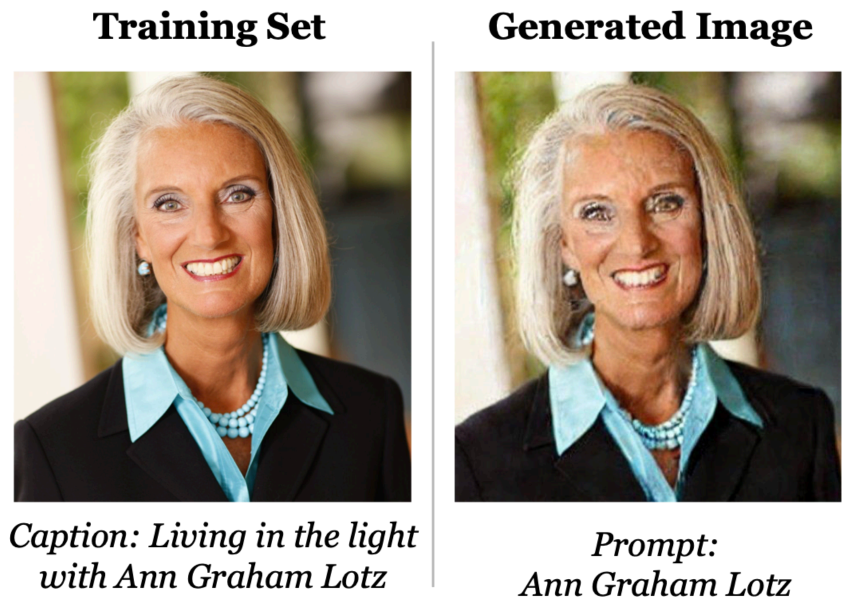

Figure 5: An example of a memorized image in Stable Diffusion, taken from Carlini et al., *Extracting Training Data from Diffusion Models* (2023).

> Wow. I sit down, fish the questions from my backpack, and go through them, inwardly cursing [MASK] for not providing me with a brief biography. I know nothing about this man I'm about to interview. He could be ninety or he could be thirty. → **Kate** (James, *Fifty Shades of Grey*).
>
> Some days later, when the land had been moistened by two or three heavy rains, [MASK] and his family went to the farm with baskets of seed-yams, their hoes and machetes, and the planting began. → **Okonkwo** (Achebe, *Things Fall Apart*).

Figure 6: Two examples of memorized text in GPT-4, taken from Chang et al., *Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4* (2023). In each case, when prompted with a sentence from a copyrighted book GPT-4 correctly fills in the name of a character.

Indeed, there is substantial evidence that many models have memorized copyrighted materials.[379] For example, Figure 5 shows how Stable Diffusion

379. Nicholas Carlini, Florian Tramèr & Eric Wallace et. al., *Extracting Training Data from Large Language Models*, *in* 2021 30TH USENIX SECURITY SYMPOSIUM (USENIX SECUhas memorized photographs. The memorized version is grainier and slightly shifted, but is immediately recognizable as the same photograph. Similarly, Figure 6 shows how GPT-4 must contain information from copyrighted books. GPT-4 can correctly fill in blanks in quotations from books; because the blanks consist of proper names of fictional characters, GPT-4 is not simply relying on its general knowledge of language.[380]

From a practical litigation perspective, a model might memorize more works or fewer.[381] But it seems clear that at least some models memorize at least some works sufficiently closely to pass the substantial-similarity test.

On this view, a sufficient condition[382] for a model to count as a substantially similar copy of a work is that the model is capable of generating that work as an output.[383] Note that this is direct infringement, not secondary.[384] The theory is not that the generation is an infringing copy, and that the model is a tool in causing that infringement in the way that a tape-duplicating ma-

---

RITY 21) 2633—2650 (2021) (GPT-2 memorizes training data); Nicholas Carlini, Jamie Hayes & Milad Nasr et al., Extracting Training Data from Diffusion Models (2023) (unpublished manuscript), https://arxiv.org/abs/2301.13188 (Stable Diffusion and Imagen memorize images); Kent K. Chang, Mackenzie Cramer, Sandeep Soni & David Bamman, Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4 (2023) (unpublished manuscript), https://arxiv.org/abs/2305.00118 (suggestive evidence that GPT-4 memorizes training data).

380. *See* Chang, Cramer, Soni & Bamman, *supra* note 379. The composition of GPT-4's training data is not public. If we don't know what the training data is, we technically cannot say that the training data was memorized with complete certainty. Filling-in-the-blank with proper names of fictional characters is highly suggestive of memorization — that copyrighted content is part of the training dataset — but does not literally satisfy the technical definition of memorization.

381. Nicholas Carlini, Daphne Ippolito & Matthew Jagielski et al., *Quantifying Memorization Across Neural Language Models*, *in* 2023 INT'L CONF. ON LEARNING REPRESENTATIONS (2023) (quantifying extent of memorization in language models); Carlini, Hayes & Nasr et al., *supra* note 379 (quantifying memorization in diffusion-based image models).

382. We write "sufficient" rather than "necessary and sufficient" because there might also be *other ways* of inspecting the model that are capable of recovering training data. Obviously, this possibility involves some speculation about technological developments, but it is worth emphasizing that, as computer scientists develop techniques that improve the interpretability of models, the copyright treatment of models and generations may well change as a result.

383. This is a sticky technical problem. Research has shown that memorization is not easily identifiable, and thus the amount of memorization in a model is not always or easily quantifiable. In particular, the choice of memorization identification technique and available information (e.g., knowledge of the training dataset, context window, etc.) affect the amount of memorization that can be identified. *See, e.g.,* Carlini, Ippolito & Jagielski et al., *supra* note 381.

384. *See infra* Part II.E (discussing direct and secondary infringement).

chine might be a tool in making infringing cassettes.[385] Rather, the theory is that the model itself is an infringing copy, regardless of whether that particular generation is ever made.[386]

Fine-Tuned Models and Aligned Models

The prior discussion about whether pre-trained models are substantially-similar copies mostly carries over to fine-tuned models and models trained with alignment – but there are a few additional considerations as well. As a starting point, fine-tuned and aligned models are influenced by the pre-trained model from which they were produced.[387] Fine-tuning may reduce the amount of memorized content from the pre-training dataset, but does not prevent all such memorization[388] and does not explicitly remove copies of training examples (i.e., particular text or images) from the trained model. Similarly, alignment may encourage models not to generate potentially infringing content, but that does not mean the copyrighted content was removed from the model.[389]

Further, the above considerations have to do with the pre-training data, not the data incorporated in these later stages in the generative-AI supply chain. Both fine-tuning and alignment bring in additional data sources — data that could also be memorized in the resulting model. As a result, just like pre-trained models, fine-tuned and aligned models could be infringing copies; but they can be infringing copies of the pre-training, fine-tuning, or alignment data.

385. *See* A & M Recs., Inc. v. Abdallah, 948 F. Supp. 1449 (C.D. Cal. 1996).

386. Alert readers will note the similarity to the debate over whether the mere act of making a work available without a download infringes the distribution right. *See* London-Sire Recs., Inc. v. Doe 1, 542 F. Supp. 2d 153 (D. Mass. 2008). *See generally* Peter S. Menell, *In Search of Copyright's Lost Ark: Interpreting the Right to Distribute in the Internet Age*, 59 J. Copyright Soc'y USA 1 (2011).

387. *See generally* Raffel, Shazeer, Roberts & Lee et al., *supra* note 59; Shayne Longpre, Gregory Yauney & Emily Reif et al., A Pretrainer's Guide to Training Data: Measuring the Effects of Data Age, Domain Coverage, Quality, & Toxicity (2023) (unpublished manuscript), https://arxiv.org/abs/2305.13169.

388. *See generally* Fatemehsadat Mireshghallah, Archit Uniyal & Tianhao Wang et. al., *An Empirical Analysis of Memorization in Fine-tuned Autoregressive Language Models*, *in* 2022 Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing 1816–1826 (2022).

389. While this is speculative, there is research indicating this may be the case. Prior work shows that models trained with alignment to be "safe" may be misaligned to produce "unsafe" content. Nicholas Carlini, Milad Nasr & Christopher A. Choquette-Choo et al., Are aligned neural networks adversarially aligned? (2023) (unpublished manuscript), https://arxiv.org/abs/2306.15447.