NYU|LAW

Journal of Intellectual Property & Entertainment Law

Statement of Purpose

Consistent with its unique development, the New York University Journal of Intellectual Property & Entertainment Law (JIPEL) is a nonpartisan periodical specializing in the analysis of timely and cutting-edge topics in the world of intellectual property and entertainment law. As NYU's first online-only journal, JIPEL also provides an opportunity for discourse through comments from all of its readers. There are no subscriptions or subscription fees; in keeping with the open-access and free discourse goals of the students responsible for JIPEL's existence, the content is available for free to anyone interested in intellectual property and entertainment law.

i

The New York University Journal of Intellectual Property & Entertainment Law is published up to three times per year at the New York University School of Law, 139 MacDougal Street, New York, New York, 10012. In keeping with the Journal's open access and free discourse goals, subscriptions are free of charge and can be accessed via www.jipel.law.nyu.edu. Inquiries may be made via telephone (212-998-6101) or e-mail (submissions.jipel@gmail.com).

The Journal invites authors to submit pieces for publication consideration. Footnotes and citations should follow the rules set forth in the latest edition of The Bluebook: A Uniform System of Citation. All pieces submitted become the property of the Journal. We review submissions through Scholastica (scholasticahq.com) and through e-mail (submissions.jipel@gmail.com).

The Journal is also available on WESTLAW, LEXIS-NEXIS and HeinOnline.

From Amchem to AI: Overcoming Certification in Generative AI Copyright Class Actions

Cindy L. Chang*

Rapid advances in generative artificial intelligence have fueled a wave of copyright class action litigation by authors, artists, and other rightsholders who allege that AI developers have unlawfully copied protected works at scale to build training datasets. But the public policy debate over fair use and innovation has largely overshadowed the threshold procedural question: whether generative AI copyright suits can be certified as damages class actions under Federal Rule of Civil Procedure 23(b)(3)—and if so, under what theory.

This Note argues that, under modern Supreme Court doctrine on predominance and standing, the vast majority of putative classes seeking actual damages are structurally uncertifiable. Given the black-box nature of LLM models and datasets, individualized inquiries about whether any given protected work was used, how it affected the LLM, and whether concrete harm can be identified fail to meet the predominance floor as enunciated in Amchem v. Windsor *and* Wal-Mart v. Dukes. *Related issues of class ascertainability and overinclusion only serve to sharpen these defects. Statutory damages under the Copyright Act, 17 U.S.C. § 504, may appear to offer a workaround— but in practice, the formal pre-registration requirement offers an inequitable foundation for aggregate adjudication because many putative classes are comprised of high-volume or resource-constrained creators.*

In response to these concerns, this Note urges federal district courts to treat class certification in this emerging area as a problem of procedural design: district judges

should more readily employ Rule 23(c)(4)(A) issue-class certification to isolate and resolve core, genuinely common liability questions early, and—where technically dense discovery and LLM model opacity threaten to distort the certification inquiry—appoint Rule 53 special masters to supply neutral expertise and manageability. Together, these targeted tools can reduce wasteful e-discovery, deter meritless claims at the outset, and clarify the path toward just resolutions for plaintiffs and defendants while Congress works toward durable substantive rules for generative AI.

INTRODUCTION

There is no hotter market than the artificial intelligence market. In recent years, the rapid development of generative artificial intelligence ("Gen-AI") has culminated in a cultural and economic frenzy not unlike the early days of the internet.[1] As with most forms of machine learning, Gen-AI models are "trained" on

---

[1] Last year, Forbes announced that the Gen-AI sector could add $4.4 *trillion* annually to the global economy. David Jones, *Harnessing Generative AI: A $4.4 Trillion Opportunity for the Global Economy,*

massive datasets of written and visual works, which they analyze to learn patterns of "expressive information."[2] These learned patterns enable the Gen-AI models to generate new content in response to a user's inputted prompt.[3] There's just a slight wrinkle—these training datasets are made up of millions, if not billions, of data files and other digital objects, each of which may be eligible for copyright protection, and many of which certainly are protected.[4] This has triggered a tidal wave of high-profile copyright infringement class action litigation.[5] The putative class members in these suits are copyright holders—authors, artists, musicians, photographers, licensors—who contend that their intellectual property rights have been violated, without permission or compensation, by the AI developers who scrape vast quantities of content from the internet to train Gen-AI systems.[6]

The merits of these suits are hotly debated. Proponents of Gen-AI argue that the "ingestion" of unlicensed copyrighted works for purposes of training Gen-AI systems constitutes fair use.[7] If trained properly, Gen-AI neural networks

---

FORBES (Aug. 7, 2024, 8:15 AM), https://www.forbes.com/councils/forbestechcouncil/2024/08/07/harnessing-generative-ai-a-44-trillion-opportunity-for-the-global-economy/ [https://perma.cc/K4CP-P5FA].

[2] Tremblay v. OpenAI, Inc., 716 F. Supp. 3d 772, 776 (N.D. Cal. 2024).

[3] *See* Authors Guild v. OpenAI, Inc., 345 F.R.D. 585, 589 (S.D.N.Y. 2024).

[4] *See* Alberto Romero, *A Complete Overview of GPT-3—The Largest Neural Network Ever Created*, TOWARDS DATA SCI. (May 24, 2021), https://towardsdatascience.com/gpt-3-a-complete-overview-190232eb25fd [https://perma.cc/J24V-7WVQ] (noting that "GPT-3 was trained with almost all available data from the Internet"). These training datasets are so vast that there's a good chance any given person has already been captured in one. *See, e.g.*, HAVE I BEEN TRAINED?, https://haveibeentrained.com [https://perma.cc/ZEB8-6CMM] (last visited Oct. 18, 2025).

[5] *E.g.*, *Authors Guild*, 345 F.R.D. at 585; Andersen v. Stability AI Ltd., 744 F. Supp. 3d 956 (N.D. Cal. 2024); *see also* Ariba A. Ahmad & Andrew M. Gross, *Generative AI Systems Tee Up Fair Use Fight*, 357 NAT. L. REV. (2024) (overview of ongoing AI copyright disputes). Courts regularly consolidate these suits against large AI developers and tech companies. *See* Kevin Madigan, *Mid-Year Review: AI Lawsuit Developments in 2024*, COPYRIGHT ALL. (July 25, 2024), https://copyrightalliance.org/ai-lawsuit-developments-2024/ [https://perma.cc/8BKZ-9E3W].

[6] *See Authors Guild*, 345 F.R.D. at 588 (noting that a writers association, authors, and a news organization seek to enforce their intellectual property against OpenAI). In addition to these direct "input side" infringements, many plaintiffs allege indirect "output side" infringement via the derivative creations of Gen-AI models. *See* Ahmad & Gross, *supra* note 5. In this Note, we will primarily refer to input-side claims as they have seen the most movement in court thus far.

[7] Copyright Act of 1976 § 101, 17 U.S.C. § 107 (2023); *see also* James Vincent, *The Scary Truth About AI Copyright is Nobody Knows What Will Happen Next*, THE VERGE (Nov. 15, 2022, 10:00 AM), https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data [https://perma.cc/44GH-UBBF] (speculating that the "vast majority of the output of generative AI models cannot be copyright protected"); Keith Kupferschmid, *Insights from Court Orders in AI Copyright Infringement Cases*,

"statistically 'learn[]' how certain … styles appear without storing exact copies of the [works] it has seen."[8] Under this theory, because they are generating novel outputs instead of creating duplicates, Gen-AI models sufficiently transform the materials on which they were trained. In response, critics emphasize that as a matter of U.S. copyright law, reproducing entire copyrighted works and storing them for more than a transitory period is *prima facie* copyright infringement.[9] There are also instances of "overfitting" or "overtraining," in which an overrepresented input in the Gen-AI training dataset causes the system to spit out a suspiciously duplicative or derivative representation of the original work.[10] And from a humanistic perspective, why should capitalism-driven innovation prevail over artists' natural rights?[11]

Substantive debate has thus far overshadowed the procedural gatekeepers that may render everything else moot. In all the noise around fair use doctrine and policy implications, there is a glaringly unanswered question: under Federal Rule of Civil Procedure 23 ("Rule 23"), can and should these AI-driven copyright class actions be certified in the first instance? To date, no motions for certification have been decided in any of the ongoing cases.[12] In other words, there is zero precedent on whether a district court should grant or deny certification to the putative classes in these contentious suits.

This Note argues that actual damages-based classes in AI copyright actions cannot be defensibly certified under existing doctrinal principles. Rule 23 in its current formulation is ill-suited to address the novel intersection of law and

COPYRIGHT ALL. (Dec. 12, 2024), https://copyrightalliance.org/ai-copyright-infringement-cases-insights/ [https://perma.cc/SXQ2-ZX2C].

[8] Benj Edwards, *Artists File Class-Action Lawsuit Against AI Image Generator Companies*, ARS TECHNICA (Jan. 16, 2023, 6:36 PM), https://arstechnica.com/information-technology/2023/01/artists-file-class-action-lawsuit-against-ai-image-generator-companies/ [https://perma.cc/K6TP-MWAN].

[9] Kupferschmid, *supra* note 7.

[10] Edwards, *supra* note 8 (pointing to Leonardo da Vinci's *Mona Lisa* as an example of an overtrained image).

[11] *See* Daryl Lim, *Innovation and Artists' Rights in the Age of Generative AI*, GEO. J. INT'L AFF. ONLINE (July 10, 2024), https://gjia.georgetown.edu/2024/07/10/innovation-and-artists-rights-in-the-age-of-generative-ai/ [https://perma.cc/HFQ4-NG9M] (comparing America's "market-driven" approach with the EU's "rights-focused" approach).

[12] Authors Guild v. Open AI, Inc., 345 F.R.D. 585, 589 (S.D.N.Y. 2024) ("[N]o motion seeking certification of a class has been filed to date."). The most procedurally advanced of the AI copyright class actions have recently begun pre-certification discovery. *See infra* Part II.

technology inherent in AI-based copyright claims.[13] The past decade-plus of Supreme Court jurisprudence on predominance and standing has been skeptical—if not outright hostile—toward any expansion of damages class certification doctrine under Rule 23(b)(3).[14] Moreover, Congress has yet to legislate on the limitations and consequences of Gen-AI as a matter of statutory copyright law.[15] In light of these obstacles to certification, this Note urges district courts to consider alternative procedural mechanisms in pursuit of efficient and just outcomes for all parties.

Part I provides a brief primer on the technical foundations of Gen-AI, as well as the Rule 23 class action framework as it relates to AI copyright litigation.

Part II explores some of the interwoven challenges faced by putative classes in currently-pending actions—chiefly around predominance, ascertainability, and standing—where uncommon questions and unmanageable criteria stack against certification in the vast majority of cases. I further argue that the requirements for a statutory cause of action for damages under the Copyright Act, 17 U.S.C. § 504, fail to address these challenges in an equitable and accessible manner.

Part III advocates for the adaptation of two procedural devices to help district courts navigate AI copyright class actions. First, given the inevitability of individualized inquiries in these types of disputes, courts should invoke Rule 23(c)(4)(A) issue subclass certification *sua sponte* to isolate and resolve foundational questions of liability at the outset. Resolving a pivotal common issue—e.g., whether a specific Gen-AI model was trained exclusively on public domain works or a mixed dataset of protected works—can distinguish meritless claims and facilitate (or extinguish) the creation of a defined putative class, reducing costs of e-discovery down the line. Second, special masters should be appointed in AI class actions to provide technical guidance for certification-related

---

[13] *See* Jasminka Kalajdzic, *AI & The End of Lawyers … Defeating Class Certification*, JOTWELL (Mar. 24, 2021), https://courtslaw.jotwell.com/ai-the-end-of-lawyers-defeating-class-certificat ion/ [https://perma.cc/HP7Y-RQXZ] (noting the difficulties presented for copyright class actions by Rule 23(b)(3)'s requirement that common issues predominate over individual ones).

[14] *See, e.g.*, Comcast Corp. v. Behrend, 569 U.S. 27 (2013) (holding that plaintiffs were improperly certified under Rule 23(b)(3) since they did not establish that damages could be measured classwide); *see also* Part II, *infra*.

[15] *See* Zeynep Ü. Kahveci, *Attribution Problem of Generative AI: A View from U.S. Copyright Law*, 18 J. INTELL. PROP. L. PRAC. 796, 797-98 (2023).

inquiries and dense e-discovery disputes.[16] Cutting edge Gen-AI systems are often so complex that not even their developers know precisely how they work.[17] Judges cannot—and should not—be expected to comprehend the inner workings of these constantly evolving deep learning models. Of course, it would be inefficient to appoint special masters in *every* ongoing AI class action. But if a case makes it to pre-certification discovery, district courts should take advantage of their broad discretion under Rule 53 to appoint technical experts capable of examining threshold questions of liability, injury, and causation through a more AI-centric lens. Together, these measures serve as practical and administrable stopgaps to protect just outcomes in AI copyright class actions.

I
GEN-AI TRAINING METHODOLOGIES AND THE RULE 23 FRAMEWORK

In a relatively short amount of time, Gen-AI technologies, particularly large language models (LLMs) like the infamous ChatGPT, have transformed the landscape of content creation and information synthesis across diverse sectors, from videogames to biomedical research.[18] Even the legal field is not immune from Gen-AI's impact—much to the chagrin of a growing number of judges receiving briefs with hallucinated caselaw.[19] These Gen-AI models, driven by sophisticated neural networks and training methodologies, have introduced unprecedented complexities in the realm of copyright class actions.

Much of this tension stems from the opacity and scale of Gen-AI training methodologies. The numbers speak for themselves. OpenAI's now-obsolete GPT-3 was trained on a dataset of millions of text-based works comprising 499 billion

---

[16] *See* FED. R. CIV. P. 53 (provisions for the appointment and use of masters).

[17] Ahmad & Gross, *supra* note 5.

[18] Ian Walker, *Microsoft Touts Generative AI Model for Recreating Video Game Visuals and Controller Inputs*, POLYGON (Feb. 19, 2025), https://www.polygon.com/news/525625/microsoft-muse-xbox-gene rative-ai-model-video-game-gameplay-bleeding-edge [https://perma.cc/SR5Y-8JR5]; *Healthcare Startup Abridge Raises $250 Million to Enhance AI Capabilities*, REUTERS (Feb. 17, 2025, 8:54 AM), https://ww w.reuters.com/business/healthcare-pharmaceuticals/healthcare-startup-abridge-raises-250-million-enhance-ai-capabilities-2025-02-17 [https://perma.cc/337V-73RJ].

[19] *See* Mata v. Avianca, Inc., 678 F. Supp. 3d 443, 466 (S.D.N.Y. 2023) (in which Judge Castel imposed sanctions on law firm for submitting brief with six fake, AI-generated case citations); Park v. Kim, 91 F.4th 610, 614-16 (2d Cir. 2024) (referring Plaintiff-Appellant's attorney to Court's Grievance Panel for citing to AI-generated authority in brief).

tokens, or forty-five terabytes of data.[20] Its (now equally obsolete) successor GPT-4 was trained on *one thousand* terabytes of data.[21] On the visual side, an early Stable Diffusion model was trained on 2.3 billion captioned images.[22] Because these Gen-AI training datasets are comprised of such vast amounts of material scraped from every corner of the internet, it becomes an insurmountable challenge to ascertain the extent of potential copyright infringement.

To appreciate the scope of these complicated legal questions, it is essential to understand the mechanics behind both Gen-AI and the federal class action device. Part I-A provides the reader with a broad overview of the technical foundations underlying Gen-AI models and training methodologies. Part I-B then summarizes the Rule 23 procedural framework that governs the class action mechanism as it relates to AI-centric copyright class action litigation.

A.   *Primer on LLMs, Diffusion Models, and Gen-AI Training Methodologies*

Generative AI is best understood as a subset of machine learning, focused on *generating* new content rather than simply analyzing and classifying data, through neural network-based architectures. Neural networks are a class of machine learning algorithms "loosely modeled on the way brains work," consisting of layers of interconnected nodes (or "neurons") that process data and learn patterns by adjusting their connections based on inputs and outputs.[23] Two of the most common—and, as this Note posits, legally problematic—neural network architectures are LLMs and diffusion models.[24]

---

[20] Matthew Sag, *Copyright Safety for Generative AI*, 61 Hous. L. Rev. 295, 315 (2023).

[21] *Id.*

[22] *Id.* (citing Andy Baio, *Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator*, Waxy (Aug. 30, 2022), https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/ [https://perma.cc/7MUH-KR7U]).

[23] Peter Jeffcock, *Neural Networks in Deep Learning*, Oracle Big Data Blog (June 1, 2020), https://blogs.oracle.com/bigdata/post/neural-networks-in-deep-learning [https://perma.cc/ZPP3-KNHZ].

[24] There are other emerging subsets of Gen-AI such as Generative Adversarial Networks (GANs), which generate highly realistic images and videos—including controversial "deepfakes"—through a dual neural network mechanism. *See* Priyanshu Prasad, *Artificial Faces: The Encoder-Decoder and GAN Guide to Deepfakes*, Medium (Apr. 2, 2024), https://medium.com/@priyanshuprasad1718/artificial-faces-the-encoder-decoder-and-gan-guide-to-deepfakes-75a1eed0e265 [https://perma.cc/H6ZT-35FJ]. For the purposes of providing a high-level overview, we will focus on LLMs and diffusion models, the two most widely used (and litigated) forms of Gen-AI.

LLMs are deep learning models trained on vast quantities of text in order to understand, process, and generate human-like language through probabilistic methods.[25] For example, if an LLM is trained on a dataset in which "*res ipsa*" is followed by "*loquitur*" 100% of the time, the model will predict and generate "*res ipsa loquitur*" with near certainty when prompted with a user input of "*res ipsa*." This training is done through a combination of pre-training (the "P" in GPT) and transformer architecture (the "T" in GPT).[26]

During pre-training, an LLM learns the building blocks of language—such as grammar, facts, and problem-solving skills—by studying billions of texts, many of which may be protected by copyright.[27] Once this linguistic foundation is established, transformer architecture works its magic by enabling the LLM to synthesize human language through the groundbreaking "self-attention" mechanism.[28] Simply put, this self-attention mechanism allows the LLM to master the subtleties of semantics and syntax by weighing the importance of each word relative to others in a sentence. For instance, in the standalone sentence "*The lawyer argued that he is innocent*," self-attention helps the LLM recognize that "*he*" refers back to "*the lawyer*." However, if we were to add a prior sentence with context like "*The lawyer is representing the defendant at trial*," then the LLM's self-attention would instead teach the model to link "*the defendant*" from the first sentence to "*he*" in the following sentence, ensuring coherent syntactical understanding and generation across discrete passages of text.

Diffusion models, on the other hand, are a distinct class of generative models trained to output images, audio, and other "continuous" data formats.[29] Those with a background in physics may recognize diffusion as "a process where something—

---

[25] OpenAI's ChatGPT is an example of a popular LLM implicated in a number of copyright class actions. *See* Partha P. Ray, *ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope*, 3 INTERNET THINGS & CYBER-PHYS. SYS. 121, 133-34 (2023).

[26] Joe Regalia, *From Briefs to Bytes: How Generative AI is Transforming Legal Writing and Practice*, 59 TULSA L. REV. 193, 201-02 (2024).

[27] *Id.* at 201.

[28] Transformer neural network architecture was pioneered in a 2017 landmark research paper by a team of Google Brain machine learning researchers. Today, this architecture continues to serve as the backbone for many traditional LLMs. *See* Ashish Vaswani et al., *Attention is All You Need*, 30 ADVS. NEURAL INFO. PROCESSING SYS. 5998 (2017).

[29] Midjourney is an example of a popular diffusion model. For additional examples, see *Introduction to diffusion models for machine learning*, SUPERANNOTATE (Feb. 28, 2025), https://www.superannotate.com/blog/diffusion-models [https://perma.cc/W6K9-QU8Z].

atoms, molecules, energy, pixels—move from a region of higher concentration to another one of lower concentration."[30] However, in the machine learning context, the object is to instead train the Gen-AI model through a *reverse* diffusion method that restores and regenerates the structure of deliberately destroyed audiovisual data.[31]

This is done through a two-step training process. In the "forward diffusion" phase, an audiovisual sample's data distribution structure is corrupted by adding Gaussian noise on a schedule until it becomes nearly pure noise.[32] In the "reverse diffusion" phase, the model learns how to gradually remove the noise step by-step until it recovers the original data structure.[33] By repeating this process billions of times on billions of samples, the diffusion model progressively learns to construct data with desired properties through the probabilistic denoising of "systematically random noise."[34] When it comes time to generate an output, the trained diffusion model utilizes its learned patterns to iteratively refine randomly sampled noise into coherent media outputs—such as the voice of Johnny Cash singing 'Barbie Girl,' or a picture of a furry feline judge holding court.[35]

Whether diffusion model or LLM, the Gen-AI training process is invariably premised on the model ingesting massive datasets of written, visual, and/or audio works in order to generate and recognize patterns. These patterns are then encoded into model weights which inform future content generation through probabilistic

---

[30] J. Rogel-Salazar, *Diffusion Models—More Than Adding Noise*, Domino.AI (Oct. 4, 2022), https://domino.ai/blog/diffusion-models-more-than-adding-noise [https://perma.cc/8K9Y-FC55] (describing the destruction of an image structure through noise in forward diffusion).

[31] *Id.* ("Unlike the physical process, the aim is to learn a reverse diffusion … process.").

[32] *See* Kemal Erdem, *Step by Step Visual Introduction to Diffusion Models*, Medium (Nov. 9, 2023), https://medium.com/@kemalpiro/step-by-step-visual-introduction-to-diffusion-models-235942d2f15c [https://perma.cc/8K9Y-FC55] (describing the destruction of an image structure through noise in forward diffusion).

[33] *Id.*

[34] Rogel-Salazar, *supra* note 30 (in-depth explanation of Markov chains in the reverse diffusion process); *accord* Ryan O'Connor, *Introduction to Diffusion Models for Machine Learning*, AssemblyAI (May 12, 2022), https://www.assemblyai.com/blog/diffusion-models-for-machine-learning-introduction/ [https://perma.cc/NBJ8-82NP].

[35] *Johnny Cash – Barbie Girl (Cover by There I Ruined It) Restoration*, YouTube (Sept. 21, 2023), https://www.youtube.com/watch?v=MAFdzBTe2lg [https://perma.cc/9BG3-N9Q2]; *Cat Judge in Courtroom* (illustration), https://www.freepik.com/premium-ai-image/cat-judge-courtroom-humorous-animal-illustration_306123647.htm [https://perma.cc/6P7Y-V7ZU].

associations.[36] Ideally, this multi-step process ensures that a Gen-AI model's output will not be duplicative or derivative of any given work, but rather an amalgamation of information synthesized from countless data sources. In reality, however, the sheer enormity and diversity of data within these training datasets ensures that when given certain prompts, Gen-AI models will output word-for-word passages or near-identical copies of copyrighted works.[37]

Research studies have documented the existence of these problematic edge cases.[38] One example is the phenomenon of verbatim memorization. As the name implies, verbatim memorization occurs when an LLM unintentionally overtrains on a specific data sample, causing it to "memorize" and generate "verbatim" sequences of text from the sample.[39] Traditional transformer-based LLMs are particularly susceptible to verbatim memorization for two main reasons. First, large training datasets often contain duplicates or near-duplicates of high visibility works found across many different sources.[40] For instance, take Martin Luther King Jr.'s iconic "*I Have a Dream*" speech.[41] The speech's fame and distinctiveness both increase the risk of overtraining because it exists on many websites, caches, and other scraped corners of the internet. If or when this risk is realized, an LLM will

---

[36] *See generally* Dan Milmo, *'Impossible' to Create AI Tools Like ChatGPT Without Copyrighted Material, OpenAI Says*, THE GUARDIAN (Jan. 8, 2024, 8:40 AM), https://www.theguardian.com/techno logy/2024/jan/08/ai-tools-chatgpt-copyrighted-material-openai [https://perma.cc/B27U-48NX] (noting inevitable presence of copyrighted material); O'Connor, *supra* note 34.

[37] *See* Antonia Karamolegkou et al., *Copyright Violations and Large Language Models*, *in* PROCEEDINGS OF THE 2023 CONF. ON EMPIRICAL METHODS IN NAT. LANGUAGE PROCESSING (Houda Bouamor et al. eds., 2023) (noting that one can ask a language model to print exact lines from a text).

[38] *See, e.g.*, Kent K. Chang et al., *Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4*, *in* PROCEEDINGS OF THE 2023 CONF. ON EMPIRICAL METHODS IN NAT. LANGUAGE PROCESSING (Houda Bouamor et al. eds., 2023) (addressing disparities and biases in which texts are most heavily memorized by language models); Karamolegkou et al., *supra* note 37.

[39] Karamolegkou et al., *supra* note 37; *see also* Edwards, *supra* note 8 (discussing the overtraining problem).

[40] *See generally* Katherine Lee et al., *Deduplicating Training Data Makes Language Models Better*, *in* PROC. OF THE 60TH ANN. MEETING OF THE ASS. FOR COMPUTATIONAL LINGUISTICS 8424 (2022) (finding that for LLMs trained on duplicate-heavy datasets, over 1% of unprompted output is copied verbatim from training data).

[41] The "I Have a Dream" speech is copyright-protected through 2058 and remains strictly enforced by the King Estate. *See* Arlen W. Langvardt, *"I Have a [Fair Use] Dream": Historical Copyrighted Works and the Recognition of Meaningful Rights for the Public*, 25 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 939, 942-46 (2015).

memorize and generate copyright-protected segments of the speech verbatim when fed a prompt about famous speeches or important civil rights moments.

Second, probabilistic generation training methods mean that LLMs are optimized to premise future outputs based on the patterns within preceding ones.[42] This can be a problem when a data sample is particularly *low* visibility. Imagine a situation where instead of a world-famous speech, we have a copyrighted string of code so obscure and hyper-specific that it can only be found in five out of the billions of sources within a dataset. Because an LLM trained on this dataset has only encountered the string of code in a unique sequence, its probabilistic pattern-recognition training will struggle to find ways to 'generalize' the code into new sequences. This makes the LLM much more likely to reproduce the copyrighted code verbatim.

To be sure, the risks of verbatim memorization are not confined to text-based LLMs. Diffusion models face similar vulnerabilities because they too learn through probabilistic associations—meaning that the more prominent *or* unique a work is within the dataset, the higher the likelihood of exact replication.[43] The visual nature of image generation also poses its own challenges as Gen-AI capabilities continues to evolve at an alarming pace. Not too long ago, machine learning scientists were proposing novel theories of "cascaded diffusion models" for high fidelity image generation.[44] Just three years later, advancements in Gen-AI such as deepfake technology are already approaching near-photorealistic fidelity, exacerbating the potential for infringement of copyrighted works.[45] One can only imagine—and dread—the advancements that will be made in Gen-AI over the next three years and beyond, and the consequences of those advancements on copyright law.

B.   Class Action Procedure Under Rule 23

Class actions in federal court are governed by Federal Rule of Civil Procedure 23.[46] Under Rule 23, a putative class must first satisfy the four gatekeeping

---

[42] *See generally* Part I-A, *infra* (overview of probabilistic LLM training process).

[43] *See generally* Lee et al., *supra* note 40 (addressing disparities in which works are overly memorized by language models).

[44] Jonathan Ho et al., *Cascaded Diffusion Models for High Fidelity Image Generation*, 23 J. Mach. Learning Rsch. 1 (2022).

[45] *See, e.g.*, Prasad, *supra* note 24 (discussing advancements in deepfake technology).

[46] Fed. R. Civ. P. 23.

prerequisites of Rule 23(a): numerosity, commonality, typicality, and adequacy of representation.[47] Next, the claims must fall within at least one of the three enumerated Rule 23(b) buckets. Rule 23(b)(1) applies where separate actions by or against individual class members would create a risk of conflicting obligations or undue prejudice, such as the prototypical "limited fund" and bankruptcy scenarios.[48] Rule 23(b)(2)—designed to combat desegregation obstructionism during the civil rights movement—applies in situations where a class seeks "generally applicable" injunctive or declaratory relief.[49]

The least structured of the three categories, Rule 23(b)(3) is a "most adventuresome innovation" that may be invoked in a variety of contexts, from public interest advocacy[50] to general damages claims.[51] Despite its penchant for creativity, there are two important limiting principles under Rule 23(b)(3): "the questions of law or fact common to class members [must] *predominate* over any questions affecting only individual members," and the class action must be "*superior* to other available methods" of adjudication.[52] The requirements of predominance and superiority aim to balance the "competing tugs of individual autonomy … and systemic efficiency."[53] Today, these two prerequisites play the biggest role in gatekeeping a variety of complex 23(b)(3) claims—including claims for actual damages in AI copyright class actions.[54]

---

[47] Specifically, the class must be so numerous that joinder of all members is impracticable; there must be questions of law and fact common to the class; the claims or defenses of the class representatives must be typical of the class; and the class representatives must fairly and adequately protect the interests of the class. *See* FED. R. CIV. P. 23(a).

[48] Amchem Prod., Inc. v. Windsor, 521 U.S. 591, 614 (1997).

[49] I had the pleasure of serving as Burt Neuborne's Civil Procedure TA in Fall 2024, where I learned about Rule 23(b)(2)'s history as a "civil rights" class action, as well as his personal involvement in such groundbreaking litigation. *See generally* Burt Neuborne, *The Gravitational Pull of Race on the Warren Court*, 2010 SUP. CT. R. 59 (2011); *accord* Suzette M. Malveaux, *The Modern Class Action Rule: Its Civil Rights Roots and Relevance Today*, 66 U. KAN. L. REV. 325 (2017).

[50] *E.g.*, Danielle Tarantolo, Guest Speaker at N.Y.U. School of Law, Class Actions Seminar (Nov. 18, 2024) (discussing her work on public service class actions at the New York Legal Assistance Group (NYLAG)).

[51] *See Amchem*, 521 U.S. at 614.

[52] FED. R. CIV. P. 23(b)(3) (emphases added).

[53] *Amchem*, 521 U.S. at 615. The non-exhaustive statutory considerations for predominance and superiority include the "extent and nature" of ongoing litigation, as well as the "desirability" and "difficulties" of managing claims under a class action device. *See* FED. R. CIV. P. 23(b)(3)(A)-(D).

[54] We will discuss this assertion in Part II, *infra*.

An action's formal scope, claims, and defenses are defined during the process of class certification, which must occur at an "early practicable" stage of the case per Rule 23(c).[55] Class certification serves as a check to ensure that only meritorious actions proceed, thereby safeguarding rights of both plaintiffs and defendants. Certification is proper when a putative class affirmatively demonstrates its compliance with the four prerequisites of Rule 23(a) and at least one Rule 23(b) subsection.[56] Any claims brought under Rule 23(b)(3) must additionally satisfy statutory notice requirements to protect absent class members.[57] Importantly, class certification is not a one-time checkbox. Once a class is certified, the court retains discretion under Rule 23(c)(1)(C) to alter or amend its certification decision at any time "before final judgment," ensuring the continued propriety of certification as new facts come to light.[58]

To date, the majority of input-based AI copyright class actions are jointly brought under Rules 23(b)(2) and 23(b)(3).[59] This means that the putative classes seek both injunctive relief to protect their intellectual property from continued or future exploitation by Gen-AI systems, as well as monetary damages for alleged violations that have already occurred. While allowed under the letter of Rule 23, these types of joint class actions are generally held to a higher level of judicial scrutiny. The Supreme Court confirmed in *Wal-Mart* that claims for individualized monetary relief, such as backpay or damages, cannot be brought under Rule 23(b)(2) due to the lack of procedural safeguards around notice and opt-out.[60] As such, hybrid putative classes seeking equitable relief under Rule 23(b)(2) and damages under Rule 23(b)(3) must demonstrate that the requirements of each subsection are independently satisfied in order to attain certification. For the

---

[55] FED. R. CIV. P. 23(c)(1)(A).

[56] *Amchem*, 521 U.S. at 592 (holding that both subdivisions (a) and (b) must be met for a class to be certified). Courts may also certify a class for settlement purposes only, but Rule 23 prerequisites must still be met. *Id.* at 620-21; *see also* FED. R. CIV. P. 23(e).

[57] *See* FED. R. CIV. P. 23(c)(2)(B).

[58] FED. R. CIV. P. 23(c)(1)(C).

[59] *See, e.g.*, *Andersen*, 2024 WL 3823234, at *2 (describing the six different classes of claims being brought under Rules 23(b)(2) and (b)(3)).

[60] Wal-Mart Stores, Inc. v. Dukes, 564 U.S. 338, 361-63 (2011). *See also* FED. R. CIV. P. 23(b)(2) advisory committee's note to 1966 amendment (stating that 23(b)(2) "does not extend to cases in which the appropriate final relief relates exclusively or predominantly to money damages").

purposes of this Note, we will focus on the subset of Rule 23(b)(3) damages claims that raise uniquely difficult questions at the certification stage.

II
CLASSWIDE CHALLENGES IN AI COPYRIGHT CLASS ACTION LITIGATION

The Rule 23(b)(3) framework has traditionally been an awkward fit for copyright class action litigation.[61] Because the relevant questions of copying, substantial similarity, and damages are highly individualized in most copyright infringement disputes, red flags often emerge as to the scope of classwide questions.[62] Still, plaintiffs in traditional non-AI class actions can almost always allege at least one concrete occurrence of infringement, *e.g.*, the distribution or reproduction of an identifiable creator's discrete creation.[63] In comparison, AI-based copyright class actions suffer an extra layer of complexity due to the opaque nature of any given Gen-AI model's training and development process, which is often proprietary and undisclosed.[64]

This lack of transparency makes it difficult to determine whether and which copyrighted works were scraped by any given training dataset—especially since we now know that not every inputted work will affect a Gen-AI model's ultimate output.[65] The complaints in AI copyright class actions generally fail to cite any given instance of infringement under the Copyright Act's traditional indicia of misuse, weakening plaintiffs' claims at the outset.[66] Even when the *potential* class

---

[61] *See* Xiyin Tang, *The Class Action as Licensing and Reform Device*, 122 COLUM. L. REV. 1627, 1645 (2022) ("[F]ew copyright class actions were filed between 1938, when Rule 23 was promulgated, and 1990.").

[62] Courts have even referred to proposed copyright classes as "Frankenstein monster[s]" when evincing skepticism toward their sufficiency under Rule 23. Football Ass'n Premier League, Ltd. v. YouTube, Inc., 297 F.R.D. 64 (S.D.N.Y. 2013). *See also generally* U.S. COURTS FOR THE NINTH CIRCUIT, MANUAL OF MODEL CIVIL JURY INSTRUCTIONS 17.17 (2024) (outlining the plaintiff's highly individualized burden of proof in copyright disputes).

[63] *E.g.*, Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015), *cert. denied*, 578 U.S. 941 (2016) [hereinafter *Google Books Case*] (alleging infringement over the defendant's scanning and indexing of copyright-protected books).

[64] *See* Edwards, *supra* note 8 (noting OpenAI's absence from one class action complaint because the company had "not publicly disclosed the exact contents of its training dataset").

[65] Zheng Dai & David K. Gifford, *Training Data Attribution for Diffusion Models*, ARXIV (June 3, 2023, 6:36 PM), https://arxiv.org/abs/2306.02174?utm [https://perma.cc/W7N5-K9ZZ] (research study demonstrating that certain inputs may have a more substantial impact on an AI diffusion model's outputs, while others contribute minimally or not at all). *See* Part I-A, *supra*.

[66] *See, e.g.*, *Andersen*, 2024 WL 3823234 (N.D. Cal. 2024), at **2-3 (finding that the majority of plaintiffs' initial claims were dismissed for unsupported and generalized allegations of infringement without "plausible

of injured copyright holders is enormous, the inherent complications posed by the nature of AI technology frustrate certification of damages classes under current Rule 23(b)(3) jurisprudence.

A. *Common Questions Do Not Predominate Under* Amchem *and* Wal-Mart

Under the foundational threshold of Rule 23(a)(2), class members must share at least one common question of law or fact that is "capable of classwide resolution—which means that determination of its truth or falsity will resolve an issue that is central to the validity of each one of the [class members'] claims in one stroke."[67] This requires a baseline demonstration "that the class members have suffered the same injury."[68] Moreover, after the landmark ruling in *Wal-Mart*, courts have effectively subsumed the "low hurdle" of Rule 23(a)(2)'s commonality requirement under the more stringent Rule 23(b)(3) requirement of predominance.[69] While Rule 23(a)(2) nominally requires the existence of a single common question of law or fact, courts assessing Rule 23(b)(3) claims apply heightened scrutiny by requiring classwide questions capable of "generat[ing] common *answers* to drive the resolution of the litigation"[70] through the acquisition of "generalized, classwide proof."[71] This raises the bar at the certification stage, particularly in complex disputes where varying levels of harm and divergent legal claims erode the cohesiveness of the class as a whole.

Amchem provides an example of Rule 23(b)(3) predominance acting as a bar on damages class certification. In *Amchem*, putative class members were exposed to the defendant's asbestos products "in different ways, over different periods, and for different amounts of time."[72] These factual differences meant that their injuries

---

facts in support"); *id.* at **5-6 (tentatively accepting plaintiffs' theory of induced infringement after rejecting their theories of direct infringement).

[67] *Wal-Mart*, 564 U.S. at 350.

[68] *Id.* at 349-50 (internal quotations omitted).

[69] *See* In re Kind LLC Litig., 337 F.R.D. 581, 594 (S.D.N.Y. Mar. 24, 2021), *class decertified in* 627 F. Supp. 3d 269 (S.D.N.Y. Sep. 9, 2022) (quoting *Amchem*, 521 U.S. at 591); *cf. Wal-Mart*, 564 U.S. at 368-69 (Ginsburg, J., concurring in part and dissenting in part) (disagreeing with the majority decision to "import[] into the Rule 23(a) determination concerns properly addressed in a Rule 23(b)(3) assessment").

[70] *Wal-Mart*, 564 U.S. at 350 (quoting Richard A. Nagareda, *Class Certification in the Age of Aggregate Proof*, 84 N.Y.U. L. REV. 97, 132 (2009)).

[71] Tyson Foods, Inc. v. Bouaphakeo, 577 U.S. 442, 453 (2016); *accord* Moore v. PaineWebber, Inc., 306 F.3d 1247, 1252 (2d Cir. 2002).

[72] *Amchem*, 521 U.S. at 609.

were dissimilar in scope, severity, and concreteness. As such, even if there had been concrete evidence that the defendant proximately caused *one* class member's cancer, such proof could not extend to the other members' injuries.[73] This combination of factual differences and individualized inquiries precluded a finding of predominant common questions.[74] Foreshadowing *Wal-Mart*, the *Amchem* Court declared that the mere identification of broadly "common" questions, such as class members' shared experience of asbestos exposure or their common interest in receiving prompt and fair compensation, was insufficient to satisfy the predominance standard under Rule 23(b)(3).[75]

We can extend the *Amchem* doctrinal analysis to damages classes in AI copyright class actions. Each putative class member contends that their specific copyright-protected work was used in the training of Gen-AI models. Like the disparate exposure periods in *Amchem*, there is no single instance of copyright infringement identified as common to the class.[76] In fact, some class members' protected works might have not been sampled at all.[77] Consequentially, the plaintiffs are unable to affirmatively establish how, when, and to what extent they were actually injured on a classwide basis. The broad spectrum of alleged harms in AI copyright suits—ranging from market impact to lost licensing opportunities to broader reputational damage—mirrors the rejected health claims in *Amchem* and discrimination claims in *Wal-Mart*.[78] And akin to the rejected plaintiffs in *Wal-Mart*, our current AI copyright plaintiffs have thus far been unable to establish any formal policy of infringement by the AI developers. Resolving crucial questions

---

[73] As a solution to causation and apportionment issues in mass torts class actions, some modern scholars propose a theory of percentage-based proportional liability aligned to a risk-based framework. Such an approach may have bridged the classwide predominance and causation gaps in *Amchem*. However, courts have yet to routinely adopt this type of theory. *See generally* Mark A. Geistfeld, *The Principle of Misalignment: Duty, Damages, and the Nature of Tort Liability*, 121 Yale L.J. 142 (2011).

[74] *See Amchem*, 521 U.S. at 609-10.

[75] *Id.* at 594; *accord Wal-Mart*, 564 U.S. at 349-50 (asserting that mere "reciting" of common questions is insufficient).

[76] *See* Kupferschmid, *supra* note 7 ("AI models … operate differently and the copyrighted works at issue … are different.").

[77] Similarly, some class members in *Amchem* "suffered no physical injury" and raised exposure-only claims. *Amchem*, 521 U.S. at 609-10.

[78] *See id.*; *Wal-Mart*, 564 U.S. at 349-50 (arguing that because there are so many potential ways to violate Title VII, the "mere claim" of a Title VII injury by multiple employees "gives no cause to believe all their claims can productively be litigated at once").

of injury-in-fact and causation for an entire class of AI copyright claimants would require innumerable (and impermissible) mini-trials.[79]

In the past, savvy litigants overcame the predominance hurdle by using statistical sampling and auditing techniques to demonstrate widespread patterns of actual harm without requiring individualized proof for every class member.[80] Essentially, a random representative sample of the putative class would be chosen, their individual questions resolved, and their individual damages determined; the resulting probability of meritorious claims was then extrapolated to calculate classwide liability and damages.[81] This method of statistical aggregation facilitated certification of large and diverse classes in complex doctrinal areas where "individual stakes are high and disparities among class members great"—*e.g.*, mass torts, employment discrimination, consumer fraud.[82]

But the *Wal-Mart* Court pejoratively rejected this "trial by formula" approach on the grounds that due process and the Rules Enabling Act preclude any interpretation of Rule 23 that deprives a defendant of their right to litigate statutory defenses to individual claims.[83] The holding in *Wal-Mart* sounded the death knell for many hard-to-certify areas of litigation, which may soon include AI copyright class actions. AI-related claims inherently implicate diffuse harms that are challenging to adjudicate on an individualized basis due to the sheer volume of data inputted and outputted by Gen-AI models.[84] In a pre-*Wal-Mart* regime, statistical sampling might have effectively and efficiently resolved issues of predominance in these AI disputes by making it possible to approximate concrete infringement patterns for the putative class as a whole.

To be sure, a minority of courts still allow the use of statistics-based "representative evidence" to establish liability for actual damages in rare contexts, chiefly antitrust and wage-and-hour disputes.[85] In *Bouaphakeo*, the Court even

---

[79] *See* Peter N. Salib, *Artificially Intelligent Class Actions*, 100 Tᴇx. L. Rᴇv. 519, 521 (2022).

[80] *See* Alexandra D. Lahav, *The Case for "Trial by Formula,"* 90 Tᴇx. L. Rᴇv. 571 (2012).

[81] Salib, *supra* note 79, at 522.

[82] *Amchem*, 521 U.S. at 625; *see also* Hilao v. Estate of Marcos, 393 F.3d 987 (9th Cir. 2004) (allowing a statistical sample of 137 claims to calculate and determine classwide damages).

[83] *Wal-Mart*, 564 U.S. at 367 (citing 28 U.S.C. § 2072(b)).

[84] *See generally* Kupferschmid, *supra* note 7 (noting adjudicatory challenges in AI class actions).

[85] *E.g.*, *Bouaphakeo*, 577 U.S. at 454-55 (upholding the use of statistical sampling in an FLSA wage-and-hour class action but declining to "establish general rules governing the use of statistical evidence . . . in all

backpedaled on its previously rigid stance by asserting that "*Wal-Mart* does not stand for the broad proposition that a representative sample is an impermissible means of establishing classwide liability."[86]

These narrow post-*Wal-Mart* applications of statistical sampling are distinguished, however, by their broadly-applicable regulatory frameworks (e.g., FLSA, Sherman Act) that establish consistent "elements of the underlying cause[s] of action."[87] Wage-and-hour actions revolve around standardized employment policies, such that each similarly situated employee could, in theory, rely on a *common* representative sample to bring an individual action for damages.[88] Antitrust cases often analyze market harm that flows from the defendant's anticompetitive conduct in a manner documented to affect all impacted class members similarly, making it easier to sustain inquiries of causation and actual damages.[89] In both veins of litigation, the threshold question of liability does not necessarily depend on the individual class members' scope of injuries so much as proof of the defendant's statutory violation in the first instance.[90] Meanwhile, in the toxic torts and copyright infringement contexts, threshold liability for actual damages can hinge entirely on individual questions of harm and causation.[91] The perceived dominance of individual questions in AI copyright cases is likely to disqualify such cases from the narrow *Bouaphakeo* construction of permissible representative evidence.[92]

---

class-action cases"). *But see Comcast*, 569 U.S. at 27 (antitrust class decertified for lack of predominance after representative evidence failed to establish damages capable of classwide measurement).

[86] *Bouaphakeo*, 577 U.S. at 457.

[87] Erica P. John Fund, Inc. v. Halliburton Co., 563 U.S. 804, 809 (2011).

[88] *Compare Bouaphakeo*, 577 U.S. at 455-59, *with Wal-Mart*, 564 U.S. at 355-56 (finding no common policy of discrimination).

[89] *Cf.* Vivek Mani & Darwin V. Neher, *Antitrust Impact in Class/Collective Actions*, CORNERSTONE RESEARCH (Nov. 13, 2020), https://www.cornerstone.com/insights/articles/antitrust-impact-in-class-collective-actions/ [https://perma.cc/CVW4-STBR%5d (discussing complications of the "common impact question" in antitrust litigation). *See generally* Clayton Act, 15 U.S.C. § 15.

[90] Of course, each individual class member must still demonstrate a concrete injury in fact to satisfy Article III standing. *See* TransUnion LLC v. Ramirez, 594 U.S. 413 (2021). However, the initial liability threshold can be much more forgiving in disputes with clear-cut, uniformly applicable statutory violations and damages.

[91] *See, e.g.*, *Amchem*, 521 U.S. at 609-11.

[92] *See Bouaphakeo*, 577 U.S. at 457-59 (distinguishing the proposed methodology in *Wal-Mart* from the instant wage-and-hour case).

In limiting the usage of statistical sampling techniques, the *Wal-Mart* Court ensured that many high-merit 23(b)(3) class actions, brought by diverse plaintiffs suffering widespread but unequal harm at the hands of powerful defendant corporations, are effectively dead in the water for lack of predominance. As Justice Ginsburg intimated in her partial dissent, this creates a procedural barrier to substantive justice by unduly focusing "on what distinguishes individual class members, rather than on what unites them."[93] One might find it disappointing for a Court that values "deeply rooted … history and tradition"[94] to so plainly forgo Rule 23(b)(3)'s historical basis in the "vindication of the rights of groups of people who individually would be without effective strength to bring their opponents into court at all."[95]

B.   Lack of Ascertainability Bars Classwide Determination of "Injury in Fact" *Under* TransUnion *Standing Doctrine*

Beyond the statutory text of Rule 23, many jurisdictions recognize an additional "implicit threshold requirement" at the certification stage.[96] Ascertainability—the requirement that "a class [be] defined using objective criteria that establish a membership with definite boundaries" at the outset—poses a problem for plaintiffs in AI copyright class actions due to the opaque nature of Gen-AI training methodologies and the inherent nuances of copyright ownership.[97]

Circuit courts are split on the exact scope of ascertainability doctrine. Some, like the Third Circuit, have adopted a rigorously heightened requirement that requires both objective criteria and an "administratively feasible" method for identifying class members.[98] Others, including the Second and Ninth

---

[93] *Wal-Mart*, 564 U.S. at 377 (Ginsburg, J., concurring in part and dissenting in part).

[94] Palko v. Connecticut, 302 U.S. 319, 325-26 (1937); *accord* Dobbs v. Jackson Women's Health Org., 597 U.S. 215-16 (2022); Timbs v. Indiana, 586 U.S. 146 (2019); McDonald v. Chicago, 561 U.S. 742, 767 (2010).

[95] *Amchem Prods., Inc*, 521 U.S. at 617 (internal quotation marks and citation omitted) (summarizing the 1966 Advisory Committee's original intent in creating the Rule 23(b)(3) cause of action).

[96] In re Petrobras Sec., 862 F.3d 250, 264 (2d Cir. 2017) (quoting Sandusky Wellness Ctr., LLC v. Medtox Sci., Inc., 821 F.3d 992, 995 (8th Cir. 2016)) (internal quotation marks and citation omitted).

[97] *Id* at 257. While this section focuses on ascertainability at the certification stage, the issues here also implicate the Rule 23(b)(3) requirements of superiority and manageability.

[98] *See, e.g.*, Carrera v. Bayer Corp., 727 F.3d 300, 308 (3d Cir. 2013) (holding that a proposed method of ascertaining class members must be reliable, administratively feasible, and must permit the defendant to challenge the evidence used to prove class membership); Marcus v. BMW of N. Am., LLC, 687 F.3d 583,

Circuits, require classes to be defined using objective criteria without extensive individualized factfinding but reject the formal obligation of administrative feasibility;[99] they instead consider manageability as an aspect of the broader 23(b)(3) predominance and superiority analysis.[100]

This circuit split informs litigation strategy. To date, almost if not all ongoing AI copyright class actions have been filed in Southern District of New York and Northern District of California, where judges not only reject the heightened test of administrative feasibility but are highly proficient at managing large dockets of technology-related complex litigation.[101] In these plaintiff-friendly jurisdictions, the "standard for ascertainability is not demanding and is designed only to prevent the certification of a class whose membership is truly indeterminate."[102] But even under the Second and Ninth Circuits' relatively permissive approach, putative classes in AI copyright cases for actual damages will likely struggle to define a body of determinate membership.[103]

The second amended complaint in *Andersen* illustrates the difficulty of fully ascertaining putative classes in AI class actions. Plaintiffs assert six different classes and subclasses, including a broad "Damages Class" under Rule 23(b)(3) that includes "[a]ll persons or entities nationalized or domiciled in the United States that own a copyright interest in *any* work that was used to train *any* version

---

593 (3d Cir. 2012) ("If class members are impossible to identify without … 'mini-trials,' then a class action is inappropriate.").

[99] *See, e.g.*, *In re Petrobras*, 862 F.3d at 264 ("We conclude that a freestanding administrative feasibility requirement is neither compelled by precedent nor consistent with Rule 23 ….."); Briseno v. ConAgra Foods, Inc., 844 F.3d 1121, 1123 (9th Cir. 2017) (holding that class representatives need not demonstrate "administratively feasible means of identifying absent class members" as a condition of class certification) (internal citations omitted).

[100] *In re Petrobras*, 862 F.3d at 268 ("Whereas ascertainability is an absolute standard, manageability is a component of the superiority analysis, which is explicitly comparative in nature.").

[101] *See, e.g.*, Bartz v. Anthropic PBC, No. 3:24-cv-05417 (N.D. Cal. filed Aug. 19, 2024); Doe 1 v. GitHub, Inc., No. 4:22-cv-06823 (N.D. Cal. filed Nov. 3, 2022); Authors Guild v. OpenAI, Inc., No. 1:23-cv-08292 (S.D.N.Y. filed Sept. 19, 2023); Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal. filed Jan. 13, 2023).

[102] Ebin v. Kangadis Food Inc., 297 F.R.D. 561, 567 (S.D.N.Y. 2014) (internal quotations omitted).

[103] Note that in an action for *statutory* damages, the class is more likely to be ascertainable because members have to be formally registered with the Copyright Office. This Note's discussion of ascertainability presumes an action for *actual* damages. See Part II-C *infra* for an examination of statutory damages in the AI copyright context.

of an AI image product[].”[104] This seems objective enough at first glance, but problems of vagueness and manageability surface upon review. The term "any work" covers an overwhelming range of copyrighted materials. To determine whether a particular work was "used to train" Gen-AI systems, one would have to comb through metadata and compare millions of outputs against individual copyright registrations.[105] Not only would this warrant lengthy discovery, but such discovery would be cost-prohibitive to conduct into the training datasets of multiple Gen-AI models.[106]

Further, the complex nature of copyright ownership—particularly in cases of joint authorship and licensing relationships—renders identification of all rightful owners improbable without individualized inquiries at the outset.[107] These identification efforts are further burdened by the fact that many works on the internet lack embedded metadata linking them to specific copyright holders or lack thereof.[108] Tracing ownership of such "orphan works" for the purposes of notice would require resource-intensive investigation in contravention of judicial economy.[109] Even if the court interprets ascertainability in a liberal manner without

---

[104] Second Amended Complaint at 8, Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal. Oct. 31, 2024) (emphasis added).

[105] In the *Andersen* complaint, Plaintiffs themselves concede that current Gen-AI dataset search and comparison tools are incomplete and unreliable for these purposes. *See id.* at 17.

[106] Discovery in *Andersen* is currently underway and will likely consume enormous amounts of time and money. One of the named defendants, Stability AI, has released at least eight new versions of its main Stable Diffusion model since August 2022. Stability AI, https://stability.ai [https://perma.cc/R53F-R4XM] (last visited Oct. 18, 2025). Earlier versions of this model were trained on subsets of the LAION-5B database, made up of "5 billion image-text pairs" that were "scraped from the web." Training Data, *Stable Diffusion*, https://stablediffusion.gitbook.io/overview/stable-diffusion-overview/technology/training-data [https://perma.cc/YAS7-TCP7]. Later versions have incorporated additional datasets; for example, the "StableLM-Tuned-Alpha" model was fine-tuned on a combination of five datasets. *StableLM-Tuned-Alpha*, https://huggingface.co/stabilityai/stablelm-tuned-alpha-3b [https://perma.cc/F7ML-8ZKM] (scroll down to Training Dataset). All in all, there may be billions of copyrighted works to sift through at discovery.

[107] *See generally* Lydia P. Loren, *The Nature of Copyright: A Law of Users' Rights*, 90 Mich. L. Rev. 1615, 1625 (1992) (noting how copyright "ownership is much more complicated than merely the 'right to copy'").

[108] *See* Jane C. Ginsburg, *Losing Credit: Legal Responses to Social Media Platforms' Stripping of Copyright Management Metadata from Photographs*, The Media Institute (May 30, 2016), https://www.mediainstitute.org/2016/05/30/losing-credit-legal-reponses-to-social-media-platforms-stripping-of-copyright-management-metadata-from-photographs/ [https://perma.cc/RLM5-CFVR] (analyzing legal implications of the removal of author-identifying and other copyright-relevant metadata).

[109] Orphan works are copyrighted materials whose owners are either unknown or cannot be contacted. *See* U.S. Copyright Office, *Orphan Works and Mass Digitization* at 1 (June 2015) https://www.copyright.gov/orphan/reports/orphan-works2015.pdf [https://perma.cc/3HMT-7MXD].

regard for lack of objective criteria,[110] certification is arguably inappropriate given the serious Rule 23(b)(3) concerns around unmanageability and inadequate notice.[111]

Post *TransUnion*, it is also critical to consider how undefined class boundaries may indirectly impact the justiciability of AI copyright class actions.[112] Although Rule 23 and Article III standing are distinct procedural doctrines, they bear overlapping practical implications for the feasibility of class certification. It is well-established that to show Article III standing in federal court, plaintiffs "bear the burden of … establishing, *inter alia*, that they have suffered an injury in fact."[113] This of course applies to the plaintiffs in putative class actions; the mere violation of a federal statute, on its own, is insufficient to get through the door.[114] In the landmark *TransUnion* decision, the Supreme Court asserted that "every class member must have Article III standing in order to recover individual damages"[115] because "Article III does not give federal courts the power to order relief to any uninjured plaintiff, class action or not."[116] This break from precedent not only signals a shaky future for statutes like the Copyright Act—which provides avenues for statutory damages without proof of actual harm—but also risks stifling the majority of AI-related copyright class actions at the outset.

It remains unclear exactly how *TransUnion* will impact AI class action standing, but its effects have already been felt in the general AI litigation context. *Raw Story Media v. OpenAI* is a recent case in which the district court granted AI defendants' motion to dismiss on a *TransUnion* rationale.[117] In *Raw Story Media*, the plaintiffs sought, *inter alia*, actual and statutory damages for OpenAI's alleged

---

[110] *See* In re MF Global Holdings Ltd. Inv. Litig., 310 F.R.D. 230, 235 (S.D.N.Y. 2015) (noting that the Second Circuit instructs district courts "to adopt a liberal interpretation of Rule 23" in favor of class certification).

[111] *See* FED. R. CIV. P. 23(b)(3)(D). *But see In re Petrobras*, 862 F.3d at 268 (noting that courts should try to avoid denying certification "on the sole ground that it would be unmanageable").

[112] *TransUnion*, 594 U.S. at 413. *See generally* Erwin Chemerinsky, *What's Standing After TransUnion LLC v. Ramirez*, 96 N.Y.U. L. REV. 269 (2021).

[113] Lujan v. Defenders of Wildlife, 504 U.S. 555, 555 (1992).

[114] *See* Aaron Moss, *Will Copyright Claims Keep Standing After New Ruling?*, COPYRIGHT LATELY (Nov. 10, 2024), https://copyrightlately.com/raw-story-copyright-lawsuit-standing/ [https://perma.cc/X3AU-72CJ].

[115] *TransUnion*, 594 U.S. at 431.

[116] *Id.* (quoting *Bouaphakeo*, 577 U.S. at 466 (Roberts, C.J., concurring)).

[117] Raw Story Media, Inc. v. OpenAI, Inc., 756 F.Supp.3d 1 (S.D.N.Y. Nov. 7, 2024).

use of their copyrighted articles in its Gen-AI training datasets. To be clear, the plaintiffs did not assert a standard claim of copyright infringement; they instead brought their claims under §1202(b) of the Digital Millennium Copyright Act (DMCA), under an injury of law theory.[118] The court summarily dismissed all claims for lack of standing. In her dismissal order, Judge McMahon in the Southern District of New York emphasized that the plaintiffs had failed to show an "injury in fact that is concrete, particularized, and actual or imminent."[119] Specifically, given the nature of Gen-AI training databases, "Plaintiffs [were unable to allege] any *actual* adverse effects stemming from this alleged DMCA violation."[120] They therefore failed to meet the injury in fact threshold requirement for standing that "*TransUnion* [made] clear."[121]

Although *Raw Story Media* is not a perfect analogue, plaintiffs in AI copyright class actions should remain cautious in light of its ruling. In the AI context, it is hard enough to plausibly allege injury in fact for a specific and ascertainable group of plaintiffs as in *Raw Story Media*. It's an entirely different challenge to do so for a broadly defined 23(b)(3) class where some members may not have suffered any injury at all. Take the proposed "Damages Class" from *Andersen*.[122] Under its nebulous definition, some members will surely fail to show injury in *law*, much less injury in fact. Perhaps their works were not scraped by Gen-AI training datasets at all; or the use was insufficient to constitute statutory infringement; or there was simply no evidence that legal injury of infringement caused actual harm. Under a strict interpretation of *TransUnion*, each and every class member must affirmatively overcome all of these threshold presumptions. Otherwise, like the *TransUnion* class members whose false credit reports were not disseminated, these putative AI copyright plaintiffs will lack the necessary "concrete harm" for standing in a post-*TransUnion* Rule 23(b)(3) damages claim.[123]

---

[118] Plaintiffs did, however, "contend that their injury bears a 'close relationship' to the tort of copyright infringement." *See id.* at 6.

[119] *Id.* at 4 (quoting *TransUnion*, 594 U.S. at 423).

[120] *Raw Story Media*, 756 F.Supp.3d at 6.

[121] *Id.* at 6 (comparing the plaintiffs' argument to that of the dissent in *TransUnion*). *See TransUnion*, 594 U.S. at 427 ([U]nder Article III, an injury in law is not an injury in fact."). *But see id.* at 453-454 (Thomas, J., dissenting) (criticizing the majority opinion that "legal injury is *inherently* insufficient to support standing").

[122] Second Amended Complaint at 8, Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal. Oct. 31, 2024).

[123] *See id.* at 433-38. These copyright plaintiffs might advance in the alternative that there is a material risk of future harm, as the disqualified plaintiffs argued in *TransUnion*. However, this type of forward-looking

In early 2025, the Supreme Court granted *certiorari* in *Laboratory Corp. v. Davis* to answer the narrow yet pivotal question of whether federal courts may certify a damages class pursuant to Rule 23(b)(3) when the class includes both injured and uninjured class members.[124] This issue is subject to a longstanding circuit split: the Ninth Circuit permits certification even where a *de minimis* number of class members lack standing, while the Second Circuit bars certification outright if any uninjured members are within the proposed class.[125] However, in a rare post-oral argument move, the writ of *certiorari* in *Laboratory Corp.* was dismissed as improvidently granted five months later.[126] Justice Kavanaugh succinctly stated in his dissent to the dismissal that "the Court simply declines to decide … the important class-action question on which we granted certiorari."[127] On his part, Justice Kavanaugh endorsed the Second Circuit's approach by which "a federal court may not certify a damages class that includes both injured and uninured members."[128] He reasoned that "Rule 23 requires that common questions predominate in damages class actions[,]" and "when a damages class includes both injured and uninjured members, common questions do not predominate."[129] But ultimately, the lack of a definitive answer in *Laboratory Corp.* heralds an uncertain future for the conception of ascertainability and standing in AI copyright class actions.

As the law stands today, overinclusive classes with uninjured members generally contravene both Rule 23 ascertainability and Article III standing

---

harm is typically only accepted in Rule 23(b)(2) actions for injunctive relief. *Id.* at 435-36 ("[A] plaintiff's standing to seek injunctive relief does not necessarily mean that the plaintiff has standing to seek retrospective damages."); *accord Wal-Mart*, 564 U.S. at 361-63 (holding that individualized damages claims cannot be brought under Rule 23(b)(2)).

[124] Lab'y Corp. of Am. v. Davis, No. 24-304, 2025 WL 288305 (Jan. 24, 2025).

[125] *Compare* Olean Wholesale Grocery Coop., Inc. v. Bumble Bee Foods LLC, 31 F.4th 651 (9th Cir. 2022) (holding that inclusion of a *de minimis* amount of uninjured class members does not defeat certification), *with* Denney v. Deutsche Bank AG, 443 F.3d 253, 264 (2d Cir. 2006) (affirming certification in instant case, but noting that "no class may be certified that contains members lacking Article III standing"). *See also* Amit Rana & Antonia I. Stabile, *Supreme Court Grants Certiorari on Important Class Certification Standards*, Venable LLP (Feb. 12, 2025), https://www.venable.com/insights/publications/2025/02/supreme-court-grants-certiorari-on-important [https://perma.cc/BYW9-KZJ3].

[126] Lab'y Corp. of Am. v. Davis, 605 U.S. 327, 328 (2025) (per curiam) (slip opinion).

[127] *Id.* at 330 (Kavanaugh, J., dissenting).

[128] *Id.* at 328 (Kavanaugh, J., dissenting).

[129] *Id.*; *see also id.* at 332 (collecting cases) (citing *Wal-Mart* and *Amchem* as "precedents [that] make this a straightforward case").

principles. Further, as a matter of practical consequence, certifying an overinclusive class may invite extensive post-certification challenges and appeals, wasting precious judicial resources.[130] And because plaintiffs' inability to demonstrate classwide harm raises doubts about classwide causation and redressability—which are already huge contentions in the AI copyright infringement context—courts may well deny certification on prudential ascertainability grounds to avoid reaching the constitutional question of standing.[131]

C.   *Statutory Damages Are an Inadequate and Inequitable Pathway to Certification*

Over the course of this Note, readers familiar with copyright class action litigation might find themselves wondering, "Well, what about statutory damages under the Copyright Act?"[132] Instead of struggling to sustain the highly individualized Rule 23(b)(3) inquiry for actual damages, a putative class could instead elect statutory damages under 17 U.S.C. § 504.[133] Since statutory damages provide a predetermined range of recovery on a per-work basis without the need to prove actual harm, this alternative seems to obviate the most pressing causation and scope of injury concerns.[134] Similar to wage-and-hour disputes, damages in AI copyright class actions could be reframed to center on the culpability of the defendant's violative conduct.[135] In a best case scenario for plaintiffs, this could pave the way for allowing representative evidence sampling techniques in AI copyright class actions. Indeed, in recognition of these substantial benefits, several amended complaints in AI copyright class actions have expressly requested

---

[130] *See generally Wal-Mart*, 564 U.S. at 367.

[131] Rule 23 might be characterized as a tool of constitutional avoidance in these types of cases. *See, e.g.*, *Amchem*, 521 U.S. at 612-13 (agreeing that class certification issues are "logically antecedent to the existence of any Article III issues," and that "it is appropriate to reach them first"); *Marcus*, 687 F.3d at 583 (vacating class certification because of individualized injury and causation inquiries).

[132] Although statutory damages are not central to the purposes of this Note, it would be remiss to omit them from the discussion entirely.

[133] 17 U.S.C. § 504 (2018).

[134] *See* 17 U.S.C. § 504(c) (2018) (outlining statutory damages for "any one work").

[135] *Compare* 17 U.S.C. § 504(c)(2) (2018) (giving courts discretion within the statutory range to tailor damages based on the infringer's guilty mens rea), *with* 29 U.S.C. § 260 (2018) (giving courts discretion to reduce or eliminate liquidated damages if the employer sustains a good faith defense).

a maximum award of statutory damages "in the alternative to actual damages and profits."[136]

Unfortunately, while statutory damages may be viable for a handful of tightly constructed and well-resourced putative classes, most plaintiffs in AI copyright class actions will not be able to avail of this remedy for one key reason: in order to be eligible for statutory damages in the first instance, copyright owners must have registered each infringed work with the U.S. Copyright Office *prior* to the commencement of the infringement.[137] If a work is unregistered or was registered only *after* being scraped into a Gen-AI training dataset, its creator is wholly precluded from statutory damages and may only pursue actual damages.[138] Plaintiffs in this posture are in a catch-22: statutory damages are barred, while actual damages are practically impossible to prove.[139]

This pre-registration requirement is problematic for AI copyright class action plaintiffs in multiple ways. For one, not everyone has the legal know-how to formally and preemptively register all of their creations with the U.S. Copyright Office. Many creators mistakenly believe that their works are protected by copyright law without any formal registration; others are forced to forego registration in light of seemingly insurmountable administrative burdens.[140] This is especially true for works created and disseminated online, where the ongoing

---

[136] *See, e.g.*, Amended Complaint at 51, Authors Guild v. OpenAI Inc., No. 1:23-cv-08292 (S.D.N.Y. Dec. 5, 2023); Second Amended Complaint at 57, 59, 64, Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal. Oct. 31, 2024).

[137] 17 U.S.C. § 412 (mandating registration with the U.S. Copyright Office as a prerequisite to certain remedies including "award of statutory damages . . . as provided by [section] 504").

[138] *See* Pamela Samuelson, *How to Think About Remedies in the Generative AI Copyright Cases*, COMM. OF THE ACM (June 11, 2024), https://cacm.acm.org/opinion/how-to-think-about-remedies-in-the-generative-ai-copyright-cases/ [https://perma.cc/HC99-JZED] ("Most class action plaintiffs will not qualify for copyright statutory damage awards, even though some will claim them anyway.").

[139] *Id.* ("Actual damages . . . may be miniscule or non-existent, although the generative AI plaintiffs argue the outputs may reduce demand for the originals."); *see generally* Part II-A, *supra* (outlining difficulties of proving actual harm).

[140] *See* Janet Fries and Jennifer T. Criss, *Debunking Copyright Myths*, AM. BAR ASS'N (Mar. 2021), https://www.americanbar.org/groups/intellectual_property_law/resources/landslide/archive/debunking-copyright-myths/ [https://perma.cc/AJ5R-UP7M] (clarifying that "true ownership in a copyrightable work requires official registration with the U.S. Copyright Office"). *But see also* Unicolors v. H&M Hennes & Mauritz, L.P., 595 U.S. 178,182 (2022) (holding that a copyright registration may not be invalidated if inaccurate information was included on the application, as long as the mistake was made in good faith and the copyright holder lacked "knowledge that it was inaccurate").

question of "*When does uploading a work to the internet constitute publication?*" continues to confound even seasoned attorneys and copyright experts.[141] Absent legal background or counsel, individual creators stand little chance to avail of the legal benefits afforded by official registration. The complexities of the copyright registration process ensure that, perhaps by design, "the population of copyrighted works is greater than registered ones."[142]

Equally important is the fact that official copyright registration is not free. The financial burden of copyright registration is a significant barrier for high-volume and/or low-income creators.[143] The standard application fee is $65 to register an original work of authorship, $55 for a discrete "group" of photographs, and $65 for a music album.[144] Pursuant to this fees schedule, high-volume individual creators simply cannot afford to register *every* protectable creation.[145] One can easily conjure a scenario where an amateur poet, hobbyist photographer, or aspiring songwriter faces the prospect of paying thousands of dollars to register their expansive œuvre. For many Americans, this level of expense is just not feasible.[146] The registration prerequisite makes it much more likely that only entities with preexisting resources—newspapers, corporations, famous published authors—can avail of statutory damages in the AI copyright context.

For the foregoing reasons, putative plaintiffs cannot consistently rely on the crutch of statutory damages to overcome certification in AI copyright class actions. As a matter of practicality, after *TransUnion*, an action built solely on statutory

---

[141] Terry Hart, *This One Weird Trick Could Improve Copyright Registration*, COPYRIGHT ALLIANCE (Jan. 28, 2019), https://copyrightalliance.org/trick-improve-copyright-registration/ [https://perma.cc/4DMW-62SV].

[142] Dotan Oliar et al., *Copyright Registrations: Who, What, When, Where, and Why*, 92 TEXAS L. REV. 2211, 2241 (2014).

[143] *See* Hart, *supra* note 141 (asserting that "creators often have to weigh the costs and benefits" of registration).

[144] Fees Schedule, U.S. COPYRIGHT OFFICE, https://www.copyright.gov/about/fees.html [https://perma.cc/8VCN-T92H] (last visited Feb. 23, 2025).

[145] *See generally* Robert Brauneis, *Properly Funding the Copyright Office: The Case for Significantly Differentiated Fees*, GWU L. Sch. Pub. L. Res. Paper No. 2017-58, GWU Legal Studies Res. Paper No. 2017-58 (July 4, 2017), available at https://ssrn.com/abstract=2997192 (proposing reforms to make registration more affordable for modest applicants).

[146] According to the U.S. Bureau of Labor Statistics, the median annual wage for "craft and fine artists" in 2024 was $56,260. *Craft and Fine Artists*, Occupational Outlook Handbook, U.S. BUREAU OF LAB. STAT., https://www.bls.gov/ooh/arts-and-design/craft-and-fine-artists.htm [https://perma.cc/FX9D-BD7T] (last visited Jan. 22, 2026).

damages could open an uncomfortable can of worms that would best be avoided by AI plaintiffs.[147] As a matter of doing justice, the potential for inequity is damning. In order to maximize odds of recovery for statutory damages, future putative classes will be forced to strategically limit membership to those with the resources and wherewithal to "timely register[ ] their copyrights in their works with the Copyright Office."[148] Consequently, many potential plaintiffs will not qualify for relief when their claims are just as meritorious as those of the preregistered class members. This disparity in AI copyright class membership will disproportionately affect the segments of artists, authors, and other creators who are most in need of the 23(b)(3) class action device—those "who individually [are] without effective strength to bring their opponents into court at all."[149]

III
LEVERAGING JUDICIAL DISCRETION TO EXPLORE ALTERNATIVE PATHWAYS TOWARD CLASS CERTIFICATION

The unique procedural challenges faced by the proposed classes in AI copyright litigation demand that courts approach certification with creativity and flexibility. District judges retain a remarkable degree of discretion over managing their dockets and adapting procedural tools to novel litigation, especially in situations that lack controlling precedent. This Note proposes that the courts presiding over AI copyright class actions should leverage their discretion under the Federal Rules to help remedy existing deficiencies at the certification stage. Of course, "[c]ourts are not free to amend [the Federal Rules of Civil Procedure] outside the process Congress ordered."[150] The letter of the law thus "limits judicial inventiveness."[151] But in the class certification context, "[t]rial courts are given substantial discretion in determining whether to grant class certification."[152] Rather than rewriting the rules of procedure wholesale, district judges can apply existing procedural mechanisms in a creative and flexible manner.

---

[147] *See generally* Moss, *supra* note 114.

[148] Amended Complaint at 48, Authors Guild v. OpenAI Inc., No. 1:23-cv-08292 (S.D.N.Y. Dec. 5, 2023). This has already happened in some well-resourced pending actions such as *Authors Guild. Id.*

[149] *Amchem*, 521 U.S. at 617 (internal quotation marks and citation omitted) (summarizing the 1966 Advisory Committee's original intent in creating the Rule 23(b)(3) cause of action).

[150] *Amchem*, 521 U.S. at 620.

[151] *Id.*

[152] *In re MF Global*, 310 F.R.D. at 235.

This final Part advances a two-pronged approach to increase the odds of certification while also distinguishing and deterring meritless claims. First, Rule 23(c)(4)(A) issue-class certification should be liberally granted to resolve common liability questions. Second, on the basis of these commonly resolved issues, courts should consider the suitability of appointing Rule 53 special masters to assist with technical complexities such as the resolution of data-intense e-discovery disputes. Federal judges have historically employed these procedural devices in other complex areas of law—such as securities fraud and antitrust litigation—as well as intellectual property disputes more broadly.[153] By extending these effective tools to AI copyright class actions, a federal judge can enhance fairness, maintain doctrinal integrity, and reserve the class action device for situations where it is particularly apt.

A. Employing Rule 23(c)(4)(A) Issue Subclass Certification as a Scalpel to Isolate and Resolve Common Liability Questions Within the Boundaries of Wal-Mart

District courts have discretion under Rule 23(c)(4)(A) to certify a partial liability-based issue subclass in situations where individualized inquiries preclude traditional certification under Rule 23(b)(3).[154] The Supreme Court has declined to define the scope of Rule 23(c)(4)(A) analysis in relation to Rule 23(b)(3) predominance, resulting in a circuit split below.[155] A growing majority of courts endorse a liberal interpretation of Rule 23(c)(4)(A) that permits certification of a particular issue subclass even when the global action fails to satisfy the broader Rule 23 certification requirements.[156] Most relevant to AI copyright class actions,

---

[153] *E.g.*, Fikes Wholesale, Inc. v. Visa USA, N.A., 62 F.4th 704, 718-19 (2d Cir. 2023) (affirming district court's appointment of Rule 53 special master to "decid[e] antitrust standing issues"); Louis Vuitton Malletier v. Dooney & Bourke, Inc., 525 F. Supp. 2d 558, 563-64 (S.D.N.Y. Jun. 15, 2007) (appointing special masters to review expert evidence and testimony in trademark infringement lawsuit).

[154] Fed. R. Civ. P. 23(c)(4) advisory committee's note to 1966 amendment (describing Rule 23(c)(4)(A) as a tool for adjudicating liability common to a class); *see* In re Nassau Cnty. Strip Search Cases, 461 F.3d 219 (2d Cir. 2006) [hereinafter *Strip Search Cases*] (holding that a court may employ Rule 23(c)(4)(A) to certify a class on an issue even if the claim as a whole does not meet predominance).

[155] Behr Dayton Thermal Prod. LLC v. Martin, 896 F.3d 405 (6th Cir. 2018), *cert. denied*, 139 S. Ct. 1319 (2019).

[156] *See* Scott Dodson, *Subclassing*, 724 W&M Faculty Pub. 2351, 2375 (2006).

the Second and Ninth Circuits do not require a certified 23(c)(4)(A) "issue" to predominate.[157]

In the *Strip Search Cases*, for instance, the plaintiffs requested to certify a Rule 23(b)(3) damages class "solely on the issue of liability" after multiple motions for certification were denied for lack of predominance.[158] The district court declined to certify the issue subclass, expressing "doubt over whether it could certify a class on the issue of liability" after the "claims, as a whole, failed the predominance test."[159] On appeal, the Second Circuit reversed. It held that as a matter of the Rule's plain text and structure, "a court may employ Rule 23(c)(4)[ ] to certify a class as to an issue *regardless of whether the claim as a whole satisfies the predominance test*."[160] This approach allows putative classes in otherwise hard-to-certify cases— such as AI copyright actions—to partially and temporarily sidestep the gatekeeper of Rule 23(b)(3) predominance by instead settling baseline theories of liability at the outset.[161]

Federal courts should embrace the pragmatic advantages of Rule 23(c)(4)(A) in the AI copyright context by certifying issue subclasses *sua sponte* to alleviate intra-class incompatibilities. Isolating a common liability question at the outset plays to the strengths of aggregated litigation. Resolving the issue in the defendant's favor can obviate the need for costly discovery and deter future meritless claims; conversely, resolving the issue in the plaintiffs' favor can streamline future litigation, be it individual suits or a newly refined putative class.[162] In this way, the merits of preclusive nonmutual collateral estoppel—namely judicial efficiency and consistency—effectively extend to the application of Rule 23(c)(4)(A) no matter the direction in which the pendulum swings.

---

[157] *See Strip Search Cases*, 461 F.3d at 231 (2d Cir. 2006); Valentino v. Carter-Wallace, Inc., 97 F.3d 1227,1234 (9th Cir. 1996).

[158] *Strip Search Cases*, 461 F.3d at 221.

[159] *Id.*

[160] *Id.* (emphasis added); *see also id.* at 226 (rejecting Fifth Circuit's "strict application" of 23(b)(3) predominance).

[161] *See* Mark A. Perry, *Issue Certification Under Rule 23(c)(4): A Reappraisal*, 62 DePaul L. Rev. 733 (2013).

[162] *See* Mejdrech v. Met-Coil Sys. Corp., 319 F.3d 910, 911 (7th Cir. 2003) ("If there are genuinely common issues ... then it makes good sense, especially when the class is large, to resolve those issues in one fell swoop while leaving the remaining, claimant-specific issues to individual follow-on proceedings.").

We can illustrate this approach with a hypothetical AI copyright dispute in which an AI developer is being sued by a proposed class of visual artists for input-side copyright infringement. Assuming the defendant's Gen-AI dataset training methodology remained uniform across all class members' claims, then determining whether such methodology inherently violated the plaintiffs' exclusive rights is a legal matter that does not require individualized consideration of factual causation or injury.[163] As a single theory of liability, this is an ideal use case for Rule 23(c)(4). A ruling in favor of the defendant developer would preclude costly and repetitive suits on the same legal question, while a ruling in favor of the plaintiff artists would establish common injury in law, which may then serve as a foundation for future showings of injury in fact.

Another reason for courts to liberally invoke Rule 23(c)(4)(A) is that it is consistent with principles of *Wal-Mart*. Let's be real—no district judge wants to adopt a new practice, only to be reversed on appeal for purported abuse of discretion.[164] Such reversal need not be a concern for judges in the Second and Ninth Circuits. Not only do these jurisdictions expressly endorse broad 23(c)(4)(A) discretion, but issue subclass certification itself does not inherently circumvent *Wal-Mart*.[165] Rather, Rule 23(c)(4)(A) acknowledges and defers to the determination that classwide treatment is presently inappropriate in the instant case. *Wal-Mart* emphasized the necessity of a "common contention" whose resolution "resolve[s] an *issue* that is central to the validity of each one of the claims in one stroke."[166] Issue subclass certification addresses and resolves precisely this type of "common contention." In the AI developer hypothetical above, the question of copyright infringement can be stripped to its essence: Did the defendant's uniform course of conduct—the act of feeding copyright-protected artworks into a Gen-AI training dataset—constitute a legally cognizable violation of the plaintiffs' exclusive rights under 17 U.S.C. § 106? If so, then the proposed class members have a *common* legal basis for their claims. If not, then the proposed class fails as a legal *collective*.

---

[163] *See* 17 U.S.C. § 106 (2023).

[164] *See* Richard A. Posner, *How Judges Think* 142 (2008) (noting that trial judges may act cautiously in discretionary matters to avoid reputational and procedural costs of appellate reversal).

[165] *See* United States v. City of N.Y., 276 F.R.D. 22, 34 (E.D.N.Y. July 8, 2011) ("[T]he Second Circuit's interpretation of Rule 23(c)(4)(A) is consistent with *Wal-Mart*'s interpretation of Rule 23(b)").

[166] *Wal-Mart*, 564 U.S. at 350.

Some courts and critics may worry that liberal grants of issue class certification under Rule 23(c)(4)(A) will lead to unmanageable, inefficient, and fragmented litigation. This critique centers on the concern that, in complex actions with multiple threshold questions of liability, courts must manage piecemeal adjudication of issues. Furthermore, conclusive determinations of liability on central issues could open the floodgates of litigation, potentially spawning hundreds of subsequent individualized cases.[167] This subverts one of the core purposes of the class action: judicial economy.

While manageability concerns are important, they are overblown in the context of AI copyright infringement litigation. In the majority of AI copyright class actions, the central liability question comes down to whether the defendant's AI training practices infringe upon the plaintiffs' copyrights.[168] Compared to cases involving diverse factual patterns or multiple liability theories, this is typically a binary, classwide issue that lends itself naturally to Rule 23(c)(4)(A) issue certification. In the rare cases with multiple complex threshold questions of liability, courts can and should exercise their innate discretion to ensure that the scope of conditional certification, if any, is appropriately tailored.[169]

As a matter of caselaw precedent, circuit courts have routinely affirmed the discretionary certification of 23(c)(4)(A) issue subclasses in the face of manageability concerns.[170] So too has the Federal Judicial Center's *Manual for Complex Litigation* urged courts to consider issue subclass certification where a case may otherwise "be unmanageable as a class action."[171] As one example, the Fourth Circuit directly addressed the "concern about [ ] manageability" in *Central*

---

[167] *See generally* Laura J. Hines, *The Dangerous Allure of the Issue Class Action*, 79 Ind. L.J. 567, 586-88 (2004) (arguing against expansive interpretation of issue subclass certification).

[168] *See generally The Battle Over AI Training Data: Copyright, Fair Use, and the Future of GenAI*, Dykema (Feb. 8, 2024), https://www.dykema.com/news-insights/the-battle-over-ai-training-data-copyright-fair-use-and-the-future-of-genai.html [https://perma.cc/LD2R-4NXY] (outlining infringement and fair use liability arguments).

[169] *Cf.* Jacks v. DirectSat USA, LLC, 118 F.4th 888, 898 (7th Cir. 2024) (decertifying issue class for overbroad scope "as to the entire cause of action" instead of limiting to the "fourteen certified issues").

[170] *E.g.*, Black v. Occidental Petroleum Corp., 69 F.4th 1161, 1169 (10th Cir. 2023) (affirming liability issue subclass certification); Cent. Wesleyan Coll. v. W.R. Grace & Co., 6 F.3d 177, 188-89 (4th Cir. 1993) (upholding issue certification order in the face of a "daunting number of individual issues" necessary to "establish liability"); *Strip Search Cases*, 461 F.3d at 221 (same). *But see* Castano v. Am. Tobacco Co., 84 F.3d 734 (5th Cir. 1996) (requiring issue subclasses to meet predominance and manageability).

[171] Manual for Complex Litigation, Third, § 30.17 (1995).

Wesleyan.[172] As is often the case in asbestos litigation, the *Central Wesleyan* plaintiffs' individual issues predominated over common ones under Rule 23(b)(3). To navigate this ordinarily action-defeating deficit, the district court certified eight common liability issues under Rule 23(c)(4)(A).[173] Defendants appealed this decision on the basis that "[certification of] individual issues will swamp the litigation and make it unmanageable."[174] The Fourth Circuit acknowledged the existence of these substantial manageability concerns in the asbestos mass tort context, but ultimately deferred to the district court's discretion under the Federal Rules: "[t]he tentative, limited nature of the conditional certification … counsels in favor of affirmance."[175]

Issue subclass certification was similarly sustained over an unmanageability objection in *Gunnells*, a dispute centered on the mismanagement and collapse of a multi-employer healthcare plan.[176] In response to classwide manageability and predominance deficiencies, the district court "[took] full advantage of the provision in [Rule 23(c)(4)(A)] permitting class treatment of separate issues … to reduce the range of disputed issues."[177] On appeal, the Fourth Circuit affirmed again. This time, they reasoned that manageability issues in issue subclasses are often offset by the likelihood that the "consolidation of recurring common issues" will reduce litigation costs in the long run, thus "conserv[ing] important judicial resources."[178] From a policy and practicality standpoint, the Fourth Circuit also noted that without their liberal approval of conditional issue class certification, "very few claims would be brought against [corporate Defendants], making 'the adjudication of [the] matter through [an issue] class action … superior to no adjudication of the matter at all.'"[179]

---

[172] *Central Wesleyan*, 6 F.3d at 190.

[173] *Id.* at 181-82 (discussing the unmanageability of asbestos litigation generally); *see also Amchem*, 521 U.S. at 624 (finding that "any overarching dispute about the health consequences of asbestos exposure cannot satisfy the Rule 23(b)(3) predominance standard").

[174] *Central Wesleyan*, 6 F.3d at 186.

[175] *Id.*

[176] Gunnells v. Healthplan Serv., Inc., 348 F.3d 417 (4th Cir. 2003).

[177] *Id.* at 426 (citing intra-circuit precedent for allowing issue class certification under Rule 23(c)(4)).

[178] *Id.* (quoting *Central Wesleyan*, 6 F.3d at 185).

[179] *Id.* (quoting 5 *Moore's Federal Practice* § 23.48[1] (1997)).

Charron provides a jurisdictionally relevant illustration within the Second Circuit.[180] Here, tenants brought a putative class action against their corporate landlord, alleging numerous RICO and NYCPA violations. Because each tenant's factual circumstance was unique, certification as a Rule 23(b)(3) damages class was plainly improper.[181] Instead, the court identified the one issue that, "contrary to Defendants' assertion," was common to all Plaintiffs—Defendants' "*same general course* of allegedly fraudulent and harassing conduct, *the same pattern* of racketeering."[182] The court certified a Rule 23(c)(4)(A) "Liability Class" to examine the core liability question of whether Defendants had violated the RICO and/or NYCPA statutes.[183] With an eye toward future manageability, the court expressly cautioned that "in the event that Defendants' liability is established," it would have to "reevaluate its options for managing the damages portion of the action."[184] Still, *Charron* further legitimized the use of Rule 23(c)(4)(A) issue subclass certification in a context analogous to that of the AI copyright actions— *i.e.*, predominant individualized causation and actual damages, with a central liability issue premised on defendants' statutory violation. There is no reason why *Charron*'s careful balancing of liability and manageability in the Rule 23(c)(4)(A) context cannot be extended to certification of AI copyright infringement issues.

In the context of manageability, it also bears reiterating that district courts retain discretion to decertify an issue class *sua sponte*. If, at any time in the litigation, "manageability and other types of problems … overwhelm the advantages of conditional certification," the court is *obligated* to decertify the Rule 23(c)(4)(A) class.[185] This obligation to decertify down the line allows courts to be more permissive of issue class certification at the outset, since it functions as a built-in safeguard for courts to reverse course if a certified action later becomes unmanageable. The stakes of initial Rule 23(c)(4)(A) certification are thus lowered. A number of courts regularly and effectively exercise this procedural safeguard

---

[180] Charron v. Pinnacle Gr. N.Y. LLC, 269 F.R.D. 221 (S.D.N.Y. 2010).

[181] *See id.* at 226 (noting proposed class members' "different apartments, different leases, different alleged misrepresentations, different modes of alleged harassment, and so forth").

[182] *Id.* at 226-227 (emphasis added).

[183] *Id.* at 240.

[184] *Id.* at 227.

[185] *Central Wesleyan*, 6 F.3d at 189 (citing 7B Wright, Miller & Kane, *Federal Practice and Procedure* § 1785, at 128-36 (1986)); *accord* In re School Asbestos Litig., 789 F.2d 996, 1011 (3d Cir. 1986) ("When, and if, the district court is convinced that the litigation cannot be managed, decertification is proper.").

to maintain integrity throughout all stages of the class action, demonstrating its workability as a judicial tool.[186]

At the end of the day, courts should not prioritize efficiency over justice. The Second Circuit said as much when it explicitly instructed district courts "to adopt a liberal interpretation of Rule 23" in favor of certification when possible.[187] This instruction naturally envelopes the penumbras of subsection (c)(4)(A), as we saw in the *Strip Search Cases*.[188] Putative plaintiffs in AI copyright class actions will be glad to know that courts in the Southern District of New York have felt regularly "empowered" to "certify a class on a particular issue," even when not affirmatively asked to do so by plaintiffs.[189] Overall, both precedent and prudency support the premise that Rule 23(c)(4)(A) issue subclass certification augments justice in procedurally deficient cases while remaining within the controlling doctrinal limits of *Wal-Mart*.

Because issue subclass certification generally favors plaintiffs, critics may also accuse Rule 23(c)(4)(A) of unfairly disadvantaging defendants.[190] It goes without saying that even if most defendants in AI class actions are large and well-resourced, they are entitled to fundamental procedural fairness. This principle is especially important in Rule 23(b)(3) damages class actions. Since the 1990s, a significant concern—and one of the motivating factors behind *Wal-Mart* and *TransUnion*—has been plaintiff-side (or, more accurately, plaintiff's attorney-side) abuse of meritless lawsuits to coerce pecuniary settlements from large companies.[191] This settlement pressure dilemma is most severe in emerging industries such as AI, where legal uncertainty exacerbates the "bet the company" stakes of adverse liability findings.[192] As such, risk-averse AI copyright defendants

---

[186] *E.g.*, Harris v. Med. Transp. Mgmt., Inc., 77 F.4th 746, 752 (D.C. Cir. 2023) (vacating and remanding lower court's issue class certification); *Jacks*, 118 F.4th at 898-99.

[187] *In re MF Global*, 310 F.R.D. at 235.

[188] *Strip Search Cases*, 461 F.3d at 221.

[189] *See* Nnebe v. Daus, 2022 WL 615039, at 17 (S.D.N.Y. Mar. 1, 2022).

[190] *See generally* Hines, *supra* note 167, at 586-87 (finding expansive issue class actions unfair and subversive).

[191] *See* S. Rep. No. 109-14, at 84 (2005) (Senate Judiciary Committee Report on the Class Action Fairness Act of 2005 (CAFA), discussing abuse of class actions to force over-settlement of "meritless" suits). *See also generally* Robert H. Klonoff, *The Decline of Class Actions*, 90 Wash. U. L. Rev. 729, 777-80 (2013) (discussing *Wal-Mart*'s purposeful chilling effect on class action litigation).

[192] *See* David M. McIntosh et al., *AI and the Copyright Liability Overhang: A Brief Summary of the Current State of AI-Related Copyright Litigation*, Ropes & Gray (Apr. 2, 2024), https://www.ropesgray.com/en/ins

may be perversely incentivized to enter into what some commentators have deemed "blackmail" settlements.[193]

But this accusation of unfairness mischaracterizes the root cause of settlement pressure. Defendants are pressured into uneven settlements when they stand to lose more than they would gain by continuing litigation. These litigation costs primarily stem from the "gargantuan scale" of class action discovery, not certification of the class.[194] As a rule, the more relevant and exclusive information is in the hands of the defendant, the greater the discovery costs.[195] Because courts often authorize discovery on the merits before ruling on Rule 23(c) certification, a corporate defendant bears the risk of heavy litigation costs regardless of the ultimate merits ruling. For defendants in this posture, dismissal on the merits *after* undergoing lengthy discovery amounts to nothing more than a Pyrrhic victory. This is often the case in antitrust class actions, where defendants stand to shell out billions in dollars and documents (in addition to the risk calculus of treble damages).[196] AI copyright class actions are similar to antitrust actions in that almost all relevant information—training datasets, methodologies, algorithms—is exclusively held by defendants, resulting in immense frontloaded discovery costs and initially hard-to-detect violations.

Because discovery costs exist largely independent of class certification, Rule 23(c)(4)(A) would not materially disadvantage defendants or exacerbate settlement pressure in AI copyright class actions. If anything, certification of an issue subclass might *alleviate* settlement pressure by streamlining pre-certification discovery to focus on a single narrow issue, rather than compelling the defendant to disclose *all* marginally relevant documents. Moreover, an early resolution of liability often *benefits* defendants in the settlement context. In *Gunnells*, for example,

---

ights/alerts/2024/04/ai-and-the-copyright-liability-overhang-a-brief-summary-of-the-current-state-of-ai-related [https://perma.cc/66NL-NECL].

[193] *Cf.* Nicholas Almendares, *The False Allure of Settlement Pressure*, 50 Loyola U. Chi. L.J. 271 (2018).

[194] *Id.* at 290; *see also* Hasan Chowdhury, *OpenAI's Lawyer Says There are Too Many Files from Ilya Sutskever and Other Employees to Share in Copyright Lawsuit*, Business Insider (Nov. 14, 2024, 9:13 AM), https://www.businessinsider.com/openai-share-files-copyright-case-authors-guild-ilya-sutskever-2024-11 [https://perma.cc/UD52-8A6S].

[195] Almendares, *supra* note 193, at 290.

[196] *See generally* Christopher R. Leslie, *De Facto Detrebling: The Rush to Settlement in Antitrust Class Action Litigation*, 50 Ariz. L. rev. 1009, 1014-15 (2008) (noting risks of "ill-gotten," "frivolous" antitrust settlements).

the Fourth Circuit noted that by "provid[ing] a single proceeding in which to determine the merits of the plaintiffs' claims," Rule 23(c)(4)(A) issue subclass certification "therefore protects *the defendant* from inconsistent adjudications."[197] A favorable determination allows defendants to preclude future litigation—including settlements—against all class members on that liability finding.

Comparatively, litigating an issue over and over again in individualized trials is much riskier because of the one-way asymmetry of collateral estoppel—plaintiffs would be able to avail of offensive issue preclusion without being bound by the defendant's victories in other judgments.[198] Out of thousands of potential trials, a single loss would force the defendant to concede that issue in all future disputes, placing them at a huge disadvantage when entering into settlement negotiations. As a consequence of the large classes and finite number of liability questions in AI copyright infringement actions, winning on a single certified issue could save the defendant millions of dollars in long-term settlements and litigation fees. In this way, Rule 23(c)(4)(A) "promotes consistency of results, giving defendants the benefit of finality and repose."[199]

We conclude by acknowledging that some courts are hesitant to certify issue classes in emerging areas of law and technology that are better served through legislative solutions. Judicial overreach is always a prudent concern when breaking new ground. AI copyright infringement disputes are particularly tricky because they implicate an enumerated power in Article I, Section 8, Clause 8 of the Constitution.[200] The Supreme Court has long read fair use exceptions into the Copyright Clause, which many defendants in AI copyright class actions have invoked in their defense.[201] However, neither the Court nor Congress has clarified whether the fair use exceptions can or should extend to copyright infringement

---

[197] *Gunnells*, 348 F.3d at 427 (citing 5 *Moore's Federal Practice* § 23.02 (1999)) (emphasis in original).

[198] *See* Parklane Hosiery Co. v. Shore, 439 U.S. 322, 331 (1979).

[199] *Gunnells*, 348 F.3d at 427.

[200] U.S. Const. art. I, § 8, cl. 8.

[201] *See* 17 U.S.C. §§ 107-122 (2018); *see also* Vincent, *supra* note 7; Kupferschmid, *supra* note 7.

in the Gen-AI context.[202] This unanswered constitutional dimension may trigger judicial instincts of avoidance.

However, it is important to remember that district courts regularly interpret and apply existing laws to unprecedented issues in the information age. This was evident in *Andersen* when the defendants filed a motion to dismiss the plaintiffs' claims of direct copyright infringement on the basis that from a technological standpoint, they only copied unprotected, non-infringing "data" stored as statistical representations in Gen-AI models.[203] Judge Orrick in the Northern District of California permitted the core copyright claims to proceed, giving the plaintiffs space to plead "claims against any defendant based on theories (if any) that are not preempted by the Copyright Act" outright.[204] However, he also cautioned that existing copyright infringement precedent may have little influence on AI cases given the unprecedented nature of the technology.[205] This even-keeled approach reflects judicial modesty without downright dismissing the plaintiffs' pleas for relief. Judge Orrick's decision to proceed case-by-case, while remaining mindful of AI's unique nature and avoiding premature doctrinal leaps, demonstrates how courts can be pragmatic—even progressive—in their discretion without overstepping the broader judicial role.

Judge Orrick's general reasoning can be extended to issue subclass certification under Rule 23(c)(4)(A). In both circumstances, the court is not creating *new* law, but rather applying *existing* principles of copyright and procedure to emerging technologies while acknowledging that doctrinal differences may emerge. We know that the common liability issues in AI copyright class actions are not so different from those certified in cases like *Charron*, making them relatively safe targets for issue subclass certification. On the other hand, the most glaringly unanswered questions in AI copyright class actions—individualized questions

---

[202] Julia Barnett, *Decoding US Copyright Law and Fair Use for Generative AI Legal Cases*, Medium (July 11, 2024), https://generative-ai-newsroom.com/decoding-us-copyright-law-and-fair-use-for-generative-ai-legal-cases-507fdfd9956c [https://perma.cc/FH9S-VMSJ] ("None of the cases have thus far had any significant rulings regarding whether the use of copyrighted works in generative AI can be considered fair use, and this is the key thing to watch going forward.").

[203] *Andersen*, 2024 WL 3823234, at 977.

[204] *Id.* at 986 n.6.

[205] *See* Kevin Madigan, *AI Lawsuit Developments in 2024: A Year in Review*, Copyright Alliance (January 9, 2025), https://copyrightalliance.org/ai-lawsuit-developments-2024-review/ [https://perma.cc/JN8F-ANGL].

of causation, damages, and standing that have few guiding principles—should continue to remain outside the scope of Rule 23(c)(4)(A). If or when Congress decides to update the law of copyright, as is their right under the Constitution, district courts will be able to adjust accordingly without needing to backtrack on sweeping doctrinal missteps. Accordingly, liberal usage of Rule 23(c)(4)(A) is not inconsistent with procedural and doctrinal integrity.

As a district judge, there is little reason to deny certification of Rule 23(c)(4)(A) issue subclasses in otherwise uncertifiable AI copyright class actions. To be sure, issue subclass certification invites more work and burns more resources than simply dismissing for classwide certification defects. But outright dismissal ignores the fact that the claims of many putative class members in AI copyright actions *do* have merit—they just struggle to prove it at the outset given the uniquely restrictive nature of the Gen-AI data training process, which compounds on the plaintiffs' relative lack of resources.[206] For this reason, Rule 23(c)(4)(A) issue subclass certification should be generously invoked to provide reasonable and just guidance in AI copyright class actions. For many deserving plaintiffs, the natural information deficit is an otherwise insurmountable impediment to certification.

B.   Appointing Rule 53 Masters to Bridge Technical Knowledge Gaps in Pre-certification e-Discovery and Beyond

Once the scope of potential litigation is clarified, district courts must assess the pros and cons of additional management resources.[207] If and when AI copyright class actions proceed to precertification discovery, Rule 53 masters are a valuable resource to bridge technical knowledge gaps and monitor alignment with certified issues. Although Rule 53 was originally geared toward magistrate-style special masters with trial duties, the most recent substantive revisions from 2003 acknowledge the modern prevalence of "masters appointed to perform a variety of pretrial . . . functions,"[208] including "outside masters [who] may prove useful when some special expertise is desired[]."[209] Under the revised Rule 53, the appointment

---

[206] *See* Ahmad & Gross, *supra* note 5.

[207] Note that both Rule 23(c)(4)(A) issue subclass certification and Rule 53 appointments should be contemplated on a case-by-case basis to address the specific needs of the pending litigation. This Note remains skeptical of a blanket one-size-fits-all approach in the novel field of AI copyright litigation.

[208] FED. R. CIV. P. 53 advisory committee's notes to 2003 amendment.

[209] FED. R. CIV. P. 53 advisory committee's notes to 1983 amendment. The current iteration of Rule 53 allows broad discretion to appoint pretrial masters, but narrow discretion to appoint trial masters.

of expert pretrial masters is now expansively governed by subsection (a)(1)(C), whereas it was previously limited by the "exceptional circumstances" condition of subsection (a)(1)(B).[210] Accordingly, modern courts have broad discretion under the text of Rule 53 to appoint expert masters who are tailored to the unique needs of the instant case.

Precedent supports the appointment of expert masters in the pretrial and discovery stages of complex litigation.[211] Even before the removal of Rule 53's "exceptional circumstances" limiting principle, courts in the Southern District of New York endorsed the appointment of expert masters to "supervise discovery proceedings where the issues are complicated or the parties recalcitrant."[212] Similarly, while upholding the appointment of an expert master in an antitrust class action, the Second Circuit recently affirmed that "district courts with loaded dockets may rely on special masters to decide thorny things."[213] Beyond case law, the Federal Judicial Center's *Manual for Complex Litigation, Fourth* strongly recommends the appointment of Rule 53 masters and other court-appointed experts to handle technical aspects of discovery and report findings of fact and law relevant to class certification.[214]

For an illustration of exactly how expert masters may be useful in AI copyright class actions, consider the ongoing e-discovery disputes in *Authors Guild*. The litigants in these consolidated suits have clashed for months over data relevance, technical feasibility, and the burden of producing enormous amounts of sensitive proprietary information.[215] This has led to a series of litigious back-

---

[210] *See id*. Hereafter, this Note will refer to the subset of pretrial masters as "expert masters."

[211] *But see* La Buy v. Howes Leather Co., 352 U.S. 249, 259 (1957). Although *La Buy* dictated that special masters should be used sparingly and only when necessarily, this was only in the context of magistrate-style masters rather than technical-expert masters. As such, it is not controlling here.

[212] Fisher v. Harris, Upham & Co., 61 F.R.D. 447, 449 (S.D.N.Y. 1973) (collecting cases).

[213] *Fikes Wholesale*, 62 F.4th at 718-19.

[214] *See* Manual for Complex Litigation, Fourth, § 11.52 (observing that the line between expert masters under Rule 53 and court-appointed experts under Federal Rule of Evidence 706 has become increasingly blurred in practice).

[215] *See* Hasan Chowdhury, *OpenAI's Lawyer Says There are Too Many Files from Ilya Sutskever and Other Employees to Share in Copyright Lawsuit*, Business Insider (Nov. 14, 2024, 9:13 AM), https://www.businessinsider.com/openai-share-files-copyright-case-authors-guild-ilya-sutskever-2024-11 [https://perma.cc/6DPM-X44R].

and-forth discovery motions.[216] An expert master with technical expertise in the fields of artificial intelligence and data science could moderate such disputes in a number of ways, saving a district judge the headache of drafting and issuing bimonthly orders. For instance, the expert master could be asked to design and oversee a machine learning algorithm to centralize relevant AI training data and minimize unnecessary production.[217] To aid the court with the technical aspects of discovery motions, the expert master could make recommendations based on their knowledge of the complex subject matter. As a neutral party, the expert master could independently survey the accuracy of proprietary data disclosures to prevent undue exposure of trade secrets, a major concern for many AI developers. Moreover, the mere presence of a non-adversarial, technically-fluent moderator may be sufficient to induce "a great tranquilizing effect" on discovery disputes.[218]

In addition to overseeing e-discovery, expert masters can advise district courts on the technical aspects of critical certification questions relating to predominance and ascertainability under Rule 23(b)(3). Of course, courts have no obligation to heed the advice of these masters when deciding the ultimate questions of law.[219] Still, it is increasingly evident that, through no fault of their own, courts are unequipped to keep up with the "black box problem" of deep learning systems and the rapidly evolving nature of Gen-AI technology.[220] This is exactly the situation that the landmark Rule 53 revisions were designed to remediate. The 2003 advisory committee expressly recognized the existence of doctrinal areas where the "court's responsibility to interpret [claims] as a matter of law … may be greatly assisted by appointing a master who has expert knowledge of the field in which the [claim] operates."[221] In its notes, the advisory committee pointed to patent litigation as a

---

[216] *See, e.g.*, Authors Guild v. Open AI, Inc., No. 23-cv-8292, 2024 WL 5047445 (S.D.N.Y. Dec. 6, 2024) (order resolving no fewer than thirteen discovery motions).

[217] *See generally* Salib, *supra* note 79 (advocating for use of machine learning algorithms in class actions).

[218] Manual for Complex Litigation, Fourth, § 11.51 (quoting E. Barrett Prettyman, *Proceedings of the Seminar on Protracted Cases for United States Circuit and District Judges*, 21 F.R.D. 395, 469 (1957)).

[219] *See* Fed. R. Civ. P. 53 advisory committee's notes to 2003 amendment (noting that while review of an expert master's findings "will be de novo under Rule 53(g)(4) … the advantages of initial determination by a master may make the process more effective and timely than disposition by the judge acting alone").

[220] *See generally* Samir Rawashdeh, *AI's mysterious 'black box' problem, explained*, U. Mich.-Dearborn News (Mar. 6, 2023), https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained [https://perma.cc/7QX2-ZWWT].

[221] Fed. R. Civ. P. 53 advisory committee's notes to 2003 amendment.

prototypical example of such a field.[222] Comparable to patent litigation, modern AI copyright class actions raise rigorous, data-driven questions that require highly specialized education and experience to resolve.[223] By leveraging the expertise of appointed expert masters, district courts can work toward a more efficient and well-balanced resolution to the complex questions around Rule 23(b)(3) certification in AI copyright class actions.

As with any procedural gamble, potential downsides exist. A prudent judge might balk at the potential expense of long-term expert masters, particularly in cases where one party is unable to shoulder the financial burden.[224] But when faced with severe funding disparities, district courts retain broad cost shifting authority under Rule 53(g)(3) "to order the nonindigent party to pay … in compelling circumstances," such as when "the indigent party's claim has merit that cannot viably be presented absent such expert assistance."[225]

AI copyright class actions implicate "compelling circumstances" that warrant the exercise of discretion under Rule 53(g)(3).[226] For one, the merits of AI infringement claims are typically impossible to show without access to, and understanding of, intricate Gen-AI algorithms, for which an expert may be necessary. Further, many plaintiffs in these actions are small artists or content creators who cannot sustain the economic impact of an expert master appointment. In *Andersen*, for instance, named plaintiffs are a small group of everyday artists from around the world, representing a huge class of other everyday artists, while defendants are large tech companies with vast resources and complex proprietary Gen-AI models.[227] This resource asymmetry provides a compelling justification for cost shifting under Rule 53(g)(3). Because AI defendants often profit from

---

[222] *Id.*

[223] *See generally* Julia Szatar, *How to Become an AI Expert: Guide & Career Paths [2025]*, Tavus (Nov. 27, 2024), https://www.tavus.io/post/how-to-become-an-ai-expert [https://perma.cc/5VXE-CATJ]. *Cf.* Salim Mamajiwalla, *A Career in Patent Law: At the Cutting Edge of Science, but Not at the Bench*, Nat. Lib. Med. (2018), https://pmc.ncbi.nlm.nih.gov/articles/PMC6028075/ [https://perma.cc/56D5-6NCN].

[224] Manual for Complex Litigation, Fourth, § 11.51.

[225] *Id.*; *see generally* Fed. R. Civ. P. 53(g)(3).

[226] Vincent, *supra* note 7 (noting that the people most affected by AI technology lack resources to launch legal challenges); *cf.* Madigan, *supra* note 5 (describing suits filed by "corporate copyright owners" who do not need or warrant judicial assistance under Rule 53(g)(3)).

[227] Second Amended Complaint at 5-6, Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal. Oct. 31, 2024) (outlining putative plaintiffs' humble and diverse backgrounds).

the fruits of Gen-AI to the tune of billions of dollars while hiding behind veiled technology, it only seems fair for them to shoulder the cost of expert master appointments, so long as the court deems it reasonable. For these defendant companies, Rule 53 expert masters would probably end up being less costly than a single junior litigation associate's billing rate. And if a putative class of plaintiffs is found to be acting in bad faith or mounting a frivolous suit, courts have discretion to dispense justice accordingly.

There may also be concerns that the appointment of expert masters will delay the already excruciatingly slow judicial process.[228] This does not have to be the case. Courts can be mindful of delays by restricting the expert master's role to a specific branch of highly technical inquiries, or by retaining the expert master for a limited and predetermined duration at important procedural junctures (*e.g.*, certification order). And when used effectively, expert masters should help alleviate the court's congested dockets by reducing the amount of time that a judge and his clerks would normally spend researching complex issues of science and technology.

Furthermore, the "appointment of a master [is] the exception and not the rule."[229] Not every AI copyright class action merits the long-term appointment of an expert master—and even if they did, AI copyright litigation is such a vanishingly small subset of the overall class action regime that there would be minimal impact on court dockets and pockets. In fact, any negative impact would be counterbalanced by the expert master's facilitation of technical efficiencies.

Finally, some judges may worry that assigning expert masters a large role in evaluating substantive pretrial issues will infringe upon the litigants' due process rights, especially if the expert masters' findings materially influence key rulings on discovery and certification. Not so. Rule 53(f) contains built-in procedural safeguards to ensure due process. For example, Rule 53(f)(1) mandates that the court "give the parties notice and an opportunity to be heard" whenever it acts on a

---

[228] *See Securities Class Action Settlements: 2023 Review and Analysis*, Cornerstone Research 14 (2023), https://www.cornerstone.com/wp-content/uploads/2024/03/Securities-Class-Action-Settlements-2023-Review-and-Analysis.pdf [https://perma.cc/V5RC-C5DC] (finding that in securities class actions, the median time from filing to settlement hearing date was 3.7 years).

[229] Fed. R. Civ. P. 53 advisory committee's notes to 2003 amendment.

master's order, report, or recommendation.[230] This gives the parties ample time to file motions to adopt, modify, or object to the master's suggestions.[231] And when a party raises an objection to the expert master's factual findings or legal conclusions, the court is generally obligated to conduct a rigorous *de novo* review.[232] The Rule 53(f) safeguards easily satisfy the "fundamental requirement[s]" of procedural due process under *Mullane*.[233]

Although there exist legitimate concerns regarding the usage of Rule 53 expert masters in AI copyright class actions, many of these concerns are mitigated by supportive precedent, practical benefits, and procedural safeguards. Most importantly, district courts retain discretion to tailor efficient appointments and mitigate financial impact on litigants under the unambiguous guidelines of Rule 53. As technology continues to evolve at an extraordinary pace, we predict that expert masters will become essential players in complex technical disputes to ensure just outcomes and judicial manageability.

Conclusion

The advent of generative artificial intelligence raises unprecedented questions at the intersection of class action procedure, intellectual authorship, and technological innovation. This challenge is exacerbated by the purposely "black box" nature of Gen-AI training mechanisms that functions as a legal smoke screen for many AI developers.[234]

As-is, the Rule 23 class action framework fails to resolve the procedural asymmetries inherent in AI copyright actions. Putative class members struggle to meet key Rule 23(b)(3) requirements like predominance and ascertainability at the outset; their highly individualized inquiries of causation and injury bar certification as an actual damages class after the Court's unfavorable rulings in *Wal-Mart* and *Amchem*. Compounding this problem is the inadequacy of statutory

---

[230] Fed. R. Civ. P. 53(f)(1).

[231] Fed. R. Civ. P. 53(f)(2) (giving parties at least 21 days to respond).

[232] *See* Fed. R. Civ. P. 53(f)(3) (mandating *de novo* review or stipulation to other terms); Fed. R. Civ. P. 53(f)(4).

[233] *See* Mullane v. Cent. Hanover Bank & Trust Co., 339 U.S. 306, 314 (1950) (requiring constitutional minimum of "notice reasonably calculated, under all the circumstances, to apprise interested parties of the pendency of the action and afford them an opportunity to present their objections").

[234] *See generally* Rawashdeh, *supra* note 220.

damages as a feasible alternative for the majority of proposed plaintiffs in these lawsuits. The majority of high-volume and/or low-income creators are precluded from the statutory remedy as a consequence of the Copyright Office's inequitable financial and knowledge barriers to access.[235] Finally, the uncertain state of Article III standing in class action litigation after *TransUnion* may also prove an insurmountable barrier at the very outset.

Beyond legal challenges, many district courts are forced to contemplate important policy considerations when making decisions in AI copyright infringement cases. Intellectual property regimes hinge on a tenuous balance between the incentivization of technological innovation and the protection of artistic creativity.[236] The rise of AI throws this balance in flux by raising uncomfortable questions about data ownership, authorship attribution, and the scope of legal protections. Copyright law operates under the presumption that *human* authors create original and expressive works.[237] AI-generated content disrupts this foundational premise. The current judicial consensus posits that creative works generated by or with AI are not protectable under copyright—but from many utilitarian and economic-minded perspectives, such treatment disincentivizes technological innovation.[238] Furthermore, some commentators even contend that from a consumer perspective, AI-generated works have the same intrinsic artistic value as human-created works and therefore should not be categorically devalued by copyright law.[239]

---

[235] *See generally* Part II-C, *supra* (discussing knowledge and financial barriers posed by the registration requirement).

[236] *See, e.g.*, Sarah Hess, *The Evolving Landscape of IP Law in the Age of AI*, ABC Legal (Aug. 19, 2024), https://www.abclegal.com/blog/the-evolving-landscape-of-ip-law-in-the-age-of-ai [https://perma.cc/6WCZ-E42G].

[237] *See* Thaler v. Perlmutter, 687 F. Supp. 3d 140, 146 (D.D.C. 2023) (holding that "[h]uman authorship is a bedrock requirement of copyright").

[238] *Cf.* Patrick Goold, *The dubious utilitarian argument for granting copyright in AI-generated works*, Kluwer Copyright Blog (Jan. 9, 2025), https://copyrightblog.kluweriplaw.com/2025/01/09/the-dubious-utilitarian-argument-for-granting-copyright-in-ai-generated-works/ [https://perma.cc/27VR-YMEM] (discussing two proposed utilitarian benefits, but concluding that these benefits probably don't outweigh the costs).

[239] *See* Benjamin Hardman & James Housel, *A Sui Generis Approach to the Protection of AI-Generated Works: Balancing Innovation and Authorship* (arguing that copyright law undervalues human inputs into AI-generated work).

But common sense dictates that judges should not be pushed to legislate from the bench. The growing legal and policy implications within the Gen-AI copyright space highlight an urgent need for legislative action. Through no fault of its own, the dated Copyright Act of 1976 is unprepared to address the starkly modern realities of AI creation. A proactive legislative response to AI must strike a delicate balance between innovation and protection—for example, a *sui generis* statutory framework that simultaneously "value[s] human input and initiative when utilizing AI to generate works [and] preserves the traditional protections reserved for human creators."[240] This framework would be constructed with explicit safeguards such as a limited term of protection, well-defined legal rights, strict registration and notice requirements, and accessible statutory causes of action for layman creators. Such safeguards would acknowledge AI's potential as a legitimate means of creation while continuing to defend and enforce the paramount rights of human creators. Most imperatively, a thorough procedural framework is needed to provide unambiguous statutory guidance for the judiciary.

When testing the uncharted waters of civil procedure, district courts judges are captains of their own fate. Rule of law is theirs to navigate within the expansive borders of common law discretion. Although this discretion may be flaunted in a number of experimental ways, as a matter of practicality and fairness—and in the practical interest of not getting overruled—district courts should consider implementing the procedural mechanisms that are proven to be effective in complex litigation.

Specifically, judges should certify—on motion and *sua sponte* in the interest of justice—issue classes under Rule 23(c)(4)(A) to pinpoint common questions of liability at the outset and inform the direction of future proceedings. If technical complexities persist after the resolution of classwide liabilities, judges should appoint expert masters under Rule 53 for their tailored subject-matter expertise, both during e-discovery and pre-certification more broadly. As a matter of this author's personal opinion, these expert masters should be presumptively financed by the corporate defendant(s) because their lack of technical transparency is what necessitates such measures in the first instance.[241] Finally, this Note urges district

---

[240] *Id.* (discussing proposed solution to the AI copyright conundrum).

[241] There may be exceptions, but the court always has full discretion to apportion costs of expert masters as needed under Rule 53(g)(3). *See* FED. R. CIV. P. 53(g)(3); Part III-B, *supra*.

judges to *liberally* enforce these mechanisms in the context of AI copyright class action litigation, to a broader extent than is typically done in established doctrinal areas. Affirmative action is necessary to counterbalance the severe informational asymmetry inherent in this nascent era of AI copyright class actions.

By strategically leveraging Rules 23(c)(4)(A) and 53, district judges can transform procedural discretion in the AI copyright space from an exercise in trial and error, into a disciplined and fair approach that compels technical transparency, streamlines litigation, and sidesteps the many inequitable barriers to justice. In an era of rapid technological advancement and stark economic inequalities, we must protect "[t]he policy at the very core of the class action mechanism": to provide everyday individuals with the ability to, as a collective, achieve justice that would not have been possible as a "solo action prosecuting his or her rights."[242]

---

[242] *Amchem*, 521 U.S. at 617.