FROM TRAINING DATA TO MACHINE MEMORY: RECONSTRUCTING COPYRIGHT LIABILITY FOR GENERATIVE AI

Gregory Koshy Thomas, Advocate & Independent Legal Researcher

ABSTRACT

The emergence of large language models and generative AI systems has fundamentally challenged traditional copyright frameworks by conflating training data ingestion with derivative similarity, machine memorization, and market harm in unprecedented ways. This paper reconstructs copyright liability for generative AI by examining how machine learning models encode and reproduce protected works through their parameters, creating a persistent form of "machine memory" that extends beyond literal reproduction to encompass stylistic appropriation and substantial similarity in expressive elements. Through comparative legal analysis of US, EU, and emerging market frameworks, we demonstrate that current copyright doctrine inadequately addresses three critical gaps: (1) the status of training data reproduction as a reproduction under copyright law; (2) the legal significance of model memorization as a distinct form of infringement; and (3) the allocation of liability among data scrapers, model developers, and end-users across the generative AI value chain. This paper proposes a reconstructed liability framework grounded in transparency obligations, tiered compensation mechanisms, and substantive similarity assessment methodologies that better align technical reality with legal principle. We further argue for extended collective licensing systems and mandatory attribution protocols to establish accountability without impeding technological innovation.

Keywords: generative artificial intelligence, copyright liability, machine memorization, training data, fair use doctrine, substantial similarity, text and data mining exceptions.

1. Introduction

1.1 Background: The Transformation of Copyright in the Age of Generative AI

The rapid proliferation of generative artificial intelligence (GenAI) technologies has precipitated a fundamental reconceptualization of copyright law's foundational assumptions. Traditional copyright frameworks were constructed around the notion of human authorship and tangible copying concepts that become unstable and multivalent when applied to machine learning systems. Generative AI systems ingest vast corpora of data, often comprising billions of copyrighted works spanning literature, music, visual art, and code, transforming these materials into distributed parameters within neural networks. This process creates what we term "machine memory": the encoding of training data patterns and stylistic markers into model weights, which can be accessed, extracted, and reproduced in ways that do not neatly conform to either traditional infringement categories or established fair use doctrines.

The technical reality of machine learning introduces legal complexity that existing copyright regimes were not designed to accommodate. When a large language model (LLM) is trained on a corpus containing copyrighted works, the model does not store those works in a discrete, retrievable format analogous to a hard drive or printed archive. Instead, training represents a lossy compression and pattern extraction process wherein the model learns statistical relationships and linguistic structures that characterize the training data.[1] Yet emerging evidence demonstrates that LLMs exhibit substantial memorization of their training data the ability to reproduce, with surprising fidelity, sequences or passages present in the original corpus. This memorization appears qualitatively different from traditional reproduction in that it is embedded in a complex, high-dimensional parameter space rather than stored explicitly; yet the legal significance of this difference remains contested.

The emergence of legal challenges including cases such as New York Times v. OpenAI, Getty Images v. Stability AI, and Bartz v. Anthropic has placed copyright liability at the intersection of technological reality and legal doctrine. These disputes have exposed fundamental misalignments between how copyright law conceptualizes reproduction, transformation, and fair use, and how machine learning systems actually encode, store, and generate content. The

[1] Directive 2019/790/EU of the European Parliament and Council of 17 April 2019, on Copyright and Related Rights in the Digital Single Market, 2019 O.J. (L 130) 92.

stakes are not merely doctrinal; they concern the future allocation of value, incentives for creation, and the distribution of risk across the technological ecosystem.

1.2 Research Problem and Objectives

The core research problem driving this paper is deceptively simple yet jurisprudentially complex: How should copyright law allocate liability for the unauthorized use of protected works in training generative AI systems when that use involves neither explicit copying nor direct market substitution in the traditional sense? More specifically, we address three interlocking questions:

(1) The Reproduction Question: Does the incorporation of copyrighted works into machine learning training datasets constitute reproduction under copyright law, and if so, how does this reproduction differ from and relate to traditional copying?

(2) The Machine Memory Question: What is the legal significance of model memorization the ability of trained models to extract and reproduce sequences from their training data for copyright liability, and how does memorization differ from stylistic appropriation or substantial similarity in expressive elements?

(3) The Liability Allocation Question: Across the generative AI value chain from data scraping and corpus assembly, through model training and release, to end-user deployment and output generation how should liability be distributed among the multiple actors, and what role should transparency, attribution, and compensation mechanisms play in reconstructing a coherent copyright framework for this technology?[2]

The overarching objective of this paper is to reconstruct copyright liability for generative AI by synthesizing technical understanding of machine learning processes, comparative legal analysis of existing copyright frameworks across multiple jurisdictions, and normative considerations about how copyright law might better balance the interests of creators, technology developers, and the broader public in the age of machine-generated content.

[2] Kneschke v. LAION e.V., 310 O 227/23 (Landgericht Hamburg Sept. 27, 2024)

2. Literature Review and Legal Context

2.1 Traditional Copyright Framework and Its Inadequacies

Copyright law in its contemporary form, codified in works such as the US Copyright Act (17 U.S.C. § 101 et seq.) and the European Directive on Copyright in the Digital Single Market (Directive 2019/790/EU), rests on several foundational principles. These include: (1) exclusive rights granted to human authors to reproduce, distribute, and create derivative works; (2) originality as a threshold requirement for copyright protection; (3) fair use and fair dealing exceptions that permit limited uses of copyrighted works without authorization; and (4) the principle of substantial similarity, which recognizes that infringement need not involve literal copying but may extend to non-verbatim appropriation of protectable expression.[3] These principles worked reasonably well in an era of discrete, human-centered creative acts and tangible copying mechanisms. However, they become strained when applied to machine learning processes that operate at unprecedented scale, with mechanisms of encoding and reproduction that are distributive, probabilistic, and non-transparent.

The challenge is not merely one of doctrinal extension. Rather, machine learning exposes deep theoretical gaps within copyright law itself. For instance, the concept of "reproduction" presupposes discrete boundaries between an original work and a reproduction boundaries that become blurred when a copyrighted work is decomposed into statistical patterns, embedded in a multi-dimensional parameter space alongside billions of other works, and then partially reconstructed through probabilistic sampling and inference. Similarly, the fair use doctrine, developed primarily in the context of human-centered secondary uses (criticism, parody, news reporting, scholarship), struggles to accommodate the functional role of training data ingestion in machine learning, where the purpose of reproduction is neither to analyze nor to criticize, but to extract statistical patterns for model learning.[4]

2.2 Copyright Doctrine in the Generative AI Context: Global Perspectives

2.2.1 United States: Fair Use and Transformative Use

In the United States, legal debate over AI copyright liability has largely centered on the doctrine

[3] Getty Images (US) Inc. v. Stability AI Ltd., No. 1:23-cv-00521 (N.D. Cal. filed Feb. 2023)
[4] The New York Times Co. v. OpenAI, Inc., No. 1:23-cv-11195 (S.D.N.Y. filed Dec. 27, 2023)

of fair use (17 U.S.C. § 107), which permits limited uses of copyrighted material without authorization where four statutory factors favor the secondary user: (1) the purpose and character of the use; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used; and (4) the effect of the use on the potential market. The "transformative use" doctrine developed primarily through cases such as Campbell v. Acuff-Rose Music, Inc. and Andy Warhol Foundation v. Goldsmith has become central to fair use analysis, with courts asking whether a secondary use adds new meaning, expression, or message to the original work, thereby transforming it and justifying reproduction without permission.[5]

Scholarly and judicial opinion regarding the application of fair use to AI training remains sharply divided. Some scholars and legal practitioners contend that machine learning qualifies as transformative use because the training process extracts statistical patterns rather than reproducing the expressive form of works, and because the resulting models generate novel outputs that do not directly substitute for the originals. Under this view, AI training serves a functional, non-expressive purpose analogous to reverse engineering or format conversion, which courts have sometimes treated favorably under fair use analysis. Others argue that transformative use doctrine, properly understood, is ill-suited to AI training because the doctrine was designed to address secondary human uses that add independent creative expression, whereas machine learning operates at an abstraction level where transformation is technical and instrumental rather than creative or communicative.[6]

The July 2025 decision in Bartz v. Anthropic marked a significant shift. The federal court recognized fair use as a potential defense for AI training on copyrighted works, reasoning that LLMs extract linguistic and statistical patterns rather than storing entire works, and that the market for training data cannot be confined exclusively to copyright holders. However, this decision has not settled the law, as earlier cases, including Authors Guild v. Google and other ongoing disputes have reached different conclusions or adopted more skeptical stances toward fair use in the AI context.

2.2.2 European Union: Text and Data Mining Exceptions and Substantial Similarity

The European Union has taken a different regulatory approach, codifying exceptions for text

[5] Ippolito, Daphne et al., Three Years of Generative AI: Ten Lessons for the Road Ahead, arXiv:2402.05462 [cs.AI] (2024).
[6] GEMA v. OpenAI, Case No. 1 HK O 195/24 (Landgericht München I Nov. 28, 2024).

and data mining (TDM) in Articles 3 and 4 of Directive 2019/790/EU. These exceptions permit reproduction of lawfully accessible works for computational analysis, subject to opt-out rights exercised "in an appropriate manner" by copyright holders. Importantly, the EU framework operates through defined exceptions rather than the flexible fair use doctrine, providing greater ex-ante clarity but less flexibility for novel uses.

A critical distinction in EU law is the emphasis on substantial similarity in protectable expressive elements plot structure, character development, narrative voice, the selection and arrangement of creative incidents rather than verbatim copying alone. Recent scholarship and legal developments, particularly the application of EU copyright standards to generative AI, emphasize that infringement may turn on stylistic appropriation and non-literal reproduction of expressive choices. This creates a potential enforcement gap: current technical safeguards in AI systems focus narrowly on detecting and suppressing verbatim memorization (through deduplication, similarity thresholds, and unlearning methods), but leave largely unaddressed the risk that fine-tuned or instruction-aligned models may systematically appropriate protected stylistic and narrative patterns from their training data.

2.2.3 China, India, and Emerging Regulatory Frameworks

China's regulatory approach remains evolving, with ongoing debates about whether AI training should qualify as fair use under Article 24 of China's Copyright Law.[7] Some scholars have argued that machine learning training should be recognized as fair use because it operates at an abstraction level where reproduction is incidental and transformative. India, by contrast, has adopted a more human-centered framework, proposing that copyright protection attach only to human creative contributions in AI-generated works, with liability extended only where AI outputs are substantially similar to protected works and likely to cause market harm. These jurisdictional differences underscore the absence of global consensus on how copyright liability should operate in the AI context.

2.3 Machine Memorization and Training Data Extraction

Recent empirical research has fundamentally altered the technical understanding of how LLMs interact with training data. Carlini et al. (2021) first demonstrated that language models exhibit

[7] Cooper, Albert F. et al., Extracting Memorized Pieces of (Copyrighted) Books from Open-Weight Language Models, arXiv:2505.12546 [cs.CL] (2025).

substantial memorization of their training data the ability to extract, through careful prompting or inference manipulation, sequences and passages present in the original corpus. Subsequent work has documented that this memorization appears largely unavoidable: it is not an artifact of careless training practices but rather an intrinsic feature of how neural networks compress high-dimensional data into learned representations.

Critically, memorization in LLMs differs from explicit storage. The model does not maintain a retrievable index of training documents; rather, memorized information is distributed and encoded across parameter values in a way that can be accessed through inference but is not easily localized or removed.[8] This technical feature creates a mismatch with traditional copyright doctrine: the model has neither explicitly copied nor stored the work in a conventional sense, yet the work's presence is functionally recoverable from the model's learned representations. This raises the question of whether memorization constitutes reproduction under copyright law a question that EU courts, particularly in cases such as Kneschke v. LAION and GEMA v. OpenAI, have begun to answer affirmatively, reasoning that encoding of protected works in model parameters constitutes reproduction regardless of whether the representation is explicit or distributed.

2.4 Substantial Similarity, Stylistic Appropriation, and Copyright Infringement

European copyright doctrine distinguishes sharply between surface-form overlap and substantial similarity in protectable expression. Infringement may occur where a secondary work exhibits substantial similarity in structure, narrative voice, character dynamics, or the selection and arrangement of creative elements, even without verbatim correspondence. This principle, established in foundational cases before the Court of Justice of the European Union, reflects a conception of copyright as protecting the author's creative choices and expressive personality rather than merely literal text.

Applied to generative AI, this doctrine suggests that a fine-tuned LLM that systematically appropriates the narrative voice, thematic concerns, or stylistic markers of protected works may constitute infringement even if the model does not reproduce specific passages verbatim. Recent scholarship utilizing automated evaluation frameworks (such as PSALM: Probing Stylistic Appropriation by Language Models) has demonstrated that instruction-tuned models

[8] CJEU, Infopaq International A/S v. Danske Dagblades Forening, Case C-5/08, 2009 ECR I-6569 (2009).

exhibit statistically significant stylistic similarity to works in their training corpus, and that fine-tuning on targeted literary corpora intensifies this similarity across multiple dimensions of protectable expression. These findings indicate that current technical safeguards focused on literal memorization detection leave a substantial compliance gap for stylistic and structural infringement.

3. Methodology and Analytical Framework

3.1 Methodological Approach

This paper employs a mixed-methods approach combining legal doctrine analysis, technical analysis of machine learning processes, and comparative legal assessment across jurisdictions. The doctrinal component involves systematic examination of copyright statutes, case law, and scholarly commentary across the United States, European Union, and emerging markets to identify core principles and articulate their application to novel AI-generated scenarios.[9] The technical component draws on peer-reviewed literature in machine learning and artificial intelligence to clarify how training data is processed, encoded, memorized, and extracted from language models. The comparative legal component contextualizes jurisdictional differences and identifies areas of potential convergence and intractable divergence.

3.2 Analytical Framework: The Generative AI Value Chain

We structure our analysis around three critical phases in the generative AI value chain:

Phase 1: The Training Phase encompassing data collection, corpus assembly, pre-processing, and model training. At this phase, the primary legal question concerns whether incorporation of copyrighted works into training corpora constitutes unauthorized reproduction and whether fair use or TDM exceptions apply.

Phase 2: The Model Storage and Distribution Phase involving the release and deployment of trained models. Here the critical questions concern the legal status of memorized content within model parameters, the liability of model distributors, and the role of unlearning and other post-hoc remediation techniques.

[9] Bartz v. Anthropic, PBC, No. 3:23-cv-04516 (N.D. Cal. July 2025) (unpublished opinion).

Phase 3: The Generation and Output Phase where end-users deploy the model to generate new content. At this phase, the relevant question is whether outputs generated by models trained on copyrighted data may infringe, and whether and how the original training data usage affects output liability.

Within this framework, we examine how liability might be allocated among multiple actors: data scrapers and corpus assemblers, model developers and trainers, model distributors and deployers, and end-users generating content through the model. We further assess how transparency obligations, attribution protocols, and compensation mechanisms might operate to clarify responsibilities and align incentives.

4. Discussion and Analysis

4.1 Reframing Reproduction in the Machine Learning Context

A fundamental reconstruction of copyright liability must begin by clarifying what reproduction means in the context of machine learning. Traditional copyright doctrine treats reproduction as copying creating a new instance of the original work, whether through mechanical duplication (photocopying), technological reformatting (digitization), or human transcription. The reproduction right is meant to prevent unauthorized creation of competing copies that substitute for the original in the marketplace and diminish its economic value.

In machine learning, the process of incorporating copyrighted works into training data involves several distinct acts: (1) retrieval of the work; (2) parsing and pre-processing to extract structured information; (3) decomposition into tokens or vector embedding's; (4) incorporation into a training corpus; and (5) gradient-based learning wherein model parameters are adjusted to minimize loss across the corpus. Viewed holistically, this process does constitute reproduction in a meaningful sense: the work's content, structure, and stylistic markers are encoded into the model's parameter space. The fact that this encoding is lossy, distributed, and not explicitly indexed does not diminish the reality that the work has been reproduced and internalized within the model.[10]

However, reproduction in the machine learning context differs substantively from traditional copying in its functional purpose and mechanism. The purpose is not to create a substitute copy

[10] Harper & Row v. Nation Enterprises, 471 U.S. 539 (1985)

competing with the original, but to extract information and patterns that will enable the model to generalize across diverse tasks and domains. The mechanism is not a faithful transcription but a statistical compression and abstraction process. Recognizing these differences suggests that the appropriate legal framework may not be one that simply extends traditional reproduction doctrine, but rather one that develops intermediate categories capturing the distinctive features of machine learning.

We propose that copyright law distinguish between: (1) Explicit Reproduction, wherein a work is intentionally copied in full or substantial part with the purpose of creating a competing instance; (2) Distributed Encoding, wherein a work is encoded as patterns within model parameters through machine learning, serving a functional rather than substitutive purpose; and (3) Extractive Reproduction, wherein encoded information is retrieved and reproduced in outputted text, potentially serving a substitutive function. This taxonomy allows for differentiated liability: explicit reproduction and extractive reproduction would constitute presumptive infringement unless justified by fair use or statutory exception; distributed encoding would fall under fair use or TDM exception analysis, with the outcome depending on whether the encoding occurs with authorization, transforms the work functionally, or harms the original market.

4.2 Memorization and the Legal Status of Machine Memory

Machine memorization poses a distinctive challenge because it operates as an intermediate category between distributed encoding and explicit reproduction. When a model memorizes and can reproduce sequences from its training data, it has not merely extracted patterns but has preserved and internalized specific textual content that can be retrieved through inference. Yet the preservation is neither intentional nor transparent; it is an artifact of the learning process rather than an explicit design choice.

We argue that memorization should be recognized as a distinct form of reproduction with its own legal implications. Specifically: (1) memorization represents unauthorized reproduction of the memorized content within the model parameters; (2) the ability to extract memorized content upon demand constitutes a latent infringement liability that accrues regardless of whether extraction is actually performed; and (3) developers who knowingly train models on protected works and retain knowledge of potential memorization bear responsibility for either obtaining authorization, ensuring that no memorization occurs (through techniques like

differential privacy), or implementing robust unlearning mechanisms to address memorized content post-hoc.

The distinction between memorization and pattern extraction is critical. A model may accurately internalize the statistical properties of text written by Shakespeare or descriptions of Parisian architecture without being able to reproduce specific passages or scenes from those works. This internalization is functionally closer to human learning absorbing stylistic influences and structural patterns and may be defensible as transformative and non-infringing. By contrast, memorization of specific passages constitutes preservation of the original expression and creates a latent capacity for substitutive reproduction. Legal accountability should track this distinction.

4.3 Stylistic Appropriation and Substantial Similarity in the EU Framework

A critical insight from European copyright doctrine is that infringement need not involve literal copying; substantial similarity in protectable expressive elements narrative structure, character development, stylistic voice may constitute infringement independently. Applied to generative AI, this principle suggests that models fine-tuned on corpora of protected works may develop systematic stylistic and structural similarities to those works that, when expressed in generated outputs, constitute infringement.

Recent empirical evidence using evaluation frameworks operationalizing EU copyright doctrine (such as PSALM) demonstrates that fine-tuned LLMs exhibit statistically significant increases in stylistic similarity to works in their fine-tuning corpus across dimensions including narrative voice, character characterization, plot architecture, and thematic structure. These similarities persist even after machine unlearning and appear to reflect abstract pattern capture rather than verbatim memorization.

We propose that developers implementing machine learning systems in the EU context face a compliance obligation to assess and mitigate substantial similarity risks through: (1) Stylistic Auditing: periodic evaluation of generated outputs using standardized evaluation frameworks that operationalize substantial similarity tests across legally protectable dimensions; (2) Training Data Transparency: disclosure of copyrighted works included in training corpora, enabling rightholders to identify potential infringement risks; and (3) Capability Restriction: design choices that limit model capacity to reproduce protected stylistic patterns, through

approaches such as differential privacy, architecture modifications, or careful corpus curation.

4.4 Liability Allocation Across the Value Chain

Reconstructing copyright liability requires clarifying which actors bear responsibility at different phases of the generative AI lifecycle:

Data Scrapers and Corpus Assemblers bear primary responsibility for identifying and obtaining authorization for copyrighted materials incorporated into training corpora. If copyrighted works are incorporated without authorization, liability attaches to the party controlling the corpus assembly process, regardless of whether the corpus is subsequently used for training.

Model Developers and Trainers bear secondary responsibility for exercising due diligence in understanding the composition of their training data and implementing technical safeguards to prevent and mitigate infringement. Where they knowingly train on protected works without authorization and implement no safeguards against memorization or stylistic appropriation, they incur liability for the resulting model.

Model Distributors who release models into commerce bear responsibility for disclosing training data sources, known memorization risks, and the results of stylistic similarity audits. This transparency obligation is critical for enabling downstream accountability.

End-Users generating outputs through models bear responsibility for ensuring that their uses comply with applicable copyright law and fair use doctrine. Where a generated output is substantially similar to protected works and likely to substitute for them in the marketplace, the end-user's use may constitute infringement, with liability potentially traceable back through the value chain to the developers who created the model.

This tiered liability structure recognizes that different actors have different capacities for control, foresight, and remediation at different stages of the process. It also preserves incentives for transparency and good-faith compliance efforts.

4.5 Fair Use and Fair Dealing in the AI Training Context

Debates over whether AI training qualifies as fair use or fair dealing will likely persist for years; however, we can identify key principles that should guide interpretation:

Transformative Purpose: Machine learning arguably qualifies as transformative in the technical sense it transforms copyrighted works into statistical patterns and model parameters that serve a distinct functional purpose from the original works. However, transformative purpose is not equivalent to transformative use as traditionally conceived. Courts should recognize that functional transformation (extracting information for machine learning) differs from expressive transformation (creating secondary creative works), and calibrate fair use analysis accordingly.

Market Harm: The most significant question concerns market harm the fourth fair use factor. Where models trained on copyrighted works are deployed to generate content in direct competition with the original works (e.g., LLMs trained on novels generating novel-like fiction), or where the training process directly substitutes for licensing and diminishes the original market, fair use analysis should weigh heavily against the secondary user. However, where the model's outputs serve distinct purposes or markets, market harm is less clear.

Authorization and Market Feasibility: A critical consideration is whether licensing markets exist and are feasible. If the copyright holder could reasonably offer licenses for training data use, and such licensing would not impede AI development significantly, fair use analysis should account for the copyright holder's right to control derivative uses and receive compensation.

We propose that fair use analysis in the AI context should focus on the commerciality of the use, the directness of market substitution, and the availability of licensing mechanisms. Uses that are non-commercial (pure research), do not directly substitute for original works in any market, and occur where licensing mechanisms are infeasible, should receive strong fair use protection. Conversely, uses that are commercially exploitative, create direct market substitution, and occur where licensing could occur, should face significant fair use obstacles.

4.6 Text and Data Mining Exceptions and Opt-Out Mechanisms

The EU's TDM exception provides a model for balancing innovation and creator rights. However, implementation challenges remain substantial. A critical ambiguity concerns "lawful access" the requirement that works be lawfully accessible to qualify for the exception. When works are scraped from the internet without explicit permission, accessed through unauthorized web crawling, or obtained from pirated collections such as Library Genesis or Books3, does

the scraper have "lawful access" in the requisite sense?

We propose that compliance with TDM exceptions should require: (1) Transparent Sourcing: clear documentation of where training data originated and confirmation of lawful access; (2) Functional Opt-Out: technical mechanisms (such as robots.txt protocols, machine-readable content negotiation headers, or registry-based opt-out systems) that enable rightholders to exclude their works from training corpora; and (3) Disclosure of Use: public disclosure of copyrighted works included in training data, enabling rightsholders to monitor unauthorized incorporation and pursue remedies where necessary.

Critically, the mere existence of opt-out mechanisms is insufficient if they are difficult to locate, implement, or enforce. Rightholders should not bear disproportionate burden to prevent unauthorized use; instead, scrapers and model developers should bear the burden of verifying lawful access and respecting clear exclusion signals.

4.7 Unlearning and Post-Hoc Remediation

Machine unlearning techniques for removing or suppressing specific knowledge from trained models has emerged as a potential mechanism for addressing memorization and copyright infringement. However, unlearning faces significant technical and legal limitations. Most unlearning techniques operate through approximate methods (gradient-based removal, self-distillation) that do not guarantee complete removal; they can leave residual memorization and stylistic patterns that preserve some of the original infringement risk.

From a legal perspective, unlearning is a remedial measure useful for demonstrating good-faith compliance efforts and reducing ongoing infringement risk. However, it should not be treated as a substitute for ex-ante authorization or as a defense to initial infringement. A model developer who trained on protected works without authorization, memorized them, and only subsequently applied unlearning techniques has still committed initial infringement. The unlearning effort may mitigate damages or demonstrate reduced risk, but does not retroactively cleanse the infringement.

We propose that unlearning should be recognized as: (1) a required element of ongoing due diligence for model developers; (2) evidence of compliance intent in determining damages calculations; and (3) a partial mitigation measure for stylistic appropriation risk. However,

unlearning should not provide complete harbor from liability or retroactively justify initial infringement.

5. Findings and Conclusion

5.1 Key Findings

This paper reaches several significant findings regarding copyright liability for generative AI:

Finding 1: Technical and Legal Misalignment: Current copyright doctrine operates through concepts reproduction, copying, market substitution that were developed for human-centered creative acts and tangible duplication mechanisms. Machine learning processes operate at fundamentally different scales and through mechanisms that do not conform neatly to these categories. This misalignment creates compliance uncertainty and enforcement gaps.

Finding 2: The Memorization Problem: Machine memorization represents a distinctive form of copyright infringement that operates as an intermediate category between distributed pattern encoding and explicit reproduction. Current technical safeguards focus narrowly on literal memorization detection but fail to address stylistic and structural appropriation, creating a significant compliance liability.

Finding 3: Substantial Similarity beyond Literal Copying: Applied to generative AI, European copyright doctrine's emphasis on substantial similarity in protectable expression reveals that fine-tuned models systematically appropriate stylistic and narrative patterns from training data, potentially constituting infringement even without literal reproduction. Current automated compliance mechanisms are inadequate to assess this dimension of infringement risk.

Finding 4: Liability Must Attach across the Value Chain: Responsibility for copyright compliance cannot rest solely on model developers or end-users; it must attach to data scrapers, corpus assemblers, model developers, distributors, and end-users, with tiered obligations reflecting their respective capacities for control and remediation at each phase.

Finding 5: Fair Use is Insufficient without Statutory Clarity: While fair use and fair dealing doctrines offer some protection for transformative AI training, their flexibility generates uncertainty that impedes investment and creates litigation risk. Statutory clarity through TDM exceptions, appropriately implemented, better serves both innovation and creator protection.

Finding 6: Transparency and Compensation Mechanisms Are Essential: Copyright liability for generative AI cannot be sustainably enforced through litigation alone. Rather, transparency obligations (requiring disclosure of training data sources, memorization assessments, and stylistic similarity audits) and compensation mechanisms (through licensing fees, royalty distributions, or extended collective licensing systems) are essential for aligning incentives and enabling fair allocation of value.

5.2 Proposed Legal Framework

Based on these findings, we propose a reconstructed copyright liability framework for generative AI comprising the following elements:

1. Tiered Classification of AI Training Activities: Distinguish between (a) non-commercial research use of lawfully accessible copyrighted materials where authorization cannot be feasibly obtained (low liability); (b) commercial AI training on copyrighted materials where licensing mechanisms exist but are economically burdensome (moderate liability, with fair use analysis turning on market harm); and (c) commercial AI training on protected materials without any authorization effort or safeguards (high liability, presumptive infringement).

2. Mandatory Transparency Obligations: Require model developers to disclose (a) significant sources of training data, including copyrighted works; (b) results of memorization assessments and stylistic similarity audits; (c) technical safeguards implemented to prevent or mitigate infringement; and (d) known residual risks. This transparency enables downstream accountability and allows rightholders to identify and pursue remedies.

3. Substantive Similarity Assessment Requirements: Impose obligations on developers deploying models in the EU context to conduct periodic audits assessing stylistic and structural similarity to protected works across legally protectable dimensions, using standardized evaluation methodologies. These assessments should inform model design choices and serve as evidence of due diligence.

4. Extended Collective Licensing System: Establish statutory or industry-based extended collective licensing (ECL) systems wherein a rights-collecting organization negotiates with AI developers for corpus-wide licensing agreements. These agreements would specify which copyrighted works may be incorporated, establish royalty rates, and distribute compensation to

rights holders. ECL systems address the practical impossibility of obtaining individual authorization from millions of worldwide rights holders while ensuring fair compensation.

5. Statutory Text and Data Mining Exception with Robust Opt-Out: Codify a TDM exception for non-commercial and transformative commercial research uses of lawfully accessible copyrighted works, subject to clear and effective opt-out mechanisms that enable rights holders to exclude their works from training corpora. Scraping and incorporation of works where opt-out signals are clear would constitute infringement.

6. Liability Allocation Rules: Establish clear rules allocating liability across the value chain: data scrapers bear responsibility for lawful access and opt-out compliance; model developers bear responsibility for due diligence regarding training data composition and implementation of technical safeguards; model distributors bear responsibility for transparency disclosures; and end-users bear responsibility for ensuring their outputs comply with applicable fair use and copyright doctrine.

7. Proportionate Remedies Framework: Establish a remedies framework distinguishing between (a) prospective injunctions restraining further infringement; (b) damages calculations accounting for developer fault, market harm, and demonstrated compliance efforts; (c) mandatory unlearning and model modification for models exhibiting high memorization or stylistic appropriation; and (d) potential registration and licensing of previously trained models to enable lawful continued deployment with ongoing royalty compensation.

5.3 Implications and Future Directions

This reconstructed framework has significant implications for AI governance, innovation incentives, and creator protection:

For Innovation: While imposing new compliance obligations, the framework provides greater clarity and certainty regarding the permissibility of AI training, thereby reducing litigation risk and enabling more confident investment in AI development. The incorporation of TDM exceptions and transparent licensing mechanisms would create stable pathways for lawful training.

For Creator Protection: By extending copyright liability across the entire value chain and establishing transparency and compensation mechanisms, the framework ensures that creators

retain meaningful control over their works and receive fair compensation when those works are incorporated into AI systems that generate commercial value.

For International Harmonization: The framework provides a foundation for international convergence by articulating principles (distributed encoding distinct from explicit reproduction, tiered liability based on actor role and capacity, transparency and compensation requirements) that can be adapted across different legal traditions.

Significant challenges remain: (1) Technical feasibility: whether effective opt-out and stylistic similarity assessment mechanisms can be implemented at scale remains uncertain; (2) Jurisdictional coordination: absent international agreement, jurisdictional arbitrage and enforcement challenges will persist; (3) Balancing innovation and protection: determining appropriate levels of restriction that preserve AI innovation while protecting creator interests will require ongoing calibration based on empirical outcomes.

5.4 Conclusion

The emergence of generative AI has precipitated a crisis in copyright law, exposing deep theoretical gaps and practical inadequacies in frameworks developed for an earlier technological era. This paper reconstructs copyright liability for generative AI by reframing core concepts reproduction, memorization, substantial similarity in light of how machine learning systems actually encode, store, and generate content. We argue that copyright liability must extend across the entire AI value chain, from data scrapers to end-users, with obligations appropriately calibrated to each actor's role and capacity for control and remediation.

The proposed framework combining tiered classification, transparency obligations, substantive similarity assessment, extended collective licensing, statutory TDM exceptions, clear liability allocation, and proportionate remedies offers a pathway for aligning copyright law with technological reality while preserving both innovation incentives and creator protection. Implementation will require sustained collaboration across legal, technical, policy, and industry communities. However, without such reconstruction, copyright law risks becoming either irrelevant (if narrowly interpreted to exclude AI training entirely) or perversely counterproductive (if interpreted so expansively as to impede legitimate innovation and adaptation).

The stakes are not merely doctrinal or economic, but concern fundamental questions about how creative value is recognized, allocated, and compensated in an age of machine-generated content. Reconstructing copyright liability for generative AI offers an opportunity to reimagine copyright's role in incentivizing creation, protecting creators, and enabling innovation in ways appropriate to the technological and social realities of the twenty-first century.

REFERENCES

1. CJEU, Infopaq International A/S v. Danske Dagblades Forening, Case C-5/08, 2009 ECR I-6569 (2009).
2. Kneschke v. LAION e.V., 310 O 227/23 (Landgericht Hamburg Sept. 27, 2024).
3. Getty Images (US) Inc. v. Stability AI Ltd., No. 1:23-cv-00521 (N.D. Cal. filed Feb. 2023).
4. The New York Times Co. v. OpenAI, Inc., No. 1:23-cv-11195 (S.D.N.Y. filed Dec. 27, 2023).
5. Ippolito, Daphne et al., Three Years of Generative AI: Ten Lessons for the Road Ahead, arXiv:2402.05462 [cs.AI] (2024).
6. GEMA v. OpenAI, Case No. 1 HK O 195/24 (Landgericht München I Nov. 28, 2024).
7. Cooper, Albert F. et al., Extracting Memorized Pieces of (Copyrighted) Books from Open-Weight Language Models, arXiv:2505.12546 [cs.CL] (2025).
8. CJEU, Infopaq International A/S v. Danske Dagblades Forening, Case C-5/08, 2009 ECR I-6569 (2009).
9. Bartz v. Anthropic, PBC, No. 3:23-cv-04516 (N.D. Cal. July 2025) (unpublished opinion).
10. Harper & Row v. Nation Enterprises, 471 U.S. 539 (1985).