

Restitution for Machine Learning: Unjust Enrichment and the Valuation of Training Data in Generative AI Litigation

(A doctrinal and restitutionary analysis of training data valuation in generative AI)

Navin Kumar Shrivastava

Associate Professor, Birla Institute of Management Technology, Greater Noida, India, Email: profnks3010@gmail.com Orcid ID: <https://orcid.org/0000-0003-2845-9115>

Abstract

Generative artificial intelligence has exposed a structural mismatch between conventional copyright damages and the economic architecture of machine learning. Copyright plaintiffs increasingly allege that protected books, news articles, photographs, illustrations, code, and other expressive works were copied into training datasets and processed to build commercially valuable models. Yet a conventional loss-based damages inquiry may understate the economic significance of training where the claimant cannot show that the model displaced sales of the original work. This article develops a restitutionary account of machine-learning training based on unjust enrichment. It asks whether the legal focus should shift, in appropriate cases, from the claimant's lost market income to the defendant's non-consensual acquisition and use of a valuable input. The analysis is grounded in the uploaded reference set, which identifies the growing literature on AI and intellectual property, comparative approaches to generative AI, and the conceptual foundations of restitution, and is supplemented by recent primary legal developments. The article proposes a valuation architecture built around a hypothetical license, adjusted for the marginal contribution of the claimant's work, dataset scarcity, substitutability, training-stage importance, downstream commercialization, and the defendant's own technological contribution. It also examines disgorgement, reasonable royalty, constructive trust, continuing royalties, and the defense of change of position. The article concludes that a generalized entitlement to model profits would be overbroad, but that a narrowly tailored restitutionary or restitution-informed remedy can be justified where acquisition was unauthorized, the defendant received a measurable benefit, the benefit was causally connected to the claimant's protected input, and retention without compensation would defeat the allocation principles of copyright and private law. The paper proposes an "unjust enrichment by extraction" framework, while emphasizing that such a doctrine should not convert copyright into a general property right over facts, ideas, or statistical learning.

Keywords: Generative AI; Training Data; Unjust Enrichment; Restitution; Copyright Remedies; Reasonable Royalty; Machine Learning.

1. Introduction

Generative artificial intelligence has changed the economic meaning of copying. In an ordinary publishing or image market, the central question is often whether an unauthorized copy substitutes for a protected work or diverts revenue from its owner.

In machine learning, the same work may be copied into a training corpus, transformed through preprocessing and numerical representation, and ultimately become one small component of a model whose commercial value arises from the interaction of millions or billions of inputs.

The legal difficulty is therefore not merely whether copying occurred. It is whether the economic value created by training can be translated into a remedy when the claimant's conventional market loss is difficult to prove.

The problem is particularly acute for generative AI. A large language model does not ordinarily deliver a stored book to a user in the same manner as a pirate website. An image model does not necessarily reproduce a single stock photograph every time it generates an image.

The defendant may consequently argue that training is analytically different from commercial substitution: the model is a new technological artifact, the original work remains available for purchase, and any subsequent revenue is attributable primarily to computing infrastructure, engineering, capital, distribution, and product design.

On this account, the plaintiff may establish unauthorized copying yet fail to establish meaningful economic loss. Recent litigation demonstrates why this remedial problem matters. In *The New York Times Company v. Microsoft Corporation* and related OpenAI litigation, plaintiffs alleged that copyrighted articles entered training datasets and that the resulting systems could sometimes reproduce or summarize protected material.

The 2025 district court decision described the alleged process as involving collection and storage of online content, numerical encoding, model training, and later output generation. The court allowed important copyright and contributory theories to proceed while dismissing some misappropriation and Digital Millennium Copyright Act theories.

The litigation thus illustrates the separation between liability theories and the difficult remedial question of how to value the contribution of training material. (*The New York Times Co. v. Microsoft Corp.*, 2025).

The court's description of the training and output stages is important because it shows that the alleged benefit is not limited to a single copied output; it can arise earlier, at the point where the data becomes an input into a commercially deployed model.

Getty Images v. Stability AI presents a parallel problem in image generation. The English High Court litigation concerned allegations involving image datasets, model training, outputs, and related intellectual property rights.

The 2025 judgment illustrates how questions of training, reproduction, distribution, knowledge, and the architecture of an AI platform may require fact-sensitive analysis rather than a single binary characterization of "copying" or "no copying." (*Getty Images (US), Inc. v. Stability AI Ltd.*, 2025).

These cases make visible a broader private-law issue: if the training use falls outside, or cannot be fully addressed by, conventional copyright infringement, what legal institution should govern the value transferred from creators to AI developers?

The reviewed literature that this is an emerging and interdisciplinary field. Awasthy et al. (2024), Azab and Ismail (2024), Jammula and Kanjula (2025), Koellner (2025), Mishra et al. (2023), Moerland (2024), Swamy (2021), and Yadav et al. (2025) approach AI and intellectual property from different comparative, doctrinal, and policy perspectives.

Their common significance is that traditional intellectual property categories are being tested by technologies that separate the input work from the later commercial artifact.

The restitution literature, including Eriksen (2018), Fronek et al. (2021), Rassool and Gibbon (2024), and Ruback et al. (2008), supplies a complementary vocabulary of restoration, benefit, obligation, and the normative significance of payment.

The combination is valuable because the AI dispute is not only a copyright problem. It is also a question about when private law requires a recipient to pay for a benefit obtained without a consensual bargain.

This article therefore focuses on the "gist of the action" problem: whether the inability to demonstrate conventional market substitution should make the economic value of unauthorized training legally invisible.

The restitutionary response shifts the inquiry from "How much did the copyright owner lose?" to "What benefit did the defendant obtain, under what circumstances, and what amount is necessary to reverse an unjust enrichment?"

That shift is not costless. Unjust enrichment is not a general damages multiplier, and a defendant's gross model value cannot automatically be attributed to a single photograph, paragraph, or book. A principled remedy must separate the claimant's contribution from the defendant's independent investments.

The central proposition of this paper is that a narrowly bounded restitutionary framework can complement copyright rather than replace it. Where training involves a legally cognizable protected work, where acquisition or use is unauthorized, where the work confers a measurable economic benefit on the model developer, and where retention of that benefit would be inequitable under the governing law, a court should be able to consider a hypothetical license or other restitutionary valuation.

The appropriate measure should normally be the value of the data input and its incremental contribution to the model, not the entire value of the model or the defendant's total profits.

The objective of the paper is to develop and critically evaluate such a framework. It asks how a single work can be valued within a massive training corpus; whether the defense of change of position should protect investments made before the legal dispute; whether a continuing royalty or constructive trust can be justified; and whether the law should recognize a narrowly defined "unjust enrichment by extraction" principle for data-intensive technologies.

The paper ultimately seeks to connect copyright's exclusionary logic with restitution's gain-based logic without allowing either to swallow the other.

2. Literature Review

The literature supplied for this study can be read as developing along three interrelated trajectories: the adaptation of intellectual property law to artificial intelligence; the comparative and institutional search for a workable balance between innovation and creator protection; and the conceptual development of restitution as a response to benefits obtained without an adequate legal basis. The literature is therefore particularly suitable for a study that moves from copyright classification to private-law remedy.

2.1 Evolution of AI and intellectual property scholarship

Early AI-IP scholarship focused on whether established doctrines could accommodate computational processes without changing their basic structure. Swamy (2021) identified the AI-IP intersection in the Indian context and highlighted emerging doctrinal questions.

Mishra et al. (2023) extended the analysis through a comparison of India and the United States, emphasizing that technological change creates cross-jurisdictional pressure on patent and intellectual property systems.

These contributions are significant because they locate AI within a broader pattern of technological disruption rather than treating generative AI as an isolated legal anomaly. Xiao et al. (2013) further illustrates the importance of IPR and technology intensity to innovation and catch-up, supporting the incentive dimension of the present analysis.

Awasthy et al. (2024) characterize AI and intellectual property as a field of both challenges and opportunities. Their contribution supports a transition from a simple infringement-centered discussion toward institutional design.

Azab and Ismail (2024), through a comparative analytical study, examine the extent to which traditional IP laws can accommodate generative AI.

Their framing is directly relevant to the present paper because the problem is not simply whether a rule exists; it is whether an inherited rule is capable of allocating rights and remedies when the technology changes the relationship between input and output.

Moerland (2024) is especially important for the present study because the discussion is situated within private law rather than copyright doctrine alone. That orientation makes it possible to ask what happens when copyright does not completely determine the legal consequences of a transfer of value.

The private-law perspective helps distinguish property rights from obligations and clarifies why restitution may become relevant when the plaintiff's claim is framed as recovery of a benefit rather than compensation for a conventional loss.

The 2025 literature deepens this movement. Jammula and Kanjula (2025) emphasize balance between innovation and creativity; Koellner (2025) addresses ethical innovation and societal benefit; Mohanta et al. (2025) examine AI-driven innovation and ethical challenges; and Yadav et al. (2025) integrate AI with patents, trademarks, copyright, and traditional knowledge.

Sector-specific studies extend the debate to agricultural research, deepfakes, fashion design, and AI-generated creativity (Kebede, 2025; Narang, 2025; Niyompatama & Lapatoura, 2026; Pala & Desai, 2026). Mahfuzzah et al. (2026) further illustrates comparative reform through an Indonesian fair-use perspective. Collectively, these works support systems-level adaptation across jurisdictions and creative sectors.

2.2 Training data and the classification problem

A recurring issue across the AI-IP literature is the distinction between input, process, and output. A training dataset may contain protected expression, but the model's learned parameters are not necessarily a collection of readable copies. The model may instead encode statistical relationships, patterns, and representations.

This technological distinction generates a legal classification problem. If the law regards training as copying, conventional infringement analysis becomes central. If training is characterized as non-expressive learning, fair use, text and data mining, or another permitted use, plaintiffs may face a remedial gap even when the developer has extracted substantial economic value from protected content.

The U.S. Copyright Office's 2025 Part 3 report on generative AI training illustrates the institutional importance of this issue.

The Office stated that its study examined the legal implications of training AI models on copyrighted works, including licensing considerations and allocation of potential liability.

The Office's initiative was informed by more than 10,000 public comments, demonstrating the scale of the policy disagreement (U.S. Copyright Office, 2025).

The report's significance for this article is not that it resolves the private-law question, but that it confirms the centrality of training data, licensing, and liability to current copyright policy.

The New York Times litigation adds a factual dimension. The 2025 court decision described training as involving collection and storage of content, numerical encoding, and adjustment of model parameters to predict subsequent words.

It also distinguished the training stage from the output stage, where the model responds to user prompts.

This distinction matters for restitution because the relevant benefit may arise at the training stage even when no particular training work appears in a later output. (*The New York Times Co. v. Microsoft Corp.*, 2025).

Getty Images similarly demonstrates that an image model can create legal questions at several stages: acquisition of images, dataset construction, training, model distribution, and outputs. The 2025 English High Court judgment also illustrates that factual knowledge and the precise mechanics of dataset creation may become important to liability. (*Getty Images (US), Inc. v. Stability AI Ltd.*, 2025).

2.3 The evolution of restitutionary thinking

The restitution references in the uploaded set provide a different but necessary conceptual vocabulary. Eriksen (2018) examines conflicting duties and restitution within a trusting relationship. Fronek et al. (2021) use restitution in a context involving intercountry adoption and demonstrate that restitution is not merely a mechanical financial calculation; terminology and normative context matter.

Rassool and Gibbon (2024) expressly emphasize the importance of distinguishing restitution from repatriation. Ruback et al. (2008), although focused on victims' perceptions of restitution in criminal justice, demonstrates that payment has a relational and normative dimension: restitution can be understood not only as a financial transfer but also as recognition of a wrong and restoration of balance.

2.4 The unresolved valuation question

The literature is much stronger on classification than on valuation. AI-IP scholarship repeatedly identifies uncertainty about training, ownership, infringement, fair use, ethical boundaries, and cross-jurisdictional coordination. Yet less attention has been devoted to the precise economic problem posed by an individual training work within a massive corpus. This is where the present paper contributes.

The conventional copyright framework already contains gain-based ideas. Section 504 permits recovery of the copyright owner's actual damages and any additional profits of the infringer attributable to the infringement.

The Supreme Court's decision in *Petrella v. Metro-Goldwyn-Mayer, Inc.* (2014) recognizes the role of profits attributable to infringement and discusses the possibility of continued use upon payment of a reasonable royalty in an appropriate equitable setting.

But AI training creates a distinctive attribution problem: the model's revenue may be the product of millions of works plus substantial independent investment. The challenge is therefore to avoid both under-compensation and over-attribution.

The literature consequently supports a research agenda that treats valuation as a separate doctrinal question. The gap is not simply "AI needs new copyright law." It is more precise: existing scholarship does not sufficiently specify how restitutionary principles could value a non-exclusive, non-market, technologically transformed use of copyrighted training material when direct loss is small or unobservable.

2.5 Synthesis of the literature

As summarized in Table 1, the literature converges on the need for doctrinal adaptation, but it leaves open the architecture of a gain-based remedy. The present paper therefore treats unjust enrichment as a bridge between property and obligation.

Copyright supplies the legal interest in protected expression; restitution supplies a method for asking whether a defendant should retain a benefit obtained through unauthorized extraction; economic valuation supplies the mechanism for translating that benefit into a defensible monetary amount.

As shown in Table 1, the literature converges on doctrinal adaptation and innovation balance, but leaves the valuation of individual training contributions insufficiently specified.

Literature stream	Representative sources	Central contribution	Unresolved issue for this study
AI and IP foundations	Swamy (2021); Mishra et al. (2023); Awasthy et al. (2024)	Identifies doctrinal disruption and cross-jurisdictional challenges	How should economic benefit be remedied when loss is difficult to prove?
Generative AI and copyright	Azab & Ismail (2024); Jammula & Kanjula (2025); Moerland (2024)	Examines accommodation, innovation, creativity, and private-law implications	How can training value be separated from output value?
Global and ethical governance	Koellner (2025); Mohanta et al. (2025); Yadav et al. (2025)	Calls for coherent, ethical, and technology-sensitive systems	What remedial architecture best aligns incentives?
Sector-specific AI-IP	Kebede (2025); Narang (2025); Niyompatama & Lapatoura (2026); Pala & Desai (2026)	Shows application across agriculture, deepfakes, fashion, and AI creativity	How should contribution be valued across different data types?
Restitution and restoration	Eriksen (2018); Fronek et al. (2021); Rassool & Gibbon (2024); Ruback et al. (2008)	Highlights restoration, benefit, payment, terminology, and normative balance	How can restitution concepts be translated into AI training disputes?

Source: Author's Own, based on the LR.

3. Theoretical Underpinning

The paper is underpinned by four complementary theoretical perspectives: the property-obligation distinction in private law; the law of unjust enrichment and restitution; the law and economics of intellectual property; and the theory of hypothetical licensing.

3.1 Property versus obligation

Copyright is commonly understood as a bundle of exclusive rights. This property-like structure allows an owner to control specified acts such as reproduction and distribution. Yet a copyright is not necessarily a property right over every form of learning from a work. A reader may observe a painting, infer a technique, or learn a historical fact without committing infringement. The private-law question becomes more difficult when computational ingestion occurs at industrial scale.

The property-obligation distinction helps avoid a conceptual error. If copyright is treated as ownership of every informational value extracted from a work, the right may become broader than its statutory design.

Conversely, if training is treated as entirely outside copyright because the model does not contain a readable copy, the law may ignore a significant commercial appropriation of the creator's economic resource. Restitution can occupy the intermediate space. It does not require treating the data as an absolute property object. Instead, it asks whether a defendant's acquisition and retention of a benefit creates an obligation to make restitution.

3.2 Unjust enrichment theory

Unjust enrichment focuses on enrichment, at the claimant's expense or through a legally significant transfer, and on the absence of a sufficient basis for retention. The core analytical advantage is that the plaintiff need not prove a mirror-image loss equal to the defendant's gain.

This does not mean that loss is irrelevant. Rather, the focus changes from compensation for a lost sale to restoration of an improperly acquired benefit.

For machine learning, the relevant benefit may include lawful access to a protected corpus, reduced data-acquisition costs, enhanced model performance, improved domain coverage, reduced error, or the ability to launch a product without negotiating licenses with individual creators. The benefit is not necessarily the model's entire market value. It is the value of the input or opportunity appropriated through the training process.

3.3 Law and economics

A law-and-economics perspective asks whether legal rules create efficient incentives for data acquisition, licensing, innovation, and creation. A rule requiring payment for every training use could impose enormous transaction costs and discourage beneficial innovation.

A rule providing no remedy for unauthorized commercial extraction could encourage systematic free riding and underinvestment in creative works.

The efficient rule therefore depends on transaction costs and substitutability. If individual licenses are impracticable but collective licensing is feasible, a collective rate may be efficient. If a work has little marginal influence on model performance, the valuation should be correspondingly small. If a curated dataset has unique and demonstrable importance, the rate should be higher. The aim is not to maximize creator compensation or minimize AI costs in isolation, but to align incentives across the innovation ecosystem.

3.4 Hypothetical licensing

A hypothetical license is particularly attractive where ex ante negotiation was impossible or did not occur. It asks what reasonable parties would have agreed upon for the relevant use. The Supreme Court's copyright jurisprudence has recognized reasonable royalty as a possible way of structuring continued use in appropriate circumstances, while copyright profit remedies already distinguish profits attributable to the protected work from returns attributable to the defendant's own enterprise (*Petrella v. Metro-Goldwyn-Mayer, Inc.*, 2014).

For AI training, however, a conventional royalty cannot simply be multiplied by the number of files. The license must be defined by scope: duration, model type, training stage, exclusivity, geography, commercial deployment, output restrictions, and risk. The theoretical model proposed in this article therefore treats hypothetical licensing as an anchor rather than a fixed number.

3.5 Restitution and remedial proportionality

The theory of remedial proportionality prevents restitution from becoming punitive. A restitutionary award should correspond to the benefit unjustly retained, subject to defenses and attribution principles.

This is particularly important because the defendant's model may contain value created by engineers, capital, computing resources, proprietary algorithms, user feedback, and many other inputs. The claimant should not receive the value of the defendant's independent contribution merely because the claimant's work was among the inputs.

The theoretical framework is summarized in Figure 1. The model proceeds from legal entitlement to enrichment, then from enrichment to valuation, and finally from valuation to a proportionate remedy.

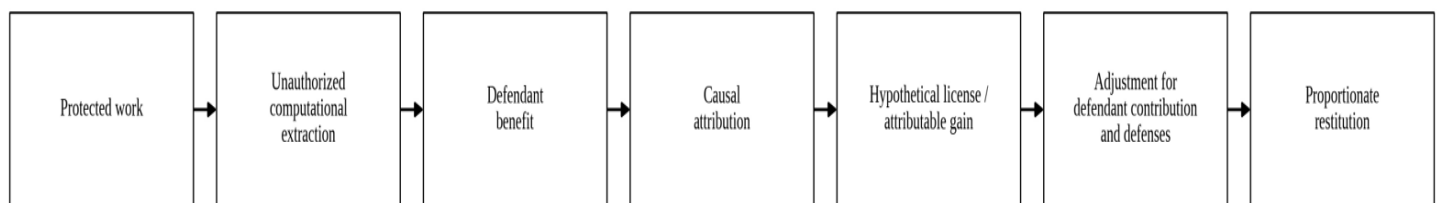
As shown in Figure 1, the theoretical framework links the legal entitlement in a protected work to enrichment, causal attribution, valuation, and a proportionate remedy. The figure is structured as a left-to-right horizontal pipeline with seven stages (as elements), rendering the analytical process as a single uninterrupted chain of reasoning.

The flow opens with "Protected work," which establishes the legal entitlement, then proceeds to "Unauthorized computational extraction," the act that triggers liability. From there it advances to "Defendant benefit" and "Causal attribution," two boxes that together isolate the gain actually traceable to the infringement rather than to the model's overall success.

The fifth box, "Hypothetical license / attributable gain," converts that benefit into an economic measure by asking what a willing licensee would have paid, while the sixth box, "Adjustment for defendant contribution and defenses," reduces or rebuts that measure to reflect value the defendant added independently and any substantive defenses available.

The seventh box, "Proportionate restitution," delivers the final remedial quantum, ensuring the award reflects only the attributable share of the benefit rather than the model's total value. The italicized note beneath the flow makes explicit the framework's two controlling principles: legal entitlement is separated from economic valuation, and the model's gross value is never automatically attributed to the claimant.

Figure 1. Conceptual framework for restitutionary valuation of generative AI training



Source: Author's Own.

The framework separates legal entitlement from economic valuation and prevents total model value from being attributed automatically to the claimant.

4. Research Gap

The reviewed literature establishes that AI challenges traditional IP categories, that comparative reform is occurring, and that restitution is conceptually capable of responding to benefits obtained without a consensual exchange. However, four specific gaps remain.

First, the literature does not sufficiently connect AI training disputes with the gain-based structure of restitution. AI-IP studies generally ask whether training is infringement, fair use, text and data mining, or another lawful use. Restitution studies generally examine relationships and transfers outside AI. The doctrinal bridge between the two remains underdeveloped.

Second, the valuation problem is under-theorized. Existing discussion often moves directly from "training data has value" to a proposed license fee or royalty without specifying how the value of one work should be isolated within a corpus containing millions or billions of works.

This paper responds by proposing a contribution-sensitive valuation model.

Third, the defense of change of position has not been sufficiently integrated into the AI training debate. AI developers may have spent enormous sums on compute, infrastructure, engineering, and model development before a court determines that a particular training use required payment.

A fair restitutionary doctrine must determine whether those expenditures are relevant to the defendant's culpability, to the amount of recovery, or to both.

Fourth, the literature has not clearly articulated when restitution should remain complementary to copyright and when it would improperly circumvent copyright's statutory boundaries.

A new cause of action cannot simply re-label infringement whenever the copyright claim is weak. The present study therefore proposes limiting principles for an "unjust enrichment by extraction" theory.

5. Research Questions

RQ1. Under what circumstances can unjust enrichment provide a coherent remedial basis for unauthorized use of copyrighted works in generative AI training?

RQ2. How should the benefit conferred by an individual training work be valued when that work is combined with millions of other inputs and transformed into model parameters?

RQ3. Can a hypothetical willing buyer/willing seller license provide a principled basis for valuation, and what adjustments are required for AI-specific use?

RQ4. When should disgorgement, a constructive trust, a reasonable royalty, or an ongoing royalty be preferred as a remedy?

RQ5. To what extent should the defense of change of position protect AI developers that invested substantial resources in model training before liability was established?

RQ6. Should private law recognize a narrowly defined doctrine of unjust enrichment by extraction for large-scale computational data use, and what safeguards would prevent overreach?

6. Research Methodology

6.1 Research Design

The study adopts a qualitative, doctrinal, and literature-based research design. It does not collect primary survey responses or experimental observations. Instead, it integrates the uploaded reference set with selected primary legal materials and contemporary institutional sources concerning generative AI training.

The design is appropriate because the research questions concern legal classification, private-law principles, valuation, and remedies rather than causal relationships among measurable behavioral variables.

6.2 Research Type

The research is exploratory and explanatory. It is exploratory because restitutionary treatment of generative AI training remains unsettled. It is explanatory because the study seeks to construct a coherent account of why a benefit may be unjustly retained and how that benefit can be valued without treating the entire AI model as the claimant's property.

6.3 Data Collection Tools

The principal data sources are secondary legal and scholarly materials. The uploaded reference set contains 20 references set, spanning AI and intellectual property, comparative legal analysis, restitution, private law, and technology governance.

The study also uses primary and institutional materials for contemporary legal context, including United States Copyright Office materials, judicial decisions concerning *The New York Times* and OpenAI-related litigation, *Getty Images v. Stability AI*, and selected Supreme Court authority concerning copyright remedies.

These materials are used to distinguish current legal developments from the conceptual propositions drawn from the uploaded academic literature.

6.4 Data Analysis Tools

Because this is a qualitative literature and doctrinal study, percentage conceptual literature as , SEM, regression, ANOVA, DEA, AHP, t-tests, and f-tests are not appropriate. The principal analytical tool is thematic analysis using Gioia-style multi-order coding.

The method identifies first-order concepts close to the terminology of the sources, groups them into second-order themes, and then develops aggregate dimensions. This structure is useful for converting a heterogeneous literature into a transparent conceptual framework.

6.5 Questionnaire Design

Not applicable. The study is not survey-based and therefore does not use a questionnaire or Likert scale.

6.6 Thematic Analysis Using GIOA Method

The literature was coded into first-order concepts, second-order themes, and aggregate dimensions. As shown in Exhibit 1, the coding moves from recurring source-level concepts such as "balance between innovation and creativity," "fair use reform," "AI and IPR intersection," "restitution," and "ethical boundaries" to broader themes of doctrinal adaptation, benefit allocation, and remedial design.

The uploaded references explicitly contain works on restitution and on AI-IP balance, which makes this coding structure directly grounded in the source set.

The coding identifies the conceptual terrain; the proposed restitutionary model is an analytical synthesis developed from that terrain and contemporary legal materials.

Exhibit 1 presents the GIOA multi-order coding structure used for the literature analysis.

First-order codes	Second-order themes	Aggregate dimension
AI-IP intersection; traditional law; emerging issues; cross-jurisdictional examination	Doctrinal adaptation	Recalibrating legal categories for AI
Balance innovation and creativity; ethical innovation; societal benefit; ethical boundaries	Institutional legitimacy	Balancing innovation and creator interests
Training data; fair use reform; infringement; ownership; AI-generated creativity	Value allocation	Determining the legal significance of computational use
Restitution; payment; restoration; terminology; trusting relationship	Restitutionary obligation	Recognizing and measuring unjust benefit
Licensing; protection; rights allocation; global systems	Remedial design	Building a workable post-liability remedy

Source: Author's Own, derived from thematic coding of the uploaded reference set.

7. Analysis

7.1 The "gist of the action" problem

The central remedial difficulty can be stated as follows. Suppose a publisher owns 100,000 copyrighted articles. An AI company copies those articles into a training corpus and uses them, along with hundreds of millions of other documents, to train a model.

The publisher proves that the copying was unauthorized. Yet the publisher cannot identify a customer who would otherwise have purchased an article but instead purchased the AI service.

Nor can the publisher show that the model's revenue is entirely attributable to its articles. A conventional lost-sales measure may therefore be close to zero.

That conclusion is incomplete because the economic transaction is not necessarily a substitute for a sale. The AI company may have acquired an input that reduced its cost of building a commercially valuable model.

The benefit is an avoided licensing cost, improved training quality, broader domain coverage, or some combination of these. The legal question is consequently not whether the plaintiff lost a sale. It is whether the defendant received a legally cognizable benefit through the unauthorized use.

The distinction can be represented as:

- Loss-based inquiry: Plaintiff's lost revenue $\approx$ compensatory damages.
- Gain-based inquiry: Defendant's benefit from unauthorized input $\approx$ restitutionary valuation.
- The two inquiries can produce different numbers without either being conceptually erroneous. The first protects the claimant's expected economic position. The second prevents the defendant from retaining an unjust benefit. The difficulty is determining when the second inquiry is legally available.

7.2 The benefit conferred by a single work

The key valuation problem is attribution. A model does not usually assign a public price to each training item. A single image may improve recognition of a visual concept by an immeasurable amount.

A single article may contribute vocabulary, factual associations, style patterns, or domain knowledge. A book may provide long-range linguistic structure.

A code repository may improve the model's ability to generate a programming pattern. The contribution can therefore be real without being individually observable.

The paper proposes five valuation filters.

First, legal relevance. Only works that fall within the protected subject matter and are implicated by the asserted legal theory should enter the restitutionary pool. Facts, ideas, public-domain material, and purely unprotected information should not be converted into private property through restitution.

Second, causal contribution. The claimant must establish a credible connection between the work and the model benefit. Evidence may include dataset inclusion, training logs, ablation studies, benchmark performance, memorization tests, domain-specific accuracy, or evidence concerning the dataset's curation.

Third, scarcity and substitutability. A work that could readily be replaced by thousands of equivalent works should normally have a lower marginal value than a unique or curated work that is difficult to substitute.

Fourth, commercial significance. The valuation should consider whether the model is commercial, the scale of deployment, the relevant product revenues, and the period of use. Non-commercial research and commercial deployment should not automatically attract the same rate.

Fifth, defendant contribution. Engineering, compute, model architecture, product development, safety systems, distribution, and capital investment must be separated from the claimant's contribution. Restitution should not transfer the defendant's own value to the claimant.

7.3 A proposed contribution-sensitive valuation equation

A practical framework can be expressed as:

Restitutionary Benefit = Base Hypothetical License Value x Scope Factor x Causation Factor x Scarcity Factor x Commercial Use Factor x Attribution Factor.

The Base Hypothetical License Value is the price that a willing buyer and willing seller would have agreed for the relevant training use if they had negotiated ex ante with adequate information.

The Scope Factor captures duration, geography, model family, training stage, exclusivity, and output rights.

The Causation Factor reflects evidence that the work materially contributed to the relevant model capability. The Scarcity Factor reflects substitutability. The Commercial Use Factor reflects deployment and monetization.

The Attribution Factor prevents the claimant from recovering value generated primarily by the defendant's independent contribution.

The equation is not intended to create false precision. It is a framework for disciplining evidence. Where exact measurement is impossible, a structured valuation range may be preferable to treating the benefit as zero.

7.4 Hypothetical license as the primary anchor

The hypothetical license performs a unifying function within the framework because it is the only remedial measure that translates the defendant's unauthorized computational extraction into an economic equivalent without requiring proof of actual market loss.

Where compensatory damages stall for want of a proven licensing market or demonstrable displacement of sales, the hypothetical license asks the more tractable question of what a reasonable willing licensor and a reasonable willing licensee would have agreed ex ante for the use actually made of the protected work.

That inquiry is restitutionary in character, measuring the defendant's incursion against the market value of permission, yet it is disciplined by the same causal and attributional filters that govern the rest of the framework, so that the notional fee captures only the proportion of the model's training corpus and downstream value genuinely traceable to the claimant's contribution.

Treating the hypothetical license as the primary anchor also reconciles the remedial ladder's two structural commitments: it preserves compensation's causal rigour while accommodating restitution's gain-based logic, and it does so through a metric courts already understand from established intellectual-property practice. Subordinate remedies then calibrate around this anchor.

Attributable profits operate as a statutory override where infringement yields identifiable gains beyond the notional fee; ongoing royalties convert an established use into a forward-looking licence once its value has been anchored; and injunctive or retraining relief reserves the most coercive response for harm that no monetary quantum can adequately repair.

Anchoring the analysis in the hypothetical license therefore prevents total model value from being attributed automatically to the claimant while still awarding a quantum that reflects what the defendant actually took.

7.5 Is disgorgement of model profits appropriate?

Full disgorgement of model profits is generally too broad. The model's revenue reflects many contributions.

Treating the entire profit stream as the defendant's unjust enrichment from one claimant would ignore causation and overcompensate the claimant.

However, disgorgement can be appropriate where evidence shows that the claimant's work was a material and identifiable source of the defendant's gain. The copyright concept of profits attributable to infringement provides a useful analogy.

In *Petrella*, the Supreme Court emphasized that the defendant may retain returns attributable to its own enterprise while profits attributable to infringement may be assessed within the statutory framework (*Petrella v. Metro-Goldwyn-Mayer, Inc.*, 2014).

The AI context therefore supports a partitioned approach: first identify the revenue stream connected to the relevant model or service; then estimate the incremental contribution of the protected training corpus; then allocate only the relevant share.

7.6 Constructive trust and continuing royalties

A constructive trust is a powerful remedy and should be used cautiously. It is conceptually attractive where a defendant holds a specific, traceable asset or identifiable proceeds subject to an equitable obligation. A machine-learning model, however, is not simply a fund containing a claimant's money. It is a composite technological artifact. A constructive trust over the entire model would therefore be disproportionate.

A more plausible approach is a constructive-trust-like accounting over identifiable proceeds attributable to the unauthorized use, or a continuing royalty that allows lawful future deployment while preventing continued uncompensated extraction. A continuing royalty is especially attractive where destroying or retraining the model would be inefficient and where the legal system prefers compensation over technological destruction.

Petrella is instructive because the Supreme Court noted the possibility of allowing continued use upon payment of a reasonable royalty in an appropriate case. That principle is not a holding that AI training always warrants a royalty. It demonstrates, however, that equitable relief can be structured to avoid both uncompensated continued use and destructive injunctions (*Petrella v. Metro-Goldwyn-Mayer, Inc.*, 2014).

7.7 The change of position defense

The change of position defense asks whether the defendant's reliance on receipt of a benefit makes restitution inequitable. AI developers can make a serious claim here. Model training may require enormous expenditures on compute, engineering, data processing, testing, and deployment. If a developer reasonably believed the training use was lawful and invested billions before a legal rule changed, requiring complete disgorgement could create severe disruption.

But change of position should not become a blanket defense. Three factors are especially important. First, knowledge. A defendant that knew the legal position was uncertain, received repeated licensing demands, or deliberately avoided available licensing information has a weaker equitable case than a defendant that reasonably relied on settled law.

Second, preventability. If the developer could have preserved datasets, tracked provenance, negotiated licenses, or designed an opt-out system at modest cost, the defense should be narrower.

Third, benefit persistence. The defense is weaker where the benefit remains embedded in a profitable model that continues to generate revenue. The fact that the defendant spent money to create the model does not mean that all later benefit has disappeared.

Change of position should therefore operate primarily as a proportionality mechanism.

It may reduce an award, alter its timing, or support a forward-looking license, rather than automatically eliminating restitution.

7.8 Should the law recognize "unjust enrichment by extraction"?

The analysis supports recognition only in a limited form. The proposed doctrine should contain six elements:

1. A legally cognizable protected input was extracted, copied, ingested, or otherwise appropriated for computational training.
2. The defendant obtained a measurable or inferable economic benefit from that use.
3. The benefit is causally connected to the claimant's input rather than merely to the defendant's independent technology.
4. The claimant lacks an adequate remedy under existing contract or copyright provisions for the particular benefit asserted.
5. Retention of the benefit is unjust under the governing law, considering authorization, notice, industry practice, public policy, and the defendant's conduct.
6. The remedy is limited to the value of the unjustly retained benefit and is reduced to account for the defendant's independent contribution and applicable defenses.

This doctrine would not create ownership over information as such. It would regulate a particular mode of commercial extraction where a protected work is used as an input to create value without a sufficient legal basis for retaining that value.

7.9 Comparative implications

The comparative literature demonstrates substantial jurisdictional diversity in AI-IP rules. The restitutionary model should therefore be treated as a private-law framework rather than a universal statutory rule.

Where a jurisdiction has an explicit text and data mining exception, restitution should not simply defeat the statutory policy; the interaction between the exception and any remuneration obligation must be examined. Where no such exception exists, restitution must still respect preemption and equivalence limits. The model is deliberately modest and does not claim that every unauthorized training use is unjust enrichment.

7.10 Analysis of the literature through GIOA coding

As shown in Exhibit 2, the first-order codes from the literature cluster into three major second-order themes: doctrinal adaptation, value allocation, and ethical or institutional legitimacy. These aggregate into one overarching dimension: "recalibrating private law for AI-enabled value extraction."

The restitution references contribute the concepts of restoration, payment, terminology, and normative balance, while the AI-IP references contribute concepts of infringement, fair use, ownership, innovation, and ethical boundaries.

The analytical result is that the literature does not support a simple "creator versus AI company" dichotomy. It supports a more nuanced allocation problem in which rights, obligations, innovation incentives, and remedial proportionality must be coordinated.

Exhibit 2 translates the coding into the proposed analytical framework.

Aggregate dimension	Analytical interpretation	Implication for proposed doctrine
Recalibrating legal categories for AI	Existing IP concepts remain relevant but require process-sensitive analysis	Do not replace copyright; distinguish training, model, and output stages
Balancing innovation and creator interests	Both under-protection and over-protection can distort incentives	Use proportionality and defendant-contribution adjustments
Determining legal significance of computational use	The economic value of training may exist without direct market substitution	Permit benefit-based analysis where legal prerequisites are met
Recognizing and measuring unjust benefit	Restitution is concerned with benefit and restoration rather than only loss	Use hypothetical license and attributable benefit
Building a workable post-liability remedy	Future use may be more efficient than destruction or retraining	Consider continuing royalty and tailored equitable relief

Source: Author's Own.

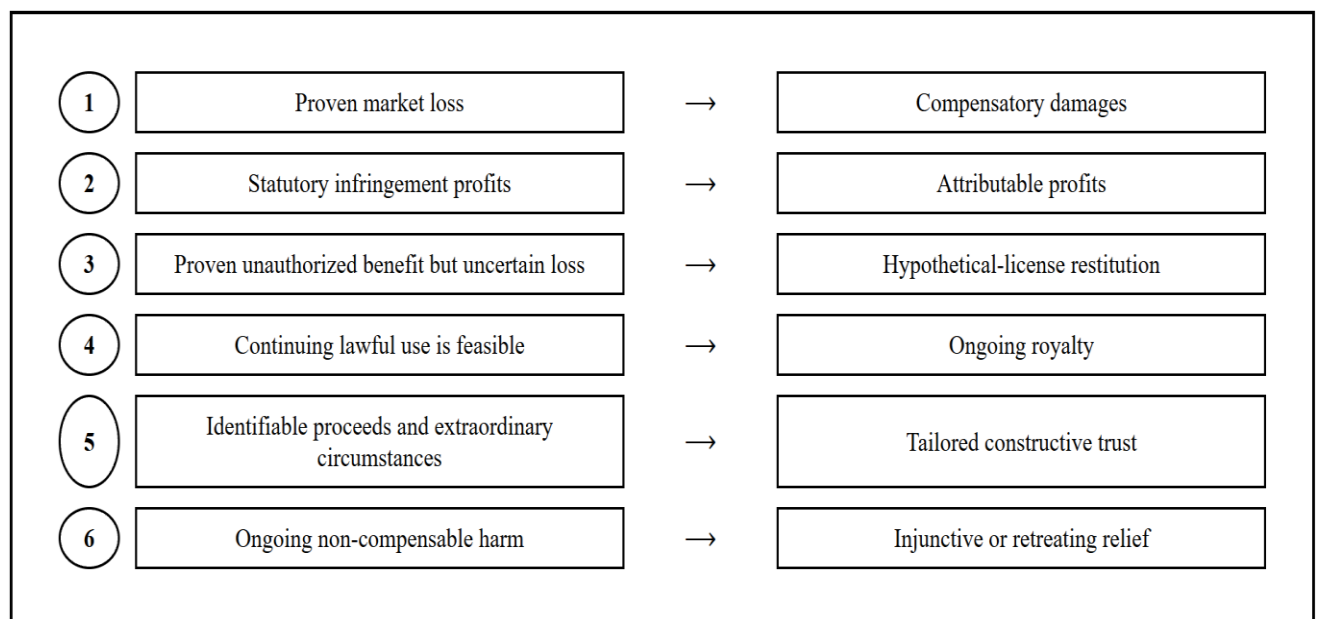
The remedial options are summarized in Figure 2. The figure is laid out as a vertical ladder of six descending rungs, each rung consisting of a numbered circle, a condition box on the left, a rightward arrow, and a remedy box on the right, with downward arrows connecting successive rungs to signal progressive fallback.

The ladder begins with the strongest and most conventional remedy: where "Proven market loss" is established, "Compensatory damages" follow at the top rung. Where direct loss is not provable but infringement profits are recoverable as a matter of statute, the second rung supplies "Attributable profits," narrowing the award to gains causally linked to the infringement.

The third rung applies where "Proven unauthorized benefit but uncertain loss" is shown, offering "Hypothetical-license restitution" as a restitutionary substitute that values the unauthorized use by reference to a notional bargain. The fourth rung accommodates cases where "Continuing lawful use is feasible," permitting an "Ongoing royalty" that converts the infringement into a forward-looking license.

The fifth rung, reserved for "Identifiable proceeds and extraordinary circumstances," authorizes a "Tailored constructive trust" that treats specific traced proceeds as held for the claimant. The sixth and final rung responds to "Ongoing non-compensable harm" with "Injunctive or retraining relief," the most coercive response, reserved for harm that monetary remedies cannot adequately address. The descending structure mirrors the substantive logic that remedies should escalate only as lower rungs fail.

Figure 2. Remedial ladder for AI training disputes



Source: Author's Own.

8. Findings

First, the conventional loss model is incomplete for machine-learning training. The absence of direct market substitution does not necessarily mean that the claimant conferred no economic value. Training can create an avoided acquisition cost and enhance model capabilities.

Second, unjust enrichment is conceptually distinct from copyright infringement. It can focus on the defendant's benefit rather than the claimant's lost sales, but it must remain bounded by the subject matter and policy of copyright law.

Third, the hypothetical license is the most administrable starting point for valuation where comparable market evidence exists. It reflects the economic value of the use without attributing the entire model's value to the claimant.

Fourth, a single work should not be valued by dividing model revenue mechanically by the number of training files. The value of a work is heterogeneous and depends on causal contribution, scarcity, substitutability, and commercial significance.

Fifth, disgorgement should be contribution-sensitive. Full model profits are ordinarily too remote from an individual training work. A calibrated accounting of attributable gains is more consistent with restitutionary proportionality.

Sixth, change of position is relevant but not absolute. Expenditures made in good faith can reduce or reshape relief, but persistent commercial benefit, prior notice, and avoidable data-governance failures weaken the defense.

Seventh, a carefully limited doctrine of unjust enrichment by extraction is theoretically defensible. Its legitimacy depends on strict elements that prevent the doctrine from becoming a backdoor property right over facts, ideas, or general statistical learning.

9. Discussion

The findings reposition the dispute from a binary copyright question toward a layered private-law analysis.

RQ1 asks when unjust enrichment can provide a remedial basis. The answer is that unjust enrichment should become relevant when the claimant can identify a legally cognizable input, an economic benefit, causal connection, and an absence of sufficient legal basis for retention.

The mere fact that a model learned from publicly accessible information should not be enough. Public availability is not necessarily equivalent to a license for industrial computational extraction, but neither does it automatically establish unjustness. The surrounding legal regime, contractual terms, statutory exceptions, industry practices, and defendant conduct must be considered.

RQ2 concerns valuation. The answer is a contribution-sensitive approach. The central unit should not be "one file" in isolation but "one legally protected input's incremental value within a defined training and commercialization context." This can be estimated through market comparables, ablation evidence, dataset substitution studies, benchmark effects, provenance records, and economic modeling.

RQ3 concerns hypothetical licensing. The analysis supports it as a primary valuation anchor because it is familiar to courts and directly addresses the counterfactual transaction. However, the hypothetical negotiation must be technically informed. It should account for training purpose, model scale, duration, commercial deployment, geographic scope, and whether the work is replaceable.

RQ4 concerns remedies. A reasonable royalty should ordinarily be preferred where future use can be separated and quantified. An accounting of attributable profits may be appropriate where the claimant proves that the protected corpus materially generated revenue.

A constructive trust should be reserved for identifiable proceeds or situations where ordinary monetary relief is inadequate. An injunction or retraining order should be considered where continued use creates ongoing harm that money cannot reasonably address.

RQ5 concerns change of position. The defense should protect reasonable reliance but not deliberate free riding. This approach is consistent with restitution's equitable character.

The defense should be assessed after liability and benefit are established, and its effect should be proportional to the defendant's reliance and the remaining benefit.

RQ6 asks whether a new doctrine is needed. The analysis suggests that a narrowly defined unjust enrichment by extraction doctrine could fill a remedial gap, but a statutory intervention may ultimately

be preferable because AI training has cross-border and systemic consequences. The doctrine should therefore be understood as a judicially disciplined interim framework, not a substitute for legislation.

The findings also connect with the uploaded literature. The recurring concern with balancing innovation and creativity (Jammula & Kanjula, 2025), adapting traditional IP to generative AI (Azab & Ismail, 2024), and building coherent systems (Koellner, 2025) supports a calibrated remedy.

The restitution sources reinforce the restorative and normative dimensions of payment, while Rassool and Gibbon (2024) show why terminology and legal basis matter. The private-law classification question is decisive. If copyright is understood as protection against expressive copying, computational learning may sometimes fall outside its core.

But deliberate commercial extraction of protected works may create a separate restitutionary obligation where the governing law recognizes unjust retention. This does not create property in learning itself; it recognizes that commercial conduct can create obligations even without conventional market substitution.

The approach also addresses the "gist of the action" concern. A plaintiff need not choose between proving a lost sale and receiving nothing, while the defendant should not lose its entire business value because one input was used. Restitution provides a middle path: identify the benefit, isolate the claimant's contribution, account for the defendant's contribution, apply defenses, and award only the amount necessary to reverse the unjust enrichment.

The approach also encourages better AI data governance. Provenance, licensing, and dataset documentation would improve both compliance and later valuation, reduce litigation uncertainty, and facilitate collective licensing.

Recent litigation supports the practical importance of this architecture. The New York Times decision shows that courts may have to examine detailed allegations about how datasets are collected, stored, processed, and used in model training.

The decision also shows that some claims can fail while others survive, underscoring the importance of matching the remedy to the legal theory. (*The New York Times Co. v. Microsoft Corp.*, 2025). Getty Images demonstrates that training disputes can involve multiple rights and technical processes across jurisdictions. (*Getty Images (US), Inc. v. Stability AI Ltd.*, 2025).

The analysis therefore favors a remedial ladder: compensation where loss is proven; statutory profits where copyright infringement supports them; hypothetical-license restitution where unauthorized benefit is proven but loss is difficult to establish; continuing royalty where future use is valuable and separable; and stronger equitable relief only where monetary measures are inadequate.

10. Conclusion

Generative AI exposes a fundamental weakness in a purely loss-centered approach to unauthorized training: a creator may confer economic value on an AI developer without losing a conventional sale.

The resulting gap is not necessarily a reason to abandon copyright, nor is it sufficient to declare that every training use is unjust. It is a reason to distinguish the questions of property, obligation, enrichment, and remedy.

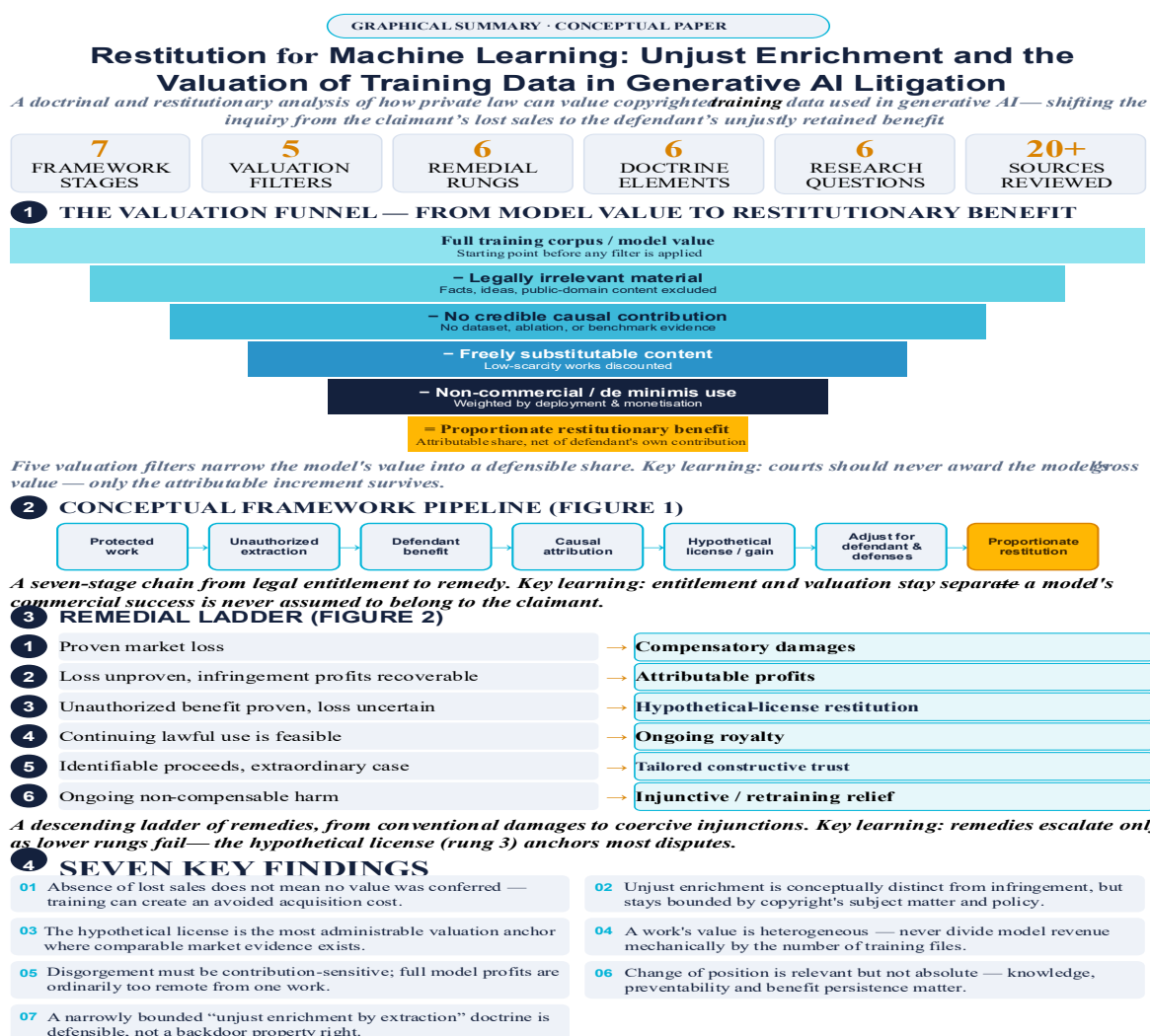
This paper has argued that restitution can provide a principled bridge. Unjust enrichment shifts attention to the benefit obtained by the AI developer, while copyright continues to determine the legally protected subject matter and the boundaries of exclusive rights.

A hypothetical license can supply an administrable valuation anchor, but the amount must be adjusted for causal contribution, scarcity, substitutability, commercial use, and the defendant's independent technological contribution.

The proposed "unjust enrichment by extraction" doctrine should therefore be narrow. It should not create property rights over facts, ideas, general knowledge, or statistical patterns. It should apply only where protected inputs were commercially extracted, an identifiable benefit resulted, causal connection can be demonstrated, no adequate legal basis for retention exists, and the remedy can be proportionately measured.

The principal contribution is conceptual and remedial: the AI training dispute is reframed as value allocation rather than only copying. This offers courts and policymakers a way to protect creator interests while preserving innovation incentives. Longer-term solutions may include legislation, collective licensing, dataset transparency, and standardized training-data markets. A detailed summary of the study is depicted in the Exhibit 1.

Exhibit 1: Study Summary



Source: Author's Own

11. Implications

11.1 Implications for industry and practice

AI developers should treat training data as an economic and legal input rather than as an invisible technical resource.

Data provenance systems should record source, license status, jurisdiction, creator information, opt-out signals, dataset version, date of ingestion, and model families trained on the data. Such records would materially improve both compliance and later valuation.

Developers should consider tiered licensing. High-value curated content can be licensed directly or collectively, while lower-value or freely licensed material can be governed through automated permissions. A uniform fee for every item is unlikely to be efficient.

Publishers, photographers, artists, and other creators should develop clearer licensing offers for machine-learning uses. A licensing market requires standardized descriptions of training scope, model purpose, duration, and downstream use. Creators should also preserve evidence of commercial licensing rates and prior negotiations because these may become important in a hypothetical-license assessment.

For litigators, claims should not rely solely on dataset inclusion. Parties should collect evidence on dataset provenance, model versions, benchmark or ablation evidence, memorization, licensing alternatives, commercial deployment, and relevant revenue.

11.2 Implications for academia and research

Academically, the paper encourages greater integration between intellectual property scholarship and restitution theory. AI-IP research should not stop after classifying training as infringement or fair use. It should ask what economic benefit was obtained and how that benefit should be measured.

Future interdisciplinary research should combine legal analysis with machine-learning experiments, causal inference, and economic valuation. Ablation studies could estimate the marginal performance effect of particular data subsets.

Economists could estimate willingness-to-pay for different training uses. Computer scientists could identify memorization and influence pathways. Private-law scholars could then translate those findings into doctrinally appropriate remedies.

The study also has implications for comparative scholarship. Because copyright exceptions, restitution doctrines, and preemption rules vary across jurisdictions, an international framework should compare not only substantive training rules but also remedial architecture. The comparative literature in the uploaded set provides a foundation for such work.

12. Limitations

The study has several limitations. First, it is doctrinal and literature-based and does not estimate an actual royalty from a dataset or litigated record. The proposed valuation equation is therefore a conceptual framework rather than a tested damages model.

Second, the uploaded reference set is interdisciplinary but relatively small and does not contain a comprehensive body of restitution case law or every major AI copyright proceeding. The analysis supplements it with selected contemporary legal materials, but the paper should not be read as an exhaustive survey of all jurisdictions.

Third, legal outcomes vary substantially by jurisdiction. The proposed doctrine may interact differently with copyright preemption, unjust enrichment elements, fair use, text and data mining exceptions, database rights, and constitutional or statutory limits.

Fourth, technical evidence about model influence is still developing. A work can affect training without being memorized, and a model can reproduce an output that does not reveal a simple causal path back to one input.

Valuation must therefore tolerate uncertainty without converting uncertainty into unsupported precision.

Finally, the paper does not attempt to resolve the normative question of whether society should permit unrestricted AI learning from copyrighted works.

Its narrower purpose is to develop a remedial framework for cases in which the legal system determines that a protected input was used in circumstances that justify restitution.

13. Scope for Future Research

Future research should proceed in six directions.

First, empirical licensing research should collect actual rates for text, image, music, code, and audiovisual training licenses. A database of real transactions would improve hypothetical-license analysis.

Second, technical-economic studies should use ablation testing to estimate marginal contribution. Instead of asking whether a work is "valuable" in the abstract, researchers could estimate how removing a defined subset changes model performance.

Third, legal scholars should compare unjust enrichment, restitution, and copyright remedies across major jurisdictions, with particular attention to preemption and statutory text and data mining exceptions.

Fourth, future work should test the change of position defense through hypothetical case simulations. Variables could include the defendant's knowledge, licensing availability, expenditure level, regulatory uncertainty, and persistence of benefit.

Fifth, researchers should examine collective licensing as an alternative to individual restitution. If millions of creators have small claims, collective mechanisms may reduce transaction costs and create predictable royalty pools.

Sixth, future research should investigate whether model cards, dataset documentation, provenance logs, and machine-readable rights information can reduce the litigation cost of valuation. Technical transparency may become a precondition for meaningful private-law remedies.

14. Acknowledgements

The author acknowledges the scholarly and institutional sources that informed this conceptual study, including the reference set supplied for the research and the public legal materials concerning artificial intelligence, copyright, restitution, and related litigation. The author is solely responsible for the analysis, interpretation, and proposed framework.

References

- American Law Institute. (2011). Restatement of the law third, restitution and unjust enrichment. American Law Institute.
- Awasthy, D., Bishnoi, A., & Meena, R. (2024). AI and intellectual property law: Challenges and opportunities in digital age. International Conference on Intelligent and Innovative Practices in Engineering and Management 2024, IIPEM 2024. <https://doi.org/10.1109/IIPEM62726.2024.10925706>
- Azab, R. S., & Ismail, H. H. (2024). Extent of getting generative AI accommodated by traditional intellectual property laws (Comparative analytical study). Pakistan Journal of Life and Social Sciences, 22(2), 7990-8005. <https://doi.org/10.57239/PJLSS-2024-22.2.00603>
- Eriksen, A. (2018). Conflicting duties and restitution of the trusting relationship. Journal of Medical Ethics, 44(11), 768-773. <https://doi.org/10.1136/medethics-2017-104682>
- Fronek, P., Rotabi-Casares, K. S., & Common, R. (2021). Inter-country adoption swimming against the tide: Restitution in Samoa. Childhood, 28(4), 525-539. <https://doi.org/10.1177/09075682211063691>
- Getty Images (US), Inc. v. Stability AI Ltd., [2025] EWHC 2863 (Ch) (High Court of Justice, Business

and Property Courts, Intellectual Property List, Chancery Division).

Jammula, S., & Kanjula, M. R. (2025). AI and IPR: Balance between innovation and creativity. *Lecture Notes in Networks and Systems*, 1148, 343-350. https://doi.org/10.1007/978-981-97-8457-8_22

Kebede, N. M. (2025). Status of Ethiopian intellectual property laws on the protection of AI-generated agricultural research outputs. *Kampala International University Law Journal*, 7(1), 119-127. <https://doi.org/10.59568/KIULJ-2025-7-1-09>

Koellner, E. (2025). AI and intellectual property law in the digital age: Creating a cohesive global system for ethical innovation and societal benefit. *Jusletter IT*, 533-541. <https://doi.org/10.38023/c0b9fea3-d2ed-41b2-9f00-031b9c4ac0df>

Mahfuzzah, Z., Siagaian, P. R., Yuswar, C. P., Ramadhan, M. C., & Turekulova, D. M. (2026). Reforming the fair use doctrine in Indonesian intellectual property law: Addressing AI and digital creativity through Islamic legal perspectives. *Jurnal Ilmiah Mizani*, 13(1), 451-467. <https://doi.org/10.29300/mzn.v13i1.9117>

Mishra, P., Thakur, V. S., Shandilya, S., Singhal, A. K., & Kishore, Y. (2023). Emerging technologies and IPR: A cross-jurisdictional examination of AI and patent laws in India and the USA. *Journal of Intellectual Property Rights*, 28(5), 413-421. <https://doi.org/10.56042/JIPR.V28I5.399>

Mohanta, A., Garikapati, K., Shreya, S., Biswal, N., & Kanungo, A. (2025). Cybernetic intelligence in intellectual property law: AI-driven innovation, detection, and ethical challenges. In *Modern perspectives on artificial intelligence and law* (pp. 89-124). <https://doi.org/10.4018/979-8-3693-9576-9.ch006>

Moerland, A. (2024). Intellectual property law and AI. In *The Cambridge handbook of private law and artificial intelligence* (pp. 362-383). <https://doi.org/10.1017/9781108980197.019>

Narang, A. (2025). Exploring the intersection of AI and IPR in the context of the emerging phenomenon of deepfakes. *Journal of Intellectual Property Rights*, 30(1), 59-64. <https://doi.org/10.56042/jipr.v30i1.9984>

New York Times Co. v. Microsoft Corp., 777 F. Supp. 3d 283 (S.D.N.Y. 2025).

Niyompatama, M. J., & Lapatoura, I. (2026). Correction to: Generative AI in fashion design creation: A copyright analysis of AI-assisted designs. *Journal of Intellectual Property Law and Practice*, 21(1), 80. <https://doi.org/10.1093/jiplp/jpaf083>

Pala, G., & Desai, K. (2026). Intellectual property law and AI-generated creativity: Addressing infringement, ownership, and ethical boundaries. 2026 IEEE 3rd International Conference on Emerging Trends in Engineering and Medical Sciences, ICETEMS 2025. <https://doi.org/10.1109/ICETEMS66917.2026.11469428>

Petrella v. Metro-Goldwyn-Mayer, Inc., 572 U.S. 663 (2014).

Rassool, C., & Gibbon, V. E. (2024). Restitution versus repatriation: Terminology and concepts matter. *American Journal of Biological Anthropology*, 184(1), e24889. <https://doi.org/10.1002/ajpa.24889>

Ruback, R. B., Cares, A. C., & Hoskins, S. N. (2008). Crime victims' perceptions of restitution: The importance of payment and understanding. *Violence and Victims*, 23(6), 697-710. <https://doi.org/10.1891/0886-6708.23.6.697>

Swamy, R. N. (2021). AI-IPR intersection: An analysis of emerging issues in the Indian context. *SCMS Journal of Indian Management*, 18(3), 95-102.

U.S. Copyright Office. (2025). Copyright and artificial intelligence, Part 3: Generative AI training (pre-publication version). U.S. Copyright Office.

Xiao, Y., Tylecote, A., & Liu, J. (2013). Why not greater catch-up by Chinese firms? The impact of IPR, corporate governance and technology intensity on late-comer strategies. *Research Policy*, 42(3), 749-764. <https://doi.org/10.1016/j.respol.2012.11.005>

Yadav, T. P., Singh, H., Rajpal, G. S., Kadu, B., Kumari, N., & Singh, V. (2025). Integrating AI with patents, trademarks, copyright and traditional knowledge: Critical analysis on IPR. 2025 International Conference on Computational Engineering, Sensing Technology and Management, ICCETM 2025. <https://doi.org/10.1109/ICCETM66557.2025.11557885>

Yen, C.-C., Hsieh, P.-T., Chen, Y.-C., & Chang, C.-K. (2025). Analysis of a generative AI-based graphical learning assistance tool in IPR courses. *Proceedings - 25th IEEE International Conference on*

Advanced Learning Technologies, ICALT 2025, 226-228.
<https://doi.org/10.1109/ICALT64023.2025.00071>