Uncertain Boundaries: Multidisciplinary Approaches to Copyright Issues in Generative AI

Archer Amon[1], Zhipeng Yin[1], Zichong Wang[1], Avash Palikhe[1], Tongjia Yu[2], and Wenbin Zhang[1*]
[1] Florida International University, Miami, FL, United States
[2] Goldman Sachs, New York, NY, United States

ABSTRACT

Generative AI is becoming increasingly prevalent in creative fields, sparking urgent debates over how current copyright laws can keep pace with technological innovation. Recent controversies of AI models generating near-replicas of copyrighted material highlight the need to adapt current legal frameworks and develop technical methods to mitigate copyright infringement risks. This task requires understanding the intersection between computational concepts such as large-scale data scraping and probabilistic content generation, legal definitions of originality and fair use, and economic impacts on intellectual property (IP) rights holders. However, most existing research on copyright in AI takes a purely computer science or law-based approach, leaving a gap in coordinating these approaches that only multidisciplinary efforts can effectively address. To bridge this gap, our survey adopts a comprehensive approach synthesizing insights from law, policy, economics, and computer science. It begins by discussing the foundational goals and considerations that should be applied to copyright in generative AI, followed by methods for detecting and assessing potential violations in AI system outputs. Next, it explores various regulatory options influenced by legal, policy, and economic frameworks to manage and mitigate copyright concerns associated with generative AI and reconcile the interests of IP rights holders with that of generative AI producers. The discussion then introduces techniques to safeguard individual creative works from unauthorized replication, such as watermarking and cryptographic protections. Finally, it describes advanced training strategies designed to prevent AI models from reproducing protected content. In doing so, we highlight key opportunities for action and offer actionable strategies that creators, developers, and policymakers can use in navigating the evolving copyright landscape.

Keywords

Generative AI, Copyright law, Intellectual property, AI policy, Copyright infringement detection

1. INTRODUCTION

The growing popularity of generative AI has reignited significant concerns around intellectual property (IP) rights, especially given that many commercial AI models rely heavily on datasets freely scraped from the internet [21]. This practice has resulted in several high-profile controversies, including real-world cases such as the legal dispute involving Stability AI's Stable Diffusion model allegedly replicating artists' styles without permission [42], and lawsuits filed by artists against companies like Midjourney and DeviantArt for unauthorized use of their works [38]. Additionally, literary authors have raised objections against large language models like ChatGPT for generating content closely resembling their copyrighted texts [109, 75, 111]. Companies have responded to these incidents in various ways: some have attempted to shift blame onto users who prompt models to create potentially infringing content [80], while others invoke broad and ambiguous interpretations of "fair use" to defend their practices, further complicating accountability and enforcement [29]. This ambiguity not only weakens the legal protections available to copyright holders but also undermines public trust and the perceived integrity of generative AI systems. Although technical measures to proactively mitigate these violations exist, such as content fingerprinting, watermarking, and cryptographic methods, their adoption has been slow and inconsistent [126]. Additionally, the rapidly evolving AI technology landscape and the absence of clear legal standards exacerbate these challenges, making it difficult for stakeholders, including creators, developers, and regulators, to effectively coordinate their efforts.

Addressing these challenges requires a holistic approach that combines technical solutions with supportive policy frameworks. To navigate these complexities, this paper aims to provide a comprehensive survey of current methods for enhancing copyright compliance in generative AI, focusing on four main goals: (1) detecting copyright violations and evaluating model performance, (2) understanding how regulatory landscapes shape technical strategies for protecting individual copyrighted works from unauthorized use, (3) protecting individual copyrighted works from being used in AI systems without authorization, and (4) designing AI models in a way that prevents generation of content violating copyright. In doing so, we adopt a uniquely multidisciplinary perspective, incorporating insights from computer science, law, policy, and economics to provide a more holistic framework for addressing copyright challenges in generative AI. Within this framework, we evaluate various methods based on their effectiveness and feasibility in preventing copyright violations, while also considering their impact on the utility of generative AI models. Through this survey, we aim to provide actionable insights for the development of technical,

*Corresponding author: wenbin.zhang@fiu.edu

legal, and policy strategies, enabling creators, developers, and policymakers to navigate the complex copyright challenges introduced by generative AI.

Paper Structure. The subsequent sections of this paper are structured as follows: Section 2 provides foundational background, explaining key concepts of generative AI and the complexities of copyright law as it applies to AI-generated content. Section 3 outlines a comprehensive taxonomy categorizing multidisciplinary methods addressing copyright issues in generative AI. Section 4 presents techniques for detecting and evaluating potential copyright infringements by AI models. Section 5 reviews regulatory frameworks and policy measures designed to manage and mitigate copyright risks associated with generative AI. Sections 6 and 7 discuss technical approaches aimed at protecting copyrighted content from unauthorized AI copying, as well as advanced training methods specifically designed to prevent generative AI models from producing infringing output. Finally, Sections 8 and 9 explore available resources, ongoing challenges, and future research directions in addressing copyright concerns within the evolving landscape of generative AI.

2. BACKGROUND

2.1 Generative AI Models

Definition. Generative AI broadly refers to artificial intelligence systems capable of creating new content, such as text, images, audio, or video, by learning patterns from extensive datasets. Compared to traditional AI systems, which focus on analyzing existing data or making decisions, generative AI models produce new outputs in response to user prompts [41]. These models are typically trained on large datasets, enabling them to generate content that mimics the style and structure of the data they were trained on, although small models are also an emerging field of research [118, 46, 59, 121]. Common model types include text-to-text, text-to-image, text-to-video, and image-to-video AI, as well as multimodal systems [50, 64, 72, 122] that integrate multiple input and output forms.

Model architecture. These models are typically trained on massive amounts of data, and utilize large architectures to encode inputs into a high-dimensional latent space and use a generator model to produce varied outputs through a stochastic behavior [41]. Most generative AI systems today are built on a transformer architecture consisting of an encoder and decoder, using a multi-head self-attention mechanism to handle long-term dependencies in data by assigning higher weights to more relevant tokens [13]. Because of the large amount of resources needed, generative models are often trained through a "pre-training" paradigm, where general purpose models are later fine-tuned for specific applications, creating a more complex chain of command where issues of indirect liability are more likely to come into play [120].

2.2 Copyright Applied to Generative AI

Fair use standards Broadly speaking, copyright allows the creator of an original work to prevent others from creating and profiting from unauthorized replication, distribution, or derivation of their works [83]. When evaluating copyright violations, US law has identified four main pillars that constitute "fair use" of copyrighted material: (1) the purpose and character of the use, (2) the nature of the copyrighted work, (3) the amount and substantiality of the portion taken, and (4) the effect of the use on the potential market for the work [28]. These questions are particularly difficult to address in the context of generative AI systems as copyrighted data is often compiled into massive datasets for model training, and the impact of a particular copyrighted work on generated outputs cannot always be traced clearly.

Analyzing violations. Given the complexity of understanding fair use, analyzing potential copyright violations in generative AI often requires mixing legal and technical understandings. While there is some legal precedent for copyright cases involving search engines, web code, and APIs [85], many questions have yet to be answered in the context of AI systems. There is no clearly delineated amount for what counts as "fair use" of a copyrighted work; for example, the copyright of certain content types such as cookbooks or dictionaries is more likely to be infringed by copying even a small part, whereas this may be acceptable for larger novels [44]. AI also frequently combines both expressive and non-expressive properties [87], and analyzing compliance often requires looking at both low-level content transformations such as n-grams and verbatim copying, and higher level concepts like themes and storylines [44].

Levels of memorization. When searching for technical evidence to explain AI copyright violations, many researchers point to the phenomenon of memorization, where generative AI models reproduce near-exact copies of training data [27]. Memorization can arise due to overfitting or the underlying distribution of data [97], and leads to both direct verbatim reproduction and more subtle forms of copying. Yet despite language models frequently committing plagiarism at the paraphrase or idea-based level [56], most existing research focuses only on verbatim copying, making more subtle forms harder to assess [19]. There is thus a need to expand AI copyright research into identifying and mitigating indirect copying of protected works.

Practical challenges for the AI context. Another factor that often complicates AI copyright issues from both a legal and technical perspective is the opaque and decentralized nature of AI development. Many generative AI models are "black-box", meaning developers and researchers cannot fully understand their internal functions and trace how copyrighted content may be appearing [61, 101]. AI development today also occurs through a highly distributed supply chain [57], and lack of coordination between involved parties reduces the chance for methods to be effective at scale or resilient across later transformations of an AI model [84].

3. TAXONOMY OUTLINE

Facing dual challenges from legal and technical environments, many researchers have called for a co-evolution of technology and law so that developments in each field may support the other [44]. Considering this need for interdisciplinary work, our paper aims to combine a detailed overview of state-of-the-art technical methods for reducing copyright violations in generative AI with an analysis of the regulatory landscape which may support these methods. Specifically, we organize our analysis using a framework that maps copyright concerns across key stages of the generative AI lifecycle, integrating legal and technical work into a coherent structure. As shown in Figure 1, we categorize recent work

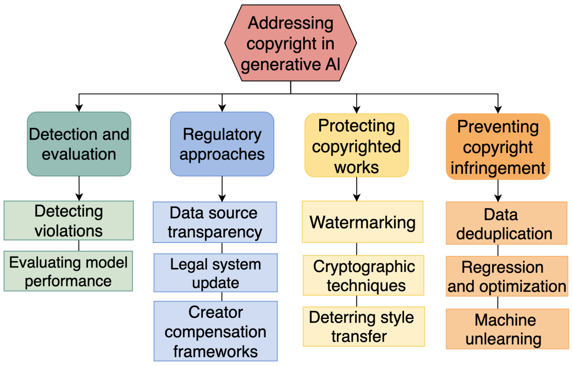

Figure 1: An overview of our proposed taxonomy.

supporting copyright compliance for generative AI into four specific focus areas, tracing copyright issues from foundational challenges to practical solutions while incorporating a diverse range of actors and development stages. We begin with **(1) detecting and assessing copyright violations**, outlining methods that identify where and how infringement occurs so that targeted solutions can be applied. Next, we discuss **(2) regulatory approaches** that can help facilitate more effective copyright protection in AI. In the following two sections, we explore technical mitigation strategies in two distinct areas: Methods for **(3) protecting copyrighted works**, which give individual creators tools to safeguard their from unauthorized use, and **(4) preventing copyright infringement**, covering model-level training strategies aimed at reducing the over likelihood of AI systems generating infringing outputs. After covering these four areas, we present resources like datasets and toolkits to support these goals, and conclude by highlighting emerging challenges in the field.

4. DETECTING AND EVALUATING VIOLATIONS

We begin by surveying methods for detecting and evaluating instances where AI models are likely to infringe on copyright. These methods are critical as they help stakeholders to identify problem sources and apply targeted remedies, assess risk and compliance, and use concrete evidence of infringement to inform legal and policy decisions. We divide the methods here into two main categories: (1) web tools for detecting unauthorized AI-enabled reproduction of copyrighted works and (2) datasets and methods for evaluating general model performance.

4.1 Detecting violations

Methods for flagging AI-enabled copyright violations draw from two primary areas of research: detecting copyright infringement and detecting AI-generated content. There is often a trade-off between the two, as methods for detecting infringement often have limited applicability once content has been altered from its original form, and AI-generated content (AIGC) detection relies on picking up special signatures from AI-generated content which may be less present in a close copy of a human creator's work.

Identifying copyright infringement. The first set of methods focuses on finding places where protected content is replicated without authorization, primarily relying on simple web tools. Reverse image search tools such as Yandex[1], Tineye[2], and Google Reverse Image Search[3], along with text search tools like Scribbr[4] and Grammarly[5] allow users to find instances where content may have been replicated. Recent models have also emerged which leverage AI to search the web for potential violations, combining BERT with DNN models to search for and flag infringing content [45]. However, many research studies on detecting copyright infringement are highly limited in scope, and may fail at detecting violations made by AI which alters the content beyond its original form [56, 19]. Special applications where the expressive and non-expressive elements of a work are closely linked, like with LLM-powered code generation, may also need unique methods for determining if a certain use counts as infringement or not [92, 119].

Identifying AI Generated Content. The next set of methods aims to detect AI generated content and trace it back to its original source to identify potential violations. A wide variety of methods have been developed to detect AI generated content across text [117, 35], image [86], video [43], and multi-modal [47, 124] content. They show that traditional methods such as logistic regression, random forest, SVM, and classifier-based methods display a fair level of accuracy at separating human from AI created content. However, not every AI generated image is one that infringes on copyright, so it is also necessary to have methods for scanning content registries to identify what images it was likely sourced from.

Fingerprinting for Comparison. Fingerprinting has been recommended for many applications to better enable identification and take-down of copyrighted material, providing a unique trace allowing digital content to be identified and attributed to its source [77]. Preetha and Bindu use a wavelet based video fingerprint to extract signatures from different images created from a video, extracting both temporal and spatial features into a compact form which can be stored in a video database and used to determine whether a query video is drawn from that database source [89]. Ning et al. develop a similar system allowing users to register content using its fingerprint rather than the original work [82]. While many different strategies for using digital fingerprints have been developed, they generally share four main characteristics: uniqueness, stability, extractability, and compactness [20]. As a result, fingerprints may provide an efficient way to identify infringing works at scale, but current research applied specifically to the context of generative AI problems is limited.

4.2 Assessing model performance

Jailbreaking methods. Several methods aim to test how easily a model can be made to generate protected content. Text-to-image models can often be prompted to generate copyrighted content even when keywords for a protected IP are replaced with a description of the image [52]. Kim et al. introduce an automatic prompt generation pipeline using LLMs to autogenerate descriptions and create revised prompts designed to induce reproduction of copyrighted content [52]. Their model can evaluate LLM copyright compli-

[1] `https://yandex.com/images`
[2] `https://tineye.com`
[3] `https://google.com/images`
[4] `https://scribbr.com/plagiarism-checker/`
[5] `https://grammarly.com`

ance without requiring access to any internal weights, allowing it to function on black-box systems. The prompts generated through their pipeline successfully jailbreak ChatGPT to generate copyrighted content 76% of the time, with an 11% block rate. Test cases can be designed for exploring a model's tendency to engage in specific copying behaviors, like verbatim vs indirect copying.

AI dataset probing. Similar methods aim to reduce the black-box nature of AI models by probing into whether certain copyrighted content is included in their training dataset. For this purpose, Duarte et al. introduce DE-COP, a benchmarking method designed to determine if a piece of copyrighted content is included in a training dataset [33]. Their approach is to probe an LLM with multiple-choice questions to complete a passage of a suspected target book, with options including both verbatim text and paraphrases. The LLMs that were tested showed substantially different performance on these tests depending on whether or not a book is in their training dataset, with the DE-COP method showing a 72% accuracy for detecting suspect books, compared to about 4% for previous models. However, information about whether or not something is in a model's dataset is not enough to prove a violation of copyright. This method should thus be combined with other strategies which examine model outputs to find violations.

Model-level risk quantification. To fulfill the need for a broad metric for evaluating copyright risk, Zhou et al. introduce CopyScope, a framework for quantifying infringement at the model level by using Fréchet Inception Distance (FID) to capture image similarity in the way that most closely mirrors human perception [135]. Using an ensemble based approach of trying different combinations of model sub-components, Zhou et al. use FID-Shapley values to calculate how much each component contributes to the final image's likeness to the training data, identifying which models are most likely to cause infringement issues and should be identified as targets for extra attention.

Regulatory Options for AI Copyright Issues. Copyright challenges in generative AI cannot be solved through advancements in computing alone; disciplines such as law and policy play a central role in shaping discussions around copyright and promoting specific values in design. Legal and policy standards not only work to clarify where violations are present [96], but also frequently serve as a prerequisite for implementing technical safeguards on a wide scale. For instance, several mitigation techniques depend on a well-documented data life cycle, requiring clear transparency about where data is sourced and how it is used [125], which better regulations can help coordinate. These frameworks also work to shape broader industry practices around data collection, creator consent, and compensation, fundamentally affecting what copyright issues emerge. Taking a uniquely interdisciplinary lens, this section explores key regulatory approaches that interact with AI copyright mitigation and enforcement, integrating perspectives from law, economics, and more.

4.3 Data source transparency

Voluntary measures and Industry standards. Transparency in AI training data is a foundational principle for advancing copyright protection in AI [95]. Transparency measures empower copyright holders to monitor and protect against misuse of their works, and promote better principles and processes for designing AI models. Inspired by similar "privacy-by-design" principles, several experts advocate for adopting measures that actively promote transparency throughout the AI development process [37]. Groups like the Coalition for Content Provenance and Authenticity and the Content Authenticity Initiative have frequently contributed to these discussions, developing technical standards for tracking data origins with features that allow rights holders to specify whether training is allowed [95]. However, the adoption of such measures remains largely dependent on the voluntary actions of individual companies, which limits their widespread effectiveness.

Standardized labels. Another approach calls for standardized labels to harmonize transparency efforts. These AI "nutrition labels", inspired by similar efforts in the food agency, call for disclosing information on an AI system's data sources, potential risks, and limitations [24]. While these models have been adapted to support labeling of generative AI [106], more work is necessary to explore how they can be applied to the specific context of copyright. A similar proposed tool is "data cards", or structured summaries that provide essential facts about datasets and explain the rationale behind decisions made in creating them [90]. Professional organizations and regulatory authorities could adopt policies to promote the adoption of these measures or integrate them within other documentation and reporting requirements.

AI system audits. Independent audits of AI systems may be proposed as part of a broader licensing framework or an industry standard certification separate from statutory policy. Auditing for copyright is typically done on model data sets to understand the data collection process and nature of the dataset [74]. However, these "data audits" typically focus on general industry data practices rather than holding dataset creators accountable, and are often divorced from other model-level audits, creating a disconnect between data understanding and deployment regulations [10]. Noting the fragmentation of the current AI audit landscape, Manheim et al. recommend creating AI audit standards boards keep auditing standards up to date with current advancements in AI and clarify which standards are suited for specific domains and applications [74]. This work also has a technical dimension, as decisions need to be made about how much access should be given to auditors. In this scope, Casper et al. propose expanding beyond traditional "black-box" access audits which only observe AI system outputs. To make audits more robust, they suggest including "white-box" access to information about model weights and inner workings, and "outside-the-box" access that provides information about training and deployment processes [14].

4.4 Copyright legal system updates

Considerations for legal change. As generative AI systems continue to increase their capabilities and become more widespread, existing doctrines of copyright law become increasingly unsuited for resolving disputes about AI [2, 53]. This creates a need for broader structural change, which is likely to be implemented through a gradual system of legislative action and interpreted rulings rather than a single policy. Many ideas revolve around strengthening protections for creators [40] or updating the copyright law system to address specific questions around data scraping and ownership in the digital age [21]. In these discussions, policymakers often need to manage the trade-off between transparency and feasibility of implementation. Beyond facing criticisms of stifling innovation which may turn government support against these measures [16], creating too harsh of standards around training data may significantly limit the amount available, and limited data is more likely to reinforce biases and stereotypes [40].

Opt-out policies. Opt-out provisions aim to reconcile issues of consent and copyright by allowing creators to opt out of having their work used for AI training [127]. However, it is often unclear how they are meant to function in practice [137]. Existing opt-out methods are often difficult and seen by creators as largely a PR stunt [87], causing a need for more research to address the fragmentation of opt-out policies in the status quo [51]. Pasquale and Sun (2024) propose a mandatory opt-out mechanism requiring AI developers to remove works from their databases upon request if infringement has been documented, and take action action to prevent the infringing content from being made again [87]. While this method creates more transparency by requiring developers to properly manage their datasets and confirm upon request if a source has been used [87], it functions more as a post-hoc right to remove copyrighted content after violation has occurred, leaving a remaining need for more preventative opt-out strategies. Many rights holders may also be unaware of unauthorized AI reproductions of their work or lack information about opt-out procedures, creating barriers to establishing a system of full consent.

Content management registries. To help streamline consent management for new AI training, Balan et al. introduce a decentralized registry for content creators to assert their right to opt in or out of AI training, combining distributed ledger technology with visual fingerprinting [7]. Their prototype model prevents a scalable way to trace generative AI training data to determine consent, similar to many of the methods discussed in the protection section. They propose such a registry also be used to track information for compensating creators that opt in to AI training.

Compensation frameworks. Many new proposals have identified that reworking current compensation frameworks may help reconcile the interests of creators and AI producers. Compensation not only ensures individual creators are treated fairly, but also establishes better data provenance tracking and reduces power differentials between AI developers and creators [87] when frameworks are well designed. Compensation systems could be arranged through either new statutory [40] or existing contract-based structures [114], working with collective rights management organizations to streamline negotiating and distributing remuneration. These models generally fall within three main categories, as illustrated in Table 1.

Table 1: Compensation frameworks for AI training data.

Compensation Model	Pays Based On
Pay to Train	Percentage of training data
Pay to Train and Inspire	Contribution to generated outputs
AI Royalties	Negotiated IP partner framework

Pay-to-train compensation. The pay-to-train model involves rewarding IP holders based on the percentage of their contribution to a dataset. Such a model could use increased dataset transparency requirements to calculate payments based on the amount of copyrighted material within AI datasets and the monetary value attributed to its use for AI training. These payments could be afforded to individual creators, or distributed into collective funds to support creators. However, dataset sources are not always well-documented, and payments may be minimal for individual creators beyond famous authors or artists whose work is more likely to comprise a large portion [26].

Pay-to-train-and-inspire compensation. This model works backwards to understand which items in a model's training data likely inspired a particular output, and distribute compensation accordingly [34]. Wang et al. combine probabilistic methods with Shapley value interpretability techniques under a game theory model to establish a base framework for compensation in this way and show its viability through practical experimentation with common data sources [114]. This method works best for AI models trained on limited data with copyright split between a smaller amount of owners. In some cases, it may be more practical to estimate aggregate payoffs for copyright owners across all generated outputs rather than tracing each output back to its source for compensation.

AI royalties. The last model, AI royalties, aims to create collaborative partnerships between IP rights holders and AI companies for compensation based on the market usage and value of their systems [34]. This model could be implemented using existing contract law systems, recognizing the exclusive right of rights holders to commercial use of substantially similar copyright or trademark uses, and granting the AI company the right to use its IP to create outputs under defined use guidelines, sharing in a portion of the overall revenue generated by the system. Ducru et al. argue that this model is the best suited to mutually benefit the interests of AI creators and IP rights holders, and eliminates the need for case-by-case determinations by allowing a broader, predefined agreement that covers all outputs generated by the system [34].

5. PROTECTING COPYRIGHTED WORKS FROM AI

We now introduce technical methods for mitigating copyright infringement in generative AI, starting with methods for protecting individual copyrighted works. These methods are primarily applied at the data collection and preprocessing level, and may be used by creators to protect their works as well as developers seeking to preserve attribution and traceability of training data used. While many protection techniques are frequently used in conjunction, we divide them into three primary categories: (1) watermarking, (2) cryptographic methods, and (3) deterring style transfer.

5.1 Watermarking

Watermarking methods serve to embed an imperceptible identification layer onto images or videos to identify ownership over the content. They can be embedded directly into inputs such as images, or their intermediate representations such as encoder/decoder feature maps [32]. We divide this section into two main approaches based on the methods used to apply the watermark: attention-based mechanisms and cryptographic methods.

Attention based mechanisms. Attention-based techniques leverage AI models' tendency to focus on specific image regions to create watermarks which remain present after a high degree of manipulation. For example, Zhang et al. introduce deep learning-based video watermarks using a custom attention mechanism designed to recover a 32-bit watermark with 99% accuracy post-transformation [128]. Similar methods can be used for image content by applying watermark algorithms based on Tchebichef moments which describe the spatial distribution of an image's intensity [36, 81, 116]. By partitioning the host image into non-overlapping blocks and calculating Tchebichef moments for each block, Ernawan and Kabir are able to prioritize watermark embedding in areas of lower visual entropy or complexity. As a result, their method demonstrates resilience against noise disruptions and JPEG2000 compression, while preserving a high peak signal-to-noise ratio (PSNR) of approximately 40 dB. While attention mechanisms like these are more resilient to various transformations, testing may be needed to ensure watermarks maintain high enough 'pattern uniformity' to be learned and reproduced by generative diffusion models [30].

Cryptography-based watermarks. In order to further increase traceability and security, several techniques for watermarking incorporate various cryptographic methods. For video content, Zheng et al. incorporate a double-layered watermarking technique combined with blockchain technology [133, 23]. The double watermarking technique embeds both robust and fragile watermarks into the video content, where the robust watermark ensures copyright protection, while the fragile watermark facilitates tamper detection and integrity verification. By integrating double watermarking with decentralization, the combined method achieved a 90% precision rate and a 95% recall rate in detecting tampered parts of watermarked videos against adversarial attacks. Sanivarapu et al. present a similarly robust digital watermarking system using cryptographic techniques to protect images from copyright infringement [99]. The system first embeds a QR code watermark using Discrete Wavelet Transform (DWT) [108]. It then layers on the transformed matrix with Singular Value Decomposition (SVD), and uses the RSA algorithm for watermark embedding, where the coefficients of DWT are modified using secret keys to embed the watermark, increasing the security and encryption of the image. While cryptographic watermarks provide better protection against specific threats like forgery and tampering, they also come with high computational overhead, and may be more perceptible than attention-based watermarks which can be optimized for processing through various generative models.

5.2 Cryptographic methods

Beyond applications in watermarks, advanced cryptographic strategies frequently aid in tracing and ownership verification of copyrighted content [31, 103]. One main type is digital signatures and hashing, which aim to authenticate content by providing a record of ownership and evidence of alterations or tampering that may be done to content, such as removing watermarks. Chain and Kuo (2013) use chaotic map transformations for generating digital signatures, using their unpredicable pattens to produce unique signatures from text for later verification [15]. By contrast, the Elliptic Curve Digital Signature Algorithm (ECDSA) used by Chandrashekhara et al. combines elliptic curve cryptography with digital signatures [17], using a SHA-256 hashing algorithm to generate a public key from a private one to authenticate the signature [5]. Commonly implemented along with hashing, blockchain methods similarly work to safeguard various works from infringement by recording transactions in cryptographically linked blocks, which are harder to tamper with than traditionally protected methods [48, 91]. The decentralized nature of blockchain enhances security by removing central points of control and creating a publicly accessible record of ownership. This can be particularly helpful for tracking data and model copyright status across each stage in the AI development life cycle, such as by integrating blockchain with contract management software for digital content [98].

5.3 Deterring style transfer

For image generation models, style transfer refers to the ability to produce the same content of a target image across a variety of artistic styles [39], which can lead to non-direct reproduction of copyrighted works. While style transfer is often explored as an intentional goal, such as for filling missing frames in an animation [12], unintentional replication of an artist's style can increase copyright risks. Adversarial layers can help protect artistic work from being copied by adding a minimally perceptible layer that distorts the ability of an AI model to recognize it as a normal image and prevents the creation of derivative works [63, 102, 134]. For example, Li et al. further develop this strategy by using a momentum-based ensemble method to enhance the ability for these protections to be generalized across AI models [62]. By altering the intermediate style representation of an image across multiple encoders and combining them through a softmax regression gradient, their method of "Neural Style Protection" can protect style transfer across both known and unknown models, while end-to-end and random noise baseline methods offer minimal protection.

6. PREVENTING COPYRIGHT INFRINGEMENT

Moving on from methods to protect individual works, this section discusses prevention methods that AI developers can use to reduce the overall risk of a model generating reproductions of copyrighted works — a significant concern for developers seeking to avoid reputation loss and legal costs [73]. These methods focus on improving the general behavior of a model, and can be useful to apply at later stages in the AI development life cycle, including fine-tuning existing models. With many current datasets lacking proper content attribution or copyright information, these solutions may serve as a short-term way to enhance copyright compliance without needing to fully construct new large-scale datasets.

6.1 Data de-duplication

De-duplication refers to the process of removing redundant data within a model's training dataset [22], and has commonly been researched for the purposes of reducing data storage space [136] and improving query performance [93]. De-duplication methods may apply fingerprints or hash values to divided data chunks in order to reduce the computational resources needed to check for duplicates in a big data context [113]. While existing research primarily focuses on the benefits of de-duplication for structured data and predictive models, Lee et al. show that de-duplication in generative AI data can improve overall performance and reduce verbatim copying in generative AI models [58]. Particularly when combined with hashing or blockchain methods, de-duplication is effective at creating cleaner datasets for training generative AI and preventing copyright issues arising from the overuse of any particular material [79, 4]. De-duplication is especially well suited for cleaning "noisy" datasets, such as data scraped from social media [60], as it can simultaneously reduce training time while improving performance and language understanding for LLMs.

6.2 Regression and optimization

In some cases, altering modeling and optimization choices can help prevent a model from reproducing copyrighted data, as Chu et al. "copyright regression" approach demonstrates [25]. By adding an inverse term to the training objective that discourages the model from generating outputs that match copyrighted data and demonstrating mathematically that it can be applied successfully on the softmax function, Chu et al. successfully create a method which helps balance between model performance and copyright protection. However, this method relies on knowledge about which training samples are copyrighted, which is not provided for most existing LLM datasets [94]. Kim et al. highlight the difficulty of applying an end-filter to remove copyrighted content, noting that no open source models are currently able to identify and differentiate copyrighted content, particularly as many datasets lack information about copyright attribution [52]. While target datasets of copyrighted content have been created for the purposes of evaluating whether a model will produce them [52], filtering out all potential violations will require larger and more comprehensive datasets that contain clear information about attribution and copyright status of works.

6.3 Machine unlearning

Unlearning techniques. For cases where an AI model has already been trained on copyrighted data, "machine unlearning" methods can be used to remove a group of samples from its training data, allowing it to act as though it has never seen the data before. This can be especially helpful from a legal perspective, allowing rules such as the GDPR's "right to be forgotten" [18] to be applied to machine learning models. Multiple methods have been developed to selectively remove content while preserving general features a model may have learned from the content. Zhang et al. introduce two different methods for unlearning: Elastic Weight Consolidation (EWC), which adds a constraint to the model's loss function to neutralize the influence of "removed" data, and Decreasing Moment Matching (DMM), which approximates the model's knowledge as a Gaussian distribution and selectively matches moments to similarly reduce reliance on data [131].

Generative AI specific strategies. While earlier research mostly focused on unlearning for classification models, newer studies have explored unlearning for generative AI [68]. Liu et al. conduct a comprehensive survey of machine unlearning for generative models, categorizing them into two main approaches: parameter optimization, which focuses on adjusting model parameters linked with target removal data, and in-context unlearning, which alters input prompts through an API aiming to steer the model away from the "unlearned" content [69].

Knowledge entanglement. One of the largest problems identified in both method types is knowledge entanglement, where data requiring removal is often closely tied to a model's knowledge of certain topics, causing a trade-off between model performance and compliance with the unlearning goal. To solve this issue, Tang et al. introduce a three-component framework to allow models to unlearn certain data without sacrificing their expressive capabilities [110]. The three components include a Knowledge Unlearning Induction module which trains the model to forget specific sequences, a Contrastive Learning Enhancement module to maintain overall performance, and an Iterative Unlearning Refinement module to iteratively update the target data for unlearning, preventing a drastic shift to the model from ocurring. Similar strategies include adopting "un-unlearning" techniques to reintroduce unlearned data in context. This can be important to prevent the unlearned data from being recreated if introduced later as input to the system [105] or retained by association with similar concepts [52]. For example, Van Gogh's *Starry Night* may still be recreated after a model attempts to unlearn "Van Gogh" as it has high correlations with the concepts *star* and *night*. For this reason, unlearning strategies will often retain a copy of the unlearned data to serve as a reference for evaluating model outputs without being used to train the generative model [110].

7. RESOURCES

7.1 Datasets

While no comprehensive dataset of all copyrighted works has been developed [52], benchmark datasets of both AI-generated and human content serve as critical resources for testing and evaluating generative model performance on copyright issues. In this section, we present some of the most commonly used datasets for AI copyright research along with a discussion of their use potential.

Human content datasets. Repositories of human-generated content may serve as a benchmark for evaluating AI models by testing if an AI model completes a passage of protected text [56] or comparing to AI-generated content for improving detection algorithms [124, 43, 86]. Popular datasets for imagery include COCO [65, 66], Flickr30K [123], and OpenImages [55], collectively offering over 9 million images with annotations describing the objects included and overall scene. Other datasets such as the the Metropolitan Museum of Art Open Access collection [6] and WikiArt [115] focus primarily on artistic contributions, which can be helpful for exploring style transfer or searching for matches between AI generations and human-created artwork. Text datasets include OpenSubtitles [67], BookCorpus [8], the WikiText collection from wikipedia articles [78], and JSTOR's library of journal articles [49] may be similarly used to scan for potential memorization, or compare human with AI-generated content.

Combined human-AI datasets. A separate category of datasets combines human-generated with AI-generated works across various art styles and writing subjects [107, 70, 9]. These datasets are primarily valuable for training models to recognize text or image pairs, which allows for both detection and mitigation of copyright violations [54, 100]. Some datasets in this category are designed for the specific purpose of comparing human copyrighted works to altered digital versions. Aboutalebi et al. introduce the Deepfake Art Challenge dataset, consisting of over 32,000 image pairs that are either forgeries, adversarially contaminated, or not [1]. The selected methods used to modify images include inpainting, style transfer, adversarial data poisoning, and cutmix, representing popular copyright violation types such as modifying painting styles or using partial image data.

Feature / Artifact based datasets. Beyond providing overall examples of human and AI creations, another category of datasets provides annotations over content to highlight specific 'artifacts' in visual media, which can be used for both detecting and preventing violations. These 'perceptual artifacts' capture subtle distortions or irregularities within an outputted image or video, which can be used for detecting the presence of AI in an image or "inpainting" to regenerate areas with potentially unwanted artifacts [32]. Datasets such as PAL4Inpaint [130] and PAL4VST [129] can be used to localize artifacts within images which may show copyright violations, such as a distorted logo within an image. However, more research is necessary to understand how methods designed for deepfake detection and authentication may be applicable to copyright [32].

7.2 Toolkits

Technical resources. While most methods for detecting and mitigating AI copyright violations have yet to be applied extensively outside of a limited resource context, a few tools have been created for public use by AI developers and creators looking to protect their work. Copyright Catcher, developed by Patronus AI researchers, is the first API which aims to test for potential copyright violations in LLMs [88]. Its method closely resembles those discussed in the evaluation methods subsection [52, 33], using a dataset sampled from popular books on Goodreads to test if a model will complete the beginning of a prompt given from protected text. While this is a good start, it may not capture more complex copying behavior such as paraphrasing [56], and may have limited performance for models trained on other types of media. Most recently, researchers at Imperial College London developed a "copyright traps" system where creators can include ficticious entities in their content to detect where LLMs may be using their content, capturing behavior beyond pure memorization [76, 11]. Resources have also been created to allow for better tracing and analysis of which data is typically used to train AI models. The Data Provenance Explorer, which allows practitioners to trace and filter on data provenance for the most popular open source data collections, is one such tool created for this purpose [71]. The interactive UI includes a tool to explore over 1800 popular datasets, viewing information such as their licenses, sources, and creators, along with additional visualizations about website protocols for scraping and AI training. However, this resource only focuses on text datasets, and more work is needed to establish provenance record tracing systems for other types of data.

Governance toolkits. Several frameworks and toolkits have been developed for analyzing the ethical impacts of AI algorithms, such as the AI and data protection risk toolkit developed by the UK Information Commissioner's Office [112] and the Ethics & Algorithms Toolkit developed in collaboration with the Center for Government Excellence at John Hopkins University [3]. However, no large-scale policy toolkits have been developed with the specific goal of increasing copyright compliance in AI. As legal and regulatory discussion around copyright in AI continues to evolve, one place where toolkits may play a helpful role is by enhancing public understanding of AI models in relation to copyright. Working to promote participatory approaches to AI and machine learning, Shen et al. established the Model Card Authoring Toolkit, a toolkit combining technical interfaces and collective decision-making protocols to empower community members to understand and make decisions about AI models in line with their collective values [104]. Applied to the issue of copyright, focus groups including authors, artists, and other rights holders could benefit from these tool kits to help understand and articulate their collective preferences surrounding AI models.

8. CHALLENGES AND FUTURE DIRECTIONS

As researchers continue to work to understand and improve the behavior of generative AI models, several directions offer potential for discovering new information about identifying and mitigating copyright violations in generative AI. On the side of detection and evaluation, a persistent challenge is *identifying indirect and ambiguous forms of copying* like paraphrasing or near-duplicate code generation [19, 34, 119]. More refined benchmarks and detection tools are needed to improve copyright enforcement for these cases. In a similar vein, *building copyright-specific datasets* is necessary to facilitate better research and detection of copyright issues at scale, track content ownership, and help facilitate certain technical mitigation strategies. For technical methods protecting copyrighted works, *strengthening watermark resilience* remains a major challenge, as there is often a trade-off between imperceptibility and security, and common modifications like compression or noise can easily weaken or remove existing markers [132]. On the side of preventing copyright infringement, more research is needed to understand how AI model parameters can reduce copyright violations while *retaining maximum utility* and *propagating to downstream models* which may be fine-tuned for different goals. Furthermore, there is a need to explore how domain-specific methods can be *transferrable across multiple types of content*, as many current strategies are limited in scope, which may hamper coordination and research. Lastly, each of these efforts continually adapt as the legal system around generative AI evolves. Stronger *consensus on the bounds of fair use* can help strengthen copyright enforcement and bridge the interests of IP rights holders and AI developers.

9. CONCLUSION

As generative AI systems trained on copyrighted data continue to proliferate in the absence of clear legal frameworks, a combination of technical and policy measures is necessary to prevent copyright infringement and ensure that generative AI is developed with fair use principles in mind. Creators, developers, policymakers, and other stakeholders should take action to carefully assess their current compliance with copyright standards and establish stronger systems of oversight, mitigation, and accountability for copyright harms caused by AI. In many cases, this will involve a continued dialogue between rights holders and AI providers to understand their demands and increase transparency surrounding technical design choices. Researchers should also expand efforts to detect violations, protect creative works, and improve the copyright performance of AI models to address the challenges previously mentioned. Through this integrated approach, creators, developers, and policymakers can collaborate to promote an AI ecosystem designed to support fair use and foster human creativity.

Acknowledgements

This work was supported in part by the National Science Foundation (NSF) under Grant No. 2404039.

10. REFERENCES

[1] Hossein Aboutalebi et al. *DeepfakeArt Challenge: A Benchmark Dataset for Generative AI Art Forgery and Data Poisoning Detection*. 2024. arXiv: 2306.01272.

[2] Jacob Alhadeff, Cooper Cuene, and Max Del Real. "Limits of algorithmic fair use". In: *Washington Journal of Law, Technology & Arts* 19.1 (2024), pp. 1–53.

[3] David Anderson et al. *Ethics & Algorithms Toolkit*. Center for Government Excellence at Johns Hopkins University, 2018.

[4] R. Aparna, Roopa G. Kulkarni, and Shilpa Chaudhari. "Secure Deduplication for Images using Blockchain". In: *IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*. 2020.

[5] Andrew W. Appel. "Verification of a Cryptographic Primitive: SHA-256". In: *ACM Transactions on Programming Languages and Systems* (2015).

[6] The Metropolitan Museum of Art. *The Met: Open Access*.

[7] Kar Balan et al. "DECORAIT - DECentralized Opt-in/out Registry for AI Training". In: *Proceedings of the 20th ACM SIGGRAPH European Conference on Visual Media Production*. CVMP '23. 2023.

[8] John Bandy and Nicholas Vincent. 2021.

[9] Bhat, Aaditya. *GPT-wiki-intro (Revision 0e458f5)*. 2023.

[10] Abeba Birhane et al. *AI auditing: The Broken Bus on the Road to AI Accountability*. 2024.

[11] Caroline Brogan and Gemma Ralton. *Hidden Data Could Reveal if an AI Model Was Trained on Copyrighted Material*. July 2024.

[12] Qiang Cai et al. "Image neural style transfer: A review". In: *Computers and Electrical Engineering* (May 2023).

[13] Yihan Cao et al. *A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT*. 2023.

[14] Stephen Casper et al. "Black-Box Access is Insufficient for Rigorous AI Audits". In: *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '24. June 2024.

[15] Kai Chain and Wen-Chung Kuo. "A new digital signature scheme based on chaotic maps". In: *Nonlinear Dynamics* 74 (Dec. 2013).

[16] Keith Jin Deng Chan, Gleb Papyshev, and Masaru Yarime. "Balancing the Tradeoff between Regulation and Innovation for Artificial Intelligence: An Analysis of Top-down Command and Control and Bottom-up Self-Regulatory Approaches". In: *Technology in Society* (Oct. 2024).

[17] J. Chandrashekhara et al. "A COMPREHENSIVE STUDY ON DIGITAL SIGNATURE". In: *International Journal of Innovative Research in Computer Science & Technology* 9.3 (May 2021).

[18] Cheng-chi Chang. "When AI Remembers Too Much: Reinventing the Right to Be Forgotten for the Generative Age". In: *Washington Journal of Law, Technology & Arts* 19 (2024).

[19] Tong Chen et al. "CopyBench: Measuring Literal and Non-Literal Reproduction of Copyright-Protected Text in Language Model Generation". In: (2024).

[20] Wendi Chen, Wensheng Gan, and Philip S. Yu. *Digital Fingerprinting on Multimedia: A Survey*. 2024. arXiv: 2408.14155.

[21] Simon Chesterman. "Good models borrow, great models steal: intellectual property rights and generative AI". In: *Policy and Society* (2024), puae006.

[22] Nipun Chhabra and Manju Bala. "A Comparative Study of Data Deduplication Strategies". In: *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*. 2018, pp. 68–72.

[23] Sribala Vidyadhari Chinta et al. "AI-driven healthcare: Fairness in AI healthcare: A survey". In: *PLOS Digital Health* 4.5 (2025), e0000864.

[24] Kasia S. Chmielinski et al. *The Dataset Nutrition Label (2nd Gen): Leveraging Context to Mitigate Harms in Artificial Intelligence*. 2022.

[25] Timothy Chu, Zhao Song, and Chiwun Yang. *How to Protect Copyright Data in Optimization of Large Language Models?* 2023.

[26] Amanda Coelho Della Giustina. "Fair Compensation for Copyrighted Data Used in AI Training". MA thesis. Tilburg Institute for Law, Technology, and Society (TILT), 2024.

[27] A. Feder Cooper and James Grimmelmann. *The Files are in the Computer: Copyright, Memorization, and Generative AI*. 2024. arXiv: 2404.12590.

[28] Copyright Advisory Services. *Fair Use*.

[29] Thomas F Cotter. "Fair use and copyright overenforcement". In: *Iowa L. Rev.* 93 (2007), p. 1271.

[30] Yingqian Cui et al. "DiffusionShield: A Watermark for Copyright Protection Against Generative Diffusion Models". In: *NeurIPS 2023 Workshop on Diffusion Models* (2023).

[31] Saad Mohamed Darwish, Mona Mahamod Abu-Deif, and Saleh Mesbah Elkaffas. “Blockchain for video watermarking: An enhanced copyright protection approach for video forensics based on perceptual hash function”. In: *PLOS ONE* (Oct. 2024), e0308451.

[32] Jingyi Deng et al. *A Survey of Defenses against AI-generated Visual Media: Detection, Disruption, and Authentication*. 2024. arXiv: 2407.10575.

[33] André V. Duarte et al. *DE-COP: Detecting Copyrighted Content in Language Models Training Data*. 2024.

[34] Pablo Ducru et al. *AI Royalties – an IP Framework to Compensate Artists & IP Holders for AI-Generated Content*. 2024. arXiv: 2406.11857.

[35] Ahmed M. Elkhatat, Khaled Elsaid, and Saeed Almeer. “Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text”. In: *International Journal for Educational Integrity* (Sept. 2023).

[36] Ferda Ernawan and Muhammad Nomani Kabir. “An Improved Watermarking Technique for Copyright Protection Based on Tchebichef Moments”. In: *IEEE Access* 7 (2019), pp. 151985–152003.

[37] Heike Felzmann et al. “Towards Transparency by Design for Artificial Intelligence”. In: *Science and Engineering Ethics* (Nov. 2020), pp. 3333–3361.

[38] Giancarlo Frosio. “Generative AI in court”. In: *Recreating Creativity, Reinventing Inventiveness*. Routledge, 2024, pp. 3–44.

[39] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. “Image Style Transfer Using Convolutional Neural Networks”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 2414–2423.

[40] Christophe Geiger and Vincenzo Iaia. “The forgotten creator: Towards a statutory remuneration right for machine learning of generative AI”. In: *Computer Law & Security Review* (Apr. 2024), p. 105925.

[41] Roberto Gozalo-Brizuela and Eduardo C. Garrido-Merchan. *ChatGPT is not all you need. A State of the Art Review of large Generative AI models*. 2023.

[42] Carol Mullins Hayes. “Generative artificial intelligence and copyright: Both sides of the black box”. In: *Available at SSRN 4517799* (2023).

[43] Peisong He et al. *Exposing AI-generated Videos: A Benchmark Dataset and a Local-and-Global Temporal Defect Based Detection Method*. 2024. arXiv: 2405.04133.

[44] Peter Henderson et al. “Foundation Models and Fair Use”. In: *Journal of Machine Learning Research* 24.400 (2023).

[45] Aldo Hernandez-Suarez et al. “Methodological Approach for Identifying Websites with Infringing Content via Text Transformers and Dense Neural Networks”. In: *Future Internet* (Dec. 2023).

[46] Cheng-Yu Hsieh et al. *Distilling Step-by-Step! Outperforming Larger Language Models with Less Training Data and Smaller Model Sizes*. 2023.

[47] Liting Huang et al. *RU-AI: A Large Multimodal Dataset for Machine Generated Content Detection*. 2024. arXiv: 2406.04906.

[48] Tao Jiang et al. “Research on the Application of Blockchain in Copyright Protection”. In: *2020 International Conference on Culture-oriented Science & Technology (ICCST)*. 2020, pp. 616–621.

[49] JSTOR. *Text-mining support*.

[50] Joohoon Kang and Youngjoo Yi. “Beyond ChatGPT: Multimodal generative AI for L2 writers”. In: *Journal of Second Language Writing* (Dec. 2023), p. 101070.

[51] Paul Keller and Zuzanna Warso. *Defining Best Practices for Opting Out of ML Training*. Tech. rep. Open Future, Sept. 2023.

[52] Minseon Kim et al. *Automatic Jailbreaking of the Text-to-Image Generative AI Systems*. 2024. arXiv: 2405.16567.

[53] John T. Kivus. “Generative AI and Copyright Law: A Misalignment That Could Lead to the Privatization of Copyright Enforcement”. In: *North Carolina Journal of Law & Technology* 25.3 (Apr. 2024), pp. 447–494.

[54] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Communications of the ACM* (May 2012).

[55] Alina Kuznetsova et al. “The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale”. In: *CoRR* abs/1811.00982 (2018).

[56] Jooyoung Lee et al. “Do Language Models Plagiarize?” In: *WWW ’23: Proceedings of the ACM Web Conference*. Apr. 2023.

[57] Katherine Lee, A. Feder Cooper, and James Grimmelmann. *Talkin’ ’Bout AI Generation: Copyright and the Generative-AI Supply Chain*. 2024.

[58] Katherine Lee et al. *Deduplicating Training Data Makes Language Models Better*. 2022.

[59] Liunian Harold Li et al. *Symbolic Chain-of-Thought Distillation: Small Models Can Also "Think" Step-by-Step*. 2024.

[60] Xianming Li and Jing Li. *Generative Deduplication For Socia Media Data Selection*. 2024.

[61] Yangzi Li. “Does black-box machine learning shift the US fair use doctrine?” In: *Journal of Intellectual Property Law & Practice* (Sept. 2021), pp. 1175–1185.

[62] Yaxin Li et al. “Neural Style Protection: Counteracting Unauthorized Neural Style Transfer”. In: *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Jan. 2024.

[63] Chumeng Liang et al. “Adversarial Example Does Good: Preventing Painting Imitation from Diffusion Models via Adversarial Examples”. In: (2023).

[64] Paul Pu Liang, Amir Zadeh, and Louis-Philippe Morency. “Foundations & Trends in Multimodal Machine Learning: Principles, Challenges, and Open Questions”. In: *ACM Comput. Surv.* (June 2024).

[65] Tsung-Yi Lin et al. *Common Objects in Context.*

[66] Tsung-Yi Lin et al. "Microsoft COCO: Common Objects in Context". In: (2015). arXiv: 1405.0312.

[67] Pierre Lison and Jörg Tiedemann. May 2021.

[68] Sijia Liu et al. *Rethinking Machine Unlearning for Large Language Models.* 2024. arXiv: 2402.08787.

[69] Zheyuan Liu et al. *Machine Unlearning in Generative AI: A Survey.* 2024. arXiv: 2407.20516.

[70] Vijini Liyanage, Davide Buscaldi, and Adeline Nazarenko. "A Benchmark Corpus for the Detection of Automatically Generated Text in Academic Publications". In: *13th Conference on Language Resources and Evaluation (LREC 2022).* 2022, pp. 4692–4700.

[71] Shayne Longpre et al. *The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI.* 2023.

[72] Ming Y. Lu et al. "A multimodal generative AI copilot for human pathology". In: *Nature* (June 2024), pp. 466–473.

[73] Nicola Lucchi. "ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems". In: *European Journal of Risk Regulation* (Aug. 2023), pp. 1–23.

[74] David Manheim et al. *The Necessity of AI Audit Standards Boards.* 2024.

[75] Gary Marcus and Reid Southen. *Generative AI Has a Visual Plagiarism Problem - IEEE Spectrum.* Jan. 2024.

[76] Matthieu Meeus et al. *Copyright Traps for Large Language Models.* 2024. arXiv: 2402.09363.

[77] David Megías, Minoru Kuribayashi, and Amna Qureshi. "Survey on Decentralized Fingerprinting Solutions: Copyright Protection through Piracy Tracing". In: *Computers* 9.2 (Apr. 2020), pp. 1–19.

[78] Stephen Merity et al. *WikiText-103 Dataset.* Sept. 2016.

[79] Dutch T. Meyer and William J. Bolosky. "A study of practical deduplication". In: *ACM Transactions on Storage* 7.4 (Jan. 2012), pp. 1–20.

[80] Midjourney. *Midjourney Terms of Service.* Feb. 2023.

[81] R. Mukundan, S.H. Ong, and P.A. Lee. "Image analysis by Tchebichef moments". In: *IEEE Transactions on Image Processing* 10.9 (2001), pp. 1357–1364.

[82] Bowen Ning et al. "Research and Development of Copyright Registration and Monitoring System Based on Digital Watermarking and Fingerprint Technology". In: *2021 International Conference on Culture-oriented Science & Technology (ICCST).* 2021, pp. 354–358.

[83] U.S. Copyright Office. *Definitions (FAQ) — U.S. Copyright Office.* 2008.

[84] Paul Ohm. "Focusing on Fine-Tuning: Understanding the Four Pathways for Shaping Generative AI". In: *SSRN Electronic Journal* (2024).

[85] David W. Opderbeck. "Copyright in AI Training Data: A Human-Centered Approach". In: *SSRN Electronic Journal* (2024).

[86] Daeeol Park, Hyunsik Na, and Daeseon Choi. "Performance Comparison and Visualization of AI-Generated-Image Detection Methods". In: *IEEE Access* 12 (2024), pp. 62609–62627.

[87] Frank A. Pasquale and Haochen Sun. "Consent and Compensation: Resolving Generative AI's Copyright Crisis". In: *SSRN Electronic Journal* (2024).

[88] Patronus AI. *Introducing CopyrightCatcher, the First Copyright Detection API for LLMs.* Mar. 2024.

[89] S. Preetha and V. R. Bindu. "A Wavelet Optimized Video Copy Detection Using Content Fingerprinting". In: *Journal of Signal Processing Systems* ().

[90] Mahima Pushkarna, Andrew Zaldivar, and Oddur Kjartansson. "Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI". In: *2022 ACM Conference on Fairness, Accountability, and Transparency.* FAccT '22. June 2022.

[91] Amna Qureshi and David Megías Jiménez. "Blockchain-Based Multimedia Content Protection: Review and Open Challenges". In: *Applied Sciences* (2021).

[92] Homayoon Rafatijo and Dennis Crouch. "When Is Software Code Copyrightable? Is Its Unauthorized Copying Excusable as a Fair Use?" In: *Preview of United States Supreme Court Cases* (2020), p. 14.

[93] B. Rasina Begum and P. Chitra. "SEEDDUP: A Three-Tier SEcurE Data DedUPlication Architecture-Based Storage and Retrieval for Cross-Domains Over Cloud". In: *IETE Journal of Research* (Feb. 2021).

[94] Jie Ren et al. "Copyright Protection in Generative AI: A Technical Perspective". In: *arXiv* (2024).

[95] Daniel Rodriguez Maffioli. "Copyright in Generative AI training: Balancing Fair Use through Standardization and Transparency". In: *SSRN Electronic Journal* (2023).

[96] Eleonora Rosati. "Infringing AI: Liability for AI-generated outputs under international, EU, and UK copyright law". In: (2024).

[97] Brendan Leigh Ross et al. *A Geometric Framework for Understanding Memorization in Generative Models.* 2024. arXiv: 2411.00113.

[98] Yilin Sai et al. *Is Your AI Truly Yours? Leveraging Blockchain for Copyrights, Provenance, and Lineage.* 2024. arXiv: 2404.06077.

[99] Prasanth Vaidya Sanivarapu et al. "Digital Watermarking System for Copyright Protection and Authentication of Images Using Cryptographic Techniques". In: *Applied Sciences* (Aug. 2022), pp. 1–13.

[100] Giordano Scarciotti and Alessandro Astolfi. "Data-driven model reduction by moment matching for linear and nonlinear systems". In: *Automatica* (2017).

[101] Johannes Schneider. "Explainable Generative AI (GenXAI): a survey, conceptualization, and research agenda". In: *Artificial Intelligence Review* ().

[102] Shawn Shan et al. "Glaze: protecting artists from style mimicry by text-to-image models". In: *Proceedings of the 32nd USENIX Conference on Security Symposium*. SEC '23. Anaheim, CA, USA: USENIX Association, 2023. ISBN: 978-1-939133-37-3.

[103] Sujie Shao et al. "WFB: watermarking-based copyright protection framework for federated learning model via blockchain". In: *Scientific Reports* (Aug. 2024).

[104] Hong Shen et al. "The Model Card Authoring Toolkit: Toward Community-centered, Deliberation-driven AI Design". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '22. 2022.

[105] Ilia Shumailov et al. *UnUnlearning: Unlearning is not sufficient for content regulation in advanced generative AI*. 2024.

[106] Meixue Si et al. *A Solution toward Transparent and Practical AI Regulation: Privacy Nutrition Labels for Open-source Generative AI-based Applications*. 2024. arXiv: 2407.15407.

[107] Ravidu Silva and Jordan J. Bird. *AI-ArtBench*. 2023.

[108] Athanassios Skodras. *Discrete Wavelet Transform: an Introduction*. Feb. 2003, pp. 1–26.

[109] Jonathan Stempel. *NY Times sues openai, Microsoft for infringing copyrighted works ...* Dec. 2023.

[110] Haoyu Tang et al. *Learn while Unlearn: An Iterative Unlearning Framework for Generative Language Models*. 2024. arXiv: 2407.20271.

[111] Stuart A. Thompson. "We Asked A.I. to Create the Joker. It Generated a Copyrighted Image." In: *The New York Times* (Jan. 2024).

[112] UK Information Commissioner's Office. *AI and Data Protection Risk Toolkit*. May 2024.

[113] K. Vijayalakshmi and V. Jayalakshmi. "Analysis on data deduplication techniques of storage of big data in cloud". In: *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. 2021.

[114] Jiachen T. Wang et al. *An Economic Solution to Copyright Challenges of Generative AI*. 2024. arXiv: 2404.13964.

[115] WikiArt. Dec. 2022.

[116] Bin Xiao, Jian-Feng Ma, and Jiang-Tao Cui. "Radial Tchebichef moment invariants for image recognition". In: *Journal of Visual Communication and Image Representation* 2 (Feb. 2012).

[117] Yaqi Xie et al. *MUGC: Machine Generated versus User Generated Content Detection*. 2024.

[118] Canwen Xu et al. *Small Models are Valuable Plug-ins for Large Language Models*. 2023.

[119] Weiwei Xu et al. *A First Look at License Compliance Capability of LLMs in Code Generation*. 2024.

[120] Rui-Jie Yew. "Break It 'Til You Make It: An Exploration of the Ramifications of Copyright Liability Under a Pre-training Paradigm of AI Development". In: *Proceedings of the Symposium on Computer Science and Law*. 2024, pp. 64–72.

[121] Zhipeng Yin et al. "AMCR: A Framework for Assessing and Mitigating Copyright Risks in Generative Models". In: *ECAI 2025*.

[122] Zhipeng Yin et al. "Uncertain Boundaries: A Tutorial on Copyright Challenges and Cross-Disciplinary Solutions for Generative AI". In: *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. 2025.

[123] Peter Young et al. "From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions". In: *Transactions of the Association for Computational Linguistics* (2014), pp. 67–78.

[124] Xiaomin Yu et al. *Fake Artificial Intelligence Generated Contents (FAIGC): A Survey of Theories, Detection Methods, and Opportunities*. 2024.

[125] Dawen Zhang et al. "Privacy and Copyright Protection in Generative AI: A Lifecycle Perspective". In: *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI*. 2024.

[126] Haoqi Zhang. "Research on Legal Risks and Responses of Generative Artificial Intelligence". In: *Mod. L. Rsch.* 5 (2024), p. 57.

[127] Hongjiao Zhang and Yahong Li. "Opt-Out Implied Licenses in Copyright Law: From Search Engines to GPAI Models". In: *GRUR International* (July 2024).

[128] Kevin A. Zhang et al. "Robust Invisible Video Watermarking with Attention." In: (2019).

[129] Lingzhi Zhang et al. *Perceptual Artifacts Localization for Image Synthesis Tasks*. 2023. arXiv: 2310.05590.

[130] Lingzhi Zhang et al. *Perceptual Artifacts Localization for Inpainting*. 2022. arXiv: 2208.03357.

[131] Yongjing Zhang et al. "Machine Unlearning by Reversing the Continual Learning". In: *Applied Sciences* 13 (Aug. 2023), pp. 1–10.

[132] Xuandong Zhao et al. *Invisible Image Watermarks Are Provably Removable Using Generative AI*. 2024. arXiv: 2306.01953.

[133] Jingjing Zheng et al. "A Novel Video Copyright Protection Scheme Based on Blockchain and Double Watermarking". In: (Dec. 2021). Ed. by Honghao Gao.

[134] Haonan Zhong et al. "Copyright Protection and Accountability of Generative AI: Attack, Watermarking and Attribution". In: *Companion Proceedings of the ACM Web Conference 2023*. WWW '23. Apr. 2023.

[135] Junlei Zhou et al. *CopyScope: Model-level Copyright Infringement Quantification in the Diffusion Workflow*. 2023.

[136] Lixi Zhou et al. "Serving deep learning models with deduplication from relational databases". In: *Proceedings of the VLDB Endowment* (June 2022), pp. 2230–2243.

[137] Gina Maria Ziaja. "The text and data mining opt-out in Article 4(3) CDSMD: Adequate veto right for rightholders or a suffocating blanket for European artificial intelligence innovations?" In: *Journal of Intellectual Property Law & Practice* (Feb. 2024), pp. 453–459.