

Copyright protection paradigm in the generative AI era: originality, attribution, and infringement

Qinru Li

China University of Political Science and Law, Beijing, China

240820438@cupl.edu.cn

Abstract. This paper examines the copyright challenges raised by generative artificial intelligence (“AI”). It begins by analyzing the generative process, arguing that the originality of text prompts can serve as a basis for the copyrightability of AI-generated outputs. Accordingly, the study then explores the complex issue of attribution based on the proposed generative process, evaluating the competing claims of users, developers, and data providers, and proposes an ownership model that reflects the degree of creative contribution. Finally, the paper addresses infringement concerns, comparing regulatory approaches in the US, EU and China, and advocates a balanced protection framework that reconciles innovation with copyright safeguards.

Keywords: AI-generated content copyright originality infringement

1. Introduction

The rapid proliferation of generative AI has precipitated unprecedented challenges to traditional copyright law frameworks. This article studies the emerging field of generated-AI works by addressing critical gaps in existing law systems, and advancing a conceptual framework for resolving the foundational question of copyright protection throughout AI generative processes.

This paper argues that existing copyright systems can adapt through a conceptual framework that determines ownership by reference to human creative input and, on the basis of clarified ownership, applies traditional doctrines of infringement in the AI era. The analysis proceeds in three main parts. First, it deconstructs the AI workflow across its different stages, analyzing the copyrightability of AI-generated content (“AIGC”), and evaluating the originality of text prompts across main jurisdictions (the US, EU, UK, and China), arguing that creative prompts constitute protectable expression. Subsequently, building on the identification of originality, the paper determines authorship by examining the competing claims of users, developers, and data providers, and proposes a multi-factorial model for attribution based on the degree of control and contribution. Finally, it examines infringement paradigms within the AI generative process, focusing on the legality of data scraping and the application of doctrines such as fair use, synthesizing insights from recent case law and regulatory developments. By doing so, this paper aims to provide a coherent framework for the copyright protection of AI-generated works while balancing the rights of stakeholders in AI generation.

Each of the three parts of this paper carries its own innovative contribution. Firstly, this paper begins with a novel “staged analysis” framework that traces the generative process from user input to final output, establishing the originality of text prompts as a legally defensible basis for claiming copyright protection of AI outputs. Secondly, it develops a multi-factor attribution model that evaluates ownership claims through the dual lenses of creative contribution and operational control, thereby moving beyond the conventional binary of human versus AI authorship. Finally, it offers a comprehensive comparative analysis of legal approaches across major jurisdictions—United States, European Union, United Kingdom, and China—identifying both convergent and divergent regulatory pathways. This article draws on recent landmark cases and extends its analysis to providing legislators with a structured framework for future policy-making, while also offering developers practical compliance guidelines and ensuring users’ creative rights receive protection.

2. A staged analysis of AI-generated works

Artificial Intelligence allows users to process a vast array of existing content, not only arts, but all forms of creative output. In a typical generative AI workflow, users input command, and then AI will produce outputs based on these instructions. Numerous generative AI models already operate in this manner. Notable examples include ChatGPT and DeepSeek. The process of how AI

produces works includes receiving instructions from users, discovering and analyzing data from their database and producing output—bears significant resemblance to traditional creative techniques such as music sampling and collage-making, both of which involve the reuse of existing materials to produce something new. This functional similarity suggests that AI-generated works may be examined within the traditional framework of copyright law.

However, the role played by AI in the creative process is fundamentally different from the direct reuse of existing works by human authors. Unlike human creators who intentionally incorporate specific pre-existing materials, AI systems passively collect these works to establish its database. These systems do not reproduce individual works directly in response to user commands, nor is there typically any commercial use of the training data itself. Instead, AI provides a technical infrastructure that facilitates the generation of new content based on statistical patterns found in its training corpus. Questions arise concerning the causal link between user input (i.e., text prompts) and AI-generated outputs, as well as the attribution and ownership of such outputs. Addressing these issues requires a stage-by-stage analysis of the generative process—from the initial user input to the final output.

The ability of AI creative capacity depends on the quality of training data—often protected by copyright—and the user-supplied instructions that guide the generation process. Moreover, the copyright during the procedure of AI selecting data and producing by algorithms remains complex and difficult to determine. To explore the intellectual property rights of generative AI works, it is necessary to examine the entire chain of generation, including the relationship between input prompts, the underlying training data, and the resulting outputs.

2.1. The originality of text prompts

As a pivotal conduit between human intention and machine execution, a text prompt translates creative directives into actionable instructions for generative AI systems, thereby serving as the foundational input that governs the content, style, and boundaries of the ensuing output [1]. For instance, a possible input for a text-generating AI model like OpenAI's Generative Pre-trained Transformer 3 (GPT-3) could be: "Compose a short story in which a time traveler uncovers a hidden civilization in the future" [2]. In comparison, for an AI system designed to generate visual art—such as DeepDream or StyleGAN—a text prompt might adopt a slightly different structure, for example: "Produce a surreal landscape painting influenced by the themes of dreams and imagination." Although such inputs are aimed at producing visual outputs rather than text, and can differ significantly in structure and detail, all text prompts fundamentally act as the initiator for AI-generated content. The core function remains consistent: to offer direction and stimulate creativity in the AI's generative process. Whether simple or elaborate, these prompts supply the AI with operational boundaries, contextual cues, and creative guidance, shaping the tone, style, and substance of the final output.

Understanding the originality of text prompts is a prerequisite and an important part of the intellectual property of AI works, for the originality of works, different jurisdictions apply distinct standards. The relevant standards for the United States, European Union, United Kingdom, and China are summarized as shown in Table 1.

Table 1. The originality standards of different countries in copyright law

Jurisdictions	Standards
United States	Modicum of creativity rather than mere effort or labour [3]
European Union	The author's own intellectual creation [4]
United Kindom	Skill, labour, and judgement [5]
China	Creativity and have a certain form of expression

The determination of whether a text prompt qualifies as an original work varies significantly across the EU, UK, US, and China, each adhering to distinct legal standards of originality. Within the EU, the criterion of "the author's own intellectual creation" necessitates that the prompt reflects personal creative choices, potentially extending protection to concise prompts provided they embody such originality. The UK, while historically emphasizing "skill, labour, and judgement", may also recognize prompts as original where they exhibit deliberate creative effort, even if diverging from the more recent *Ifopaq* standard. Conversely, both the US and China prioritize "creativity" as the central element for copyrightability [6]. Here, text prompts must demonstrate a minimal degree of creative expression to be eligible for protection, thereby excluding mundane or purely functional instructions, while safeguarding those that convey unique and individualized formulation.

Despite jurisdictional nuances, a common thread across these frameworks is the nexus between originality, creativity, and tangible expression—elements that serve to differentiate one work from another. Text prompts, by their nature, often entail thoughtful construction, embedding user-defined context, strategic direction, and specific parameters. This process inherently involves intellectual engagement and personalized decision-making. Consequently, a substantial subset of text prompts is likely to satisfy these generalized originality thresholds, thereby attaining eligibility for protection as copyrightable expressions.

2.2. The copyrightability of generative AI outputs: examining the correlation between text prompts and outputs

The mere fact that someone can claim copyright over a text prompt does not automatically confer protection over the AI output generated using that prompt. Copyright protection extends only to the expression of ideas, not the ideas themselves [7]. The emergence of AI interrupts the traditional link between human and work they create, raising the issue about authorship and whether AI-generated content can qualify as a “personal intellectual creation”—a concept central to the recognition of copyright protection.

Traditionally, AI works are not regarded as protectable intellectual properties. The predominant scholarly stance insists on an indispensable link between copyrightability and direct human authorship.¹ This human-centred approach is reflected in the Berne Convention, Article 2(1), by using the notion of “original works” [8]. European jurisdictions such as the German Copyright Act (Sec. 2), require a human creation as the basis for the work, it is crucial whether a work created by an AI can still be qualified as a protectable work. Thus, the International Association for the Protection of Intellectual Property (AIPPI) concluded in a recent study that most jurisdictions reject any copyright attributed to AI-assisted works [9]. This anthropocentric view, however, is increasingly strained by the reality of AI-generated content that exhibits emergent creativity, prompting calls for a recalibration of the authorship concept to account for supervisory or contributory human agency.

Given that AI-generated outcome cannot be considered to be inaugurated by a human being, an important question arises: Is it therefore impossible to protect generative AI works through copyright of text prompts? The answer may not be so absolute. Rather than focusing solely on whether AI itself can be qualified as a legal subject, it may be more productive to assess the copyrightability of AI works by examining the process of generation. In author’s opinion, today under the current technology it still seems impossible for a truly independent-thinking intelligence coming, so we can think about the copyrightability of AI works through the correlation of prompts issued by human to the work.

As indicated above, text prompts may be eligible for copyright protection when they fulfill the requisite criteria of originality. This possibility invites a reconsideration of copyright discourse, emphasizing the connection between human creative input—expressed through prompts—and machine-generated outputs. Given that text prompts themselves can be deemed copyrightable subject matter, it is meaningful to examine whether and how this human creativity carries over to the content produced by AI.

From a technical standpoint, the relationship between a text prompt and the AI’s output is influenced by multiple factors. Even when provided with an identical prompt, an AI system may not generate the same result consistently [10]. This variability can be attributed to stochastic elements embedded in the algorithm, as well as differences in model architecture, training data, and inference methods.

In the context of text-based models—such as large-scale language models—outputs often reflect a recognizable degree of thematic or stylistic consistency, yet still allow for noticeable variation. For generative models producing images, music, or other artistic forms, the translation of linguistic prompts into visual or auditory outputs involves additional layers of interpretation, incorporating subjective and creative agency by the AI system [11].

Thus, by integrating insights from both AI mechanisms and copyright theory, we may develop a more nuanced understanding of creativity, authorship, and ownership in the context of AI-generated works.

2.3. Application of the idea-expression dichotomy

Building on the standards of originality and technical considerations previously discussed, this section applies the idea-expression dichotomy—a core principle in copyright law that separates non-protectable ideas from protectable forms of expression—to assess the originality of both text prompts and AI-generated outputs. This framework helps determine whether copyright may subsist in the expressive elements contained within prompts and subsequent AI-produced content.

Text prompts reflect deliberate creative choices and linguistic skill on the part of the human author, representing a unique expression intended to guide the AI. Similarly, AI-generated outputs, while algorithmically produced, often exhibit originality through their interpretation of the prompt and synthesis of training data. By acknowledging the author’s creative contribution in shaping both the input and influencing the output, copyright law can offer protection to the intellectual effort and artistic judgment exercised by humans throughout the AI-assisted creative process.

That said, we still cannot assert that it is solely the originality of prompts that determines the copyrightability of AI works. AI generation is a complex process that not only users themselves are not limited to prompts: users may fine-tune models with materials to target specific outputs, and the inputs used by AI cannot always be regarded as entirely created by the individual who issued the prompt. However, we can’t deny that in the era of weak AI, copyright protection and copyright acknowledgement of generative AI works through the originality of text prompts is necessary. AI works need to be protected, and the law system must be adapted to the reality of the situation, and perhaps we should pay attention to the means of protection and attribution of AI works.

3. Copyright attribution of generative AI works

The question of whether generative AI works are copyrightable, as discussed in Part 2, remains unresolved, particularly in terms of identifying authorship and ownership. While the originality of text prompts may suggest a basis for the copyrightability of outputs, this does not necessarily confer ownership to users. In the AI era, AI systems typically rely on existing works to establish its database and develop a technical infrastructure that facilitates the generation of new content based on statistical patterns. This process complicates the identification authorship and ownership, as the contributions of both database and AI system developers are significant.

Moving forward, the paper will explore the question of ownership and responsibility in the context of AI-generated works. It will examine the possible ownership structures of AI-generated works and analyze potential frameworks to identify the authorship in order to provide guidance for protections of AI works.

3.1. The possibilities of the ownership of AI works

The process of AI-generated works involves multiple stakeholders, and the ownership of the final output is a complex issue. As discussed in Part 2, the production of AI works can be divided into two stages, from inputs to outputs. This part will follow this classification and analyze the possibilities of ownership in these two stages.

Some existing laws and regulations have stipulated the copyright ownership and judgment criteria for determining the copyright ownership of AI-generated works. In the common law systems, the principle of “human authorship” generally denies copyright of AI-generated works that lack human creations. For example, if an AI tool is used to enhance a photograph or to generate software with AI assistance, the output may be considered copyrightable if the input (e.g., text, images, audio, or video) is substantially retained in the output.² This is referred to as “expressive inputs”. UK copyright law adopts a distinctive approach by attributing authorship of computer-generated works to the individual who undertakes the “arrangements necessary for the creation of the work” (CDPA 1988, s 9(3)). This provision effectively establishes a legal fiction of authorship, vesting rights in the human agent who orchestrated the generative process, rather than in the AI tool itself. This means that English law attributes the work to those who make the substantive arrangement behind the scenes, which may include the programmer or the user who operates the system. This approach recognizes the copyrightability of computer-generated works, though the protection period is limited to 50 years after publication, which is shorter than that of ordinary works.³ This reflects a compromise model that grants limited copyright to contributors.

In China, current judicial practice holds that authors are limited to natural persons, legal persons or unincorporated organizations, and the artificial intelligence itself cannot be considered authors.⁴ In cases of disputes between developers and users, courts prioritize the protection of those who actually use AI for creation. The 2023 ruling by the Beijing Internet Court (Beijing Li case) marked a significant jurisprudential shift. The court’s recognition of the plaintiff’s copyright in an AI-generated image, predicated on the creative intellectual input embodied in the crafting of prompts and the selection of parameters, not only incentivizes user creativity but also imposes a corresponding duty of care on users to avoid infringing third-party rights through their prompts.⁵ Thereby the plaintiff was the author of the work, it establishes the “user first” ownership allocation rule and major AI platforms in China state in their user agreements that users own the copyright of their generated content.

3.2. Application of the possibilities of the ownership

In a legal context, this issue can be divided into two scenarios. First, if clear agreements exist, for example some AI platforms will specify the copyright ownership of works in detail in their user agreements. In this way, the ownership of the generated works is clearly defined and disputes are avoided. However, when no such agreements exist or the terms are not explicitly defined, users and developers or other contributors may be seen as the authors of AI works, then who is the real owner?

From the input stage, users often play the most important role. They conceive the idea, formulate specific prompts, guide the AI through iterations, and make crucial selection and refinement decisions. Their creative vision and labor shape the final output. For instance, a graphic designer meticulously prompting an image generator to achieve a unique visual style for a client project invests significant creative effort. In such cases, the ownership aligns with traditional copyright principles, which focus on the human author who bring the work into its final form. Moreover, the originality of prompts suggests that users could be seen as the “author” of the unique expression derived from the AI’s capabilities.

At the output stage, it is generally not considered that AI itself can be considered an author under current laws or practices. However, during the AI generation process, data providers and AI system developers may be contributors for an AI-generated work. This means AI works are not simply belong to users because generative AI models are trained on massive datasets—text, images, code, music—often scraped from the web or licensed from material providers. If the database was developed without the permission of original authors, the use of database may raise ownership disputes. Unlike computer software works where developers can claim rights, computer-generated works are directly played by AI systems, making it difficult for developer to

express the creativity and be regarded as an author. Additionally, the practice of data scraping raises ethical and legal questions. Especially in the AI generation process where developers have an advantageous position to use existing data available, this issue must be discussed. While in the UK, the government proposed an exception to copyright by which data scraping is allowed, but data owners can object to it if they want to reserve their rights. The government seeks to establish a standardized mechanism to declare the objection and this kind of protection will be further discussed in Part 4 next.

A further issue concerns whether, once data is authorized for use by the developers, can the original authors especially ones who enjoy complete copyrights claim the relative rights and the copyright of the AI-generated work? This argument rests on the principle that the AI's output is fundamentally derivative [12]. The original creators of the training data, such as artists, photographers, writers, publishers, and data repositories, may reasonably claim rights over those data scraped by AI system. Yet unresolved questions remain. AI-generated works are not database; they are new creations and do not reflect the ideological expression of the data providers. For those providers, they do not have direct impacts on AI-generated works, it is more reasonable to limit their rights to their own works that have been used in training, rather than extending them to the AI-generated content itself.

Above all, it can be concluded that the most direct claim often lies with the users. However, the programmer's rights in the tool and the data provider's rights in the training data create significant complications: While we mainly focused on text prompts and users, the role of AI developers and original data providers in the creation of AI generated works raises important concerns regarding authorship and ownership.

Currently, some parts of the copyright of AI-generated works remain in a gray area, particularly regarding attribution. Different jurisdictions may adopt divergent approaches, and with the rapid iteration of AI technology, the issue is becoming increasingly complex. Across both common law and civil law systems, current copyright frameworks—exemplified by the position of the US Copyright Office—continue to require human authorship for protection, generally attributing rights to users if their creative input is substantial.

3.3. A possible pattern to identify the attribution of AI works

A possible pattern may be provided to determine the ownership of certain AI work. Consider the following example: an artist trains an AI on a selection of Rembrandt's paintings, resulting in the generation of a new work in Rembrandt's style. The determination of authorship depends critically on the extent of the artist's influence over the output. If the artist intentionally curates a specific subset of paintings—rather than using the entire corpus—their creative control becomes substantially more meaningful, and the resulting work may reasonably be attributed to them [13]. By contrast, if the training data includes all available works by Rembrandt or even paintings by other artists, the artist's influence over the final output is considerably diluted [14]. This illustrates how users can shape their creative contribution through deliberate choices in dataset selection and prompt design. The level of creative detail embedded in these decisions significantly affects the appearance and originality of the final output.

Therefore, when determining attribution, priority should be given to the degree of authorial control and the intentional design of the creative process. The person who exercises meaningful creative choices and directs the generative process should be recognized as the author—with the possibility of joint authorship should multiple individuals make substantial creative contributions [15]. Typically, the programmer of the AI system is excluded from authorship, since most general-purpose AI models are not designed to produce a single specific work. However, if the AI is explicitly programmed for a particular creative task, the programmer may also qualify as an author based on their autonomous decisions in designing the system [16].

Consequently, when addressing attribution questions in AI-generated works, the degree of human control over the creative process proves to be a more decisive factor than abstract notions of creativity. To evaluate whether copyright protection applies to specific human contributors, it is essential to examine the creative input at each stage of production and assess whether sufficient originality is embodied in the final work.

Little harmonization exist at the global level. Ownership of AI-generated works resides in a complex, contested space where the claims of users, AI programmers, and material providers intersect and often clash. The user provides the creative spark and direction, the programmer builds the essential engine, and the material provider supplies the foundational knowledge. Measuring these contributions may suggest a potential solution: allocating coauthorship or co-ownership between users and the developers or owners of AI systems. What is needed, however, is a systematic and concise model for attribution, and this model must be universal so that people who do not learn about AI algorithms or the copyrights of AI works can apply it.

In this pattern, the principle of “user first” (mentioned above in part 3.1) should serve as the starting point because the direct influence in AI works from users. In China, this principle not only protects the rights and interests of ordinary users who are in a relatively vulnerable position in creation, but also enables AI system providers to clarify the ownership arrangement fairly and reasonably in the service agreement. However, there may be exceptions to this principle, that is when the input lose its originality and do not contribute much to the AI-generated work. For example, users simply input some existing works into AI and ask for similar text or simply convert them into graphics or music. At this time, although parts of these prompts may be original, it is difficult to rule out their association with existing works, at which point the user's contribution to the final output work becomes

obscured. Also, one certain AI work may exist more than one users of AI system. Considering this, following the degree of contribution is necessary to examine whether the instructions entered by the user are sufficiently original and creative. Thus, contribution analysis remains necessary: the author should be identified as the party whose expression of ideas through prompts is most clearly reflected in the final work. Similarly, the contribution principle extends to the output stage, where developers may also have certain rights to the generated works when the relevance of the prompts cannot be confirmed. This approach acknowledges the significant contribution of the AI in the creative process while also recognising the role of the developer in creating and maintaining the AI technology.

Current intellectual property laws struggle to resolve this trilemma. Further solutions will likely require nuanced approaches: clearer legal frameworks acknowledging layered contributions, sophisticated licensing models that define rights and revenue sharing explicitly, and ethical guidelines ensuring fair recognition. Ultimately, navigating this new frontier requires acknowledging that the authorship of an AI work is often a collaborative, albeit involuntary, effort involving multiple stakeholders. Addressing this challenge requires innovative thinking and a balanced approach to ownership in the algorithmic age.

4. Copyright protection of generative AI works

The legal landscape is no stranger to the complexities introduced by artificial intelligence. Earlier scholarly and policy debates already engaged with the legal status of autonomous software agents in forming contracts. A key distinction, however, lay in the fact that such agents typically operated under substantial human direction from their programmers. Consequently, any rights, responsibilities, or liabilities stemming from the actions of these artificial entities were predominantly assigned to their creators. This framework has undergone significant transformation. Contemporary AI systems function with a high degree of autonomy, effectively blurring the lines of ownership and authorship for AI-generated outputs—a shift from traditional software-created works that complicates the attribution and proof of infringement.

This evolution has triggered a vibrant global discourse among scholars and policymakers, particularly concerning the applicability of established tort law and adjacent legal doctrines to AI-related harms [17]. A considerable portion of this literature investigates which legal grounds or theories of tort are most suited to addressing the unique challenges posed by AI systems. The debate continually returns to a fundamental question: who bears responsibility for this technology, and which parties should be held liable for harms resulting from its operation? Commentators have put forward diverse, often competing, viewpoints supported by doctrinal analysis, policy considerations, and normative reasoning [18]. As a result, although tort law offers a range of conceptual tools for determining liability, assigning responsibility in practice remains a profoundly complex undertaking [19].

4.1. The significance of infringement issues in AI works through copyright law

As argued above, AI-generated works generally do not qualify for copyright protection, making it difficult for harmed entities to seek remedies other than seeking compensation through AI-specific legislation and tort law. This drives more actors to exploit the copyright uncertainty of AI works to evade infringement liability. Meanwhile, the AI-related infringement differs from traditional tort law because tort liability is grounded in human action, where individuals are expected to act reasonably to avoid harm to others, while AI-related infringements mainly arise from AI generations or the extensive use of pre-existing works. Accordingly, AI-related infringement paradigm can be divided into two categories according to different objects between its paradigm and the traditional infringement paradigm: (1) the infringement between AI-generated works themselves, and (2) infringements of AI-generated works against original human-authored works, particularly those scraped into AI training databases. This division clarifies the complexity of the copyright attributions of AI works and incorporate the responsibilities and obligations of AI developers mentioned in Part 3, thereby contributing to a more comprehensive model for protecting both AI and human creations.

The first category—AI works against AI works—centres on the copyrightability of AI-generated material. As discussed in Part 3, the first AI-related infringement case heard by Beijing Internet Court (the Beijing Li case) addressed this issue. The court held that once the AI-generated material is recognized as a work, questions of proof content and liability follow the same principles as in traditional tort law. Since both parties' works belong to the same category, the complexity is greatly reduced, and the key focus is on whether the AI-generated works themselves can be included in intellectual property protection. The importance of this infringement paradigm lies in its ability to integrate AI creation into the intellectual property framework, thus adapting to the new challenges posed by AI development.

The second category—AI works against human works—raises more difficult questions. AI training typically requires large datasets that contain copyrighted text, images, music, and videos. Using such material without authorisation may constitute copyright infringement. Key questions include: does scraping and using copyrighted content for AI training constitute infringement? Can data owners object to the scraping? National laws slightly diverge on this issue. The US applies a “fair use” doctrine, while the UK recognizes “fair dealing”, and has also proposed an exception to copyright allowing data scraping. Fair

use allows copyrighted data to be used without the owner's consent, provided specific conditions are met. Whether training AI models falls under fair use is a fundamental issue in this infringement paradigm. Different answers to this question determine whether developers or users can avoid tort liability. In practice, fair use embodies a balancing act between data utilization and copyright protection to some extent. The following section will examine how current legislative systems in different jurisdictions apply tort law and intellectual property law to AI-related infringements, with particular attention to the scope and limits of fair use exceptions.

4.2. Regulations of AI infringement issues and protections

On February 11, 2025, the U.S. District Court for the District of Delaware rendered a partial summary judgment in *Thomson Reuters v. Ross*, holding that Ross's unauthorized use of copyrighted works to train AI legal search tools constituted copyright infringement and did not qualify as fair use.⁶ This is the first case in the United States on AI training data. Before this, in 2023, the same court had denied the plaintiffs' motion for summary judgment, holding that the originality and fair use of desk annotations required jury determination. Two years later, the court amended its previous judgment, finding that the plaintiff's claim was in partial favor and clarifying that the defendant's unauthorized use of Westlaw's 2,243 desk annotations for AI training constituted copyright infringement and was not fair use.

Fair use is an important defense mechanism in copyright law, aiming to balance copyright protection with social interests. The court considered the fair use defense under the four elements: (1) the purpose and nature of the use, (2) the nature of the copyrighted work, (3) quantity and quality used, (4) market impact. Ultimately, *Thomson Reuters* prevailed on the two most important elements—the first and fourth—leading the court to reject Ross's fair use defence. The court held that the defendant's use of the desk annotations was commercial in nature, directly competitive with the plaintiff's products, and lacked a "transformative purpose". It further highlighted that market harm (factor four) was the most important consideration.

However, whether this case can fully guide the issue of infringement of AI data use. As the judge said in the judgment, this is not a typical generative AI copyright case. In this case, the defendant's AI legal search only displays "existing judicial judgment opinions". Therefore, the court found that the defendant's use of the plaintiff's copyrighted work was the same as the plaintiff's and lacked a "convertible" purpose of use. In this regard, if this case is a typical generative AI copyright infringement case—the defendant AI company unauthorized capture and copying of the plaintiff's copyrighted works for the purpose of training AI data, and its AIGC-generated content is new content, then does the purpose of training AI data by the AI company satisfy the analysis of "conversion use" and may be judged as non-infringing? At a minimum, fair use does provide a relatively appropriate template for proving AI infringement issues.

Litigation over AI training data is proliferating. U.S. technology firms such as OpenAI, Meta, Stability AI, and Anthropic are facing multiple lawsuits from copyright owners, most of which remain pending.⁷ In China, on June 20, 2024, the Beijing Internet Court held an online hearing to hear four copyright infringement cases in which artists sued AI painting software developers and operators, where Chinese courts follow the traditional copyright infringement rules, including stopping the infringement, compensating for losses, and protecting the author's rights. In the *Beijing Li* case, the court awarded economic damages of 500 yuan and ordered a public apology to protect authorship. Once AI-generated works are recognized as works under copyright law, infringement remedies in China do not differ from those applicable to human-authored work: unauthorized use requires legal liability. This adjudication logic not only protects the enthusiasm of creators, but also delineates a legal red line for all parties in society to use AIGC content correctly. However, rudely denying data scraping also poses some problems. AI-generated content is more difficult for rights holders to detect infringement and prove ownership. In addition, this also limits the use of data, which may curb further learning and development of AI.

Prior to the emergence and widespread adoption of generative AI technologies such as ChatGPT, Stable Diffusion, MidJourney, DALL-E, GitHub Copilot, or Udio, the European Union had already established legal mechanisms addressing certain uses of protected content. Directive (EU) 2019/790 on Copyright in the Digital Single Market (CDSM Directive), enacted in April 2019, introduced two specific exceptions to copyright and related rights to facilitate text and data mining (TDM): one applicable for scientific research purposes, and another for other uses, with rightsholders retaining the ability to opt out. More recently, the Regulation (EU) 2024/1689 of 13 June 2024, known as the AI Act, which is in its final negotiation stages, imposes obligations on providers of general-purpose AI (GPAI) models. These include ensuring transparency regarding the datasets used for training and implementing a standardized policy to accommodate opt-out requests from copyright and related rights holders [20].

The EU's regulatory strategy emphasizes balancing the facilitation of data access with the protection of intellectual property, achieved mainly through data-sharing agreements and industry-led governance, rather than establishing new intellectual property rights. Overall, the EU maintains a cautious stance toward copyright issues in AI-generated content. While future discussions may explore concepts such as a "right of data producers" and the protection of AI-generated works based on economic impact assessments, short-term efforts will likely focus on flexible interpretation and specialized regulation within the current legal framework.

However, structural tensions exist between the CDSM Directive—a private law instrument—and the AI Act, which operates as public law. This divergence raises several challenges in aligning both legislative regimes, particularly concerning the territorial scope of obligations and their enforcement mechanisms. Clarification will be needed to ensure coherent application across legal domains.

Taken together, jurisdictions exhibit divergent models of protection, as shown in Table 2:

Jurisdictions	Patterns	Drawbacks
US	Fair use	Ambiguous and difficult to define
China	Traditional tort law	Excessive tendency to protect the original author
EU	Exceptions and special legislation	Cumbersome documents in both public and private law

For the purpose of this article, the process of AI generation is fundamental. Attribution and liability should be examined through the lens of developers and users, leveraging existing tort law and intellectual property law frameworks. Taken together Parts 2 and 3, assessing contribution and relevance of prompts and outputs to prove attribution of AI works can help to identify and prove who is responsible for infringements. Importantly, AI works, especially the relationship between users and developers makes it difficult to exclude one of the parties from the burden of proof. Even if the AI work ownership is clearly defined, developers may withhold information on training data or generation processes, disadvantaging claimants. Therefore, when users face liability, developers should be required to cooperate in evidentiary matters. Clarifying rules of proof in AI infringement cases is therefore crucial: general principles must be respected, and the defendant or relevant third party should provide evidence to assist if necessary.

Exception also remain central, such as fair use and exceptions. When considering these exceptions, it is important to pay attention to the degree which means there are many factors to consider and should be based on the four elements adopted by the U.S. fair use rule: (1) Share of the original works used: the more of it that is used into the training, the less strong the fair use defence; (2) Commercial: The more commercial the use of the AI output, the less strong the fair use defence; (3) Transformative: The more transformative the output of the AI model compared to the data inputs, the more likely it is that will count as fair use; (4) Effect on market: Whether the AI model is generating works in a similar style or category as the original work, thereby diluting the market for that work. The more the AI output can result in “lost sales” for the input work, the less strong the fair use.

Copyright governance for AIGC should not rely solely on legislation. Admittedly, it is helpful to characterize AI infringements not only through tort theories but also from the perspective of intellectual property rights. In addition, industry self-discipline and standard formulation are also important. Possible directions include: establishing a filing and authorization platform for AI-created content, which is similar to digital copyright registration and could serve as preliminary evidence of ownership when defending rights. At the same time, the platform can act as an intermediary for licensing transactions, facilitating the convenient conclusion of licensing agreements between content users and rights holders, and simplifying the current disorderly circulation of copyright of AI works. Another industry effort may be developing specification guidelines for the use of AI training data and output. There can be industry conventions that stipulate that AI-generated works retain the original logo or signature information when disseminating and shall not be deleted or modified without authorization. With the popularity of open source of deep learning models, there is also a need for community-level protocol support to clarify the rights framework for model training data and output. It is foreseeable that industry standards and platform mechanisms will play a preemptive role outside of official legislation, and can also be elevated to legal rules once certain practices prove to be effective. Through the dual track of legislation and industry standards, the AIGC copyright ecosystem can become more transparent, orderly, minimize disputes, and encourage creation.

5. Conclusion

This paper has examined the copyright protection paradigm for AI-generated content through a staged analysis of the generative process, focusing on originality, attribution, and infringement. The analysis demonstrates that while generative AI poses significant challenges to traditional copyright frameworks, existing legal principles can be adapted to accommodate this new technology by focusing on human creative contributions.

The originality of text prompts serves as a critical anchor for establishing the copyrightability of AI outputs. As shown in the comparative analysis, although jurisdictions apply varying standards of originality, a well-crafted prompt that reflects deliberate creative choices can meet the threshold for copyright protection. This, in turn, provides a basis for linking human input to machine-generated output, thereby addressing the authorship gap created by AI's autonomous capabilities.

Attribution in AI-generated works remains complex due to the involvement of multiple stakeholders—users, developers, and data providers. The proposed multi-factor model, which prioritizes the degree of creative control and contribution, offers a pragmatic approach to determining ownership. The “user first” principle, as evidenced in recent jurisprudence like the Beijing Li case, provides a starting point, but must be tempered by an assessment of the actual creative input at both the input and output stages. Joint authorship should be considered where multiple parties make substantial contributions.

On infringement, the divide between AI-vs-AI and AI-vs-human works clarifies the application of traditional tort and copyright doctrines. The fair use defense, particularly in the U.S., and exceptions for text and data mining in the EU, offer flexible mechanisms to balance innovation with rights protection. However, as litigation proliferates, clearer guidelines and industry standards are needed to ensure predictability and fairness.

In conclusion, a balanced copyright framework for the generative AI era should recognize protectable human expression in the AI process, allocate rights based on creative contribution, and employ flexible exceptions to foster both innovation and rights protection. Future efforts should focus on harmonizing approaches across jurisdictions, clarifying evidence rules for infringement cases, and promoting industry-led solutions alongside legislative reforms. By doing so, the law can keep pace with technological advancement while safeguarding the interests of all creators.

References

- [1] Oppenlaender, J. (2022). The Creativity of Text-to-Image Generation. In *The 25th International Academic Mindtrek Conference* (pp. 256-264). ACM. <https://dl.acm.org/doi/10.1145/3569219.3569352>
- [2] Wu, T., He, S., Liu, J., Sun, S., Liu, K., & Han, Q.-L. (2023). A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development. *IEEE/CAA Journal of Automatica Sinica*, 10(5), 1-15.
- [3] Ginsburg, J. C. (1992). No Sweat Copyright and Other Protection of Works of Information after Feist v. Rural Telephone. *Columbia Law Review*, 92(2), 338-388.
- [4] Rosati, E. (2018). Why Originality in Copyright Is Not and Should Not Be a Meaningless Requirement. *Journal of Intellectual Property Law & Practice*, 13(9), 724-735. <https://doi.org/10.1093/jiplp/jpy084>
- [5] Rahmatian, A. (2013). Originality in UK Copyright Law: The Old "Skill and Labour" Doctrine Under Pressure. *IIC - International Review of Intellectual Property and Competition Law*, 44(1), 4-34.
- [6] Mazzi, F. (2024). Authorship in artificial intelligence-generated works: Exploring originality in text prompts and artificial intelligence outputs through philosophical foundations of copyright and collage protection. *Journal of World Intellectual Property*, 27(3), 410-427.
- [7] Samuels, E. (1988). The Idea-Expression Dichotomy in Copyright Law. *Tennessee Law Review*, 56(2), 321-364.
- [8] Deltorn, J.-M., & Macrez, F. (2018). Authorship in the Age of Machine Learning and Artificial Intelligence. Center for International Property Studies.
- [9] Nordemann, J. B. (2019, November 21). AIPPI: No Copyright Protection for AI Works without Human Input, but Related Rights Remain. Kluwer Copyright Blog. <http://copyrightblog.kluweriplaw.com/2019/11/21/aippi-no-copyright-protection-for-ai-works-without-human-input-but-related-rights-remain/>
- [10] Liu, V., & Chilton, L. B. (2022). Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-23).
- [11] Chang, M., Druga, S., Fiannaca, A. J., Vergani, P., Kulkarni, C., Cai, C. J., & Terry, M. (2023). The prompt artists. In *Proceedings of the 15th Conference on Creativity and Cognition* (pp. 75-87). Association for Computing Machinery. <https://doi.org/10.1145/3591196.3593515>
- [12] Senftleben, M., & Buijelaar, L. (2020). Robot Creativity: An Incentive-Based Neighboring Rights Approach. *IIC - International Review of Intellectual Property and Competition Law*, 51(5), 553-581.
- [13] Guadamuz, A. (2021). Do Androids Dream of Electric Copyright? In *Artificial Intelligence and Intellectual Property*. Oxford University Press.
- [14] European Commission. (2020, June 7). Trends and Developments in Artificial Intelligence - Challenges to the Intellectual Property Rights Framework. <https://digital-strategy.ec.europa.eu/en/library/trends-and-developments-artificial-intelligence-challenges-intellectual-property-rights-framework>
- [15] Goldstein, P., & Hugenholtz, P. B. (2019). *International Copyright Law: Principles, Law, and Practice* (4th ed.). Oxford University Press.
- [16] Senftleben, M., & Buijelaar, L. (2020). Robot Creativity: An Incentive-Based Neighboring Rights Approach. *IIC - International Review of Intellectual Property and Competition Law*, 51(5), 553-581.
- [17] Giuffrida, I. (2019). Liability for AI Decision-Making: Some Legal and Ethical Considerations. *European Journal of Risk Regulation*, 10(1), 41-48.
- [18] Lynch, H. F., Vayena, E., & Gasser, U. (Eds.). (2018). *Big Data, Health Law, and Bioethics*. Cambridge University Press.
- [19] Heverly, R. (2020). More Is Different: Liability of Compromised Systems in Denial of Service Attacks. *Harvard Journal of Law & Technology*, 33(2), 567-610.
- [20] Dusollier, S., Kretschmer, M., Margoni, T., Mezei, P., Quintais, J. P., & Rognstad, O.-A. (2025). Copyright and Generative AI: Opinion. *JIPITEC*, 16(1), 121-127.