

Fair Use and AI Training: A Comparative Study between the U.S. and Uzbekistan

Abdusattorov Shokhjakhon Jurabek ugli

Penn state law LL.M, Legal Assistant, Pennsylvania, USA

Article information:

Manuscript received: 28 Apr 2025; **Accepted:** 27 May 2025; **Published:** 30 Jun 2025

Abstract: This article provides a comparative legal analysis of how the United States and Uzbekistan approach the issue of copyright law in the context of artificial intelligence (AI) training. As AI models increasingly rely on large datasets—often containing copyrighted material—questions about the legality of using such data without authorization have become central to intellectual property debates. The study explores the U.S. doctrine of fair use, which offers flexibility and potential legal protection for AI developers, contrasting it with Uzbekistan’s more restrictive copyright regime that lacks broad exceptions for innovative or transformative uses. By examining legal frameworks, case law, policy gaps, and technological implications, the article highlights the opportunities and challenges each country faces. The analysis concludes with recommendations for legal reform in Uzbekistan to better accommodate AI development while preserving the rights of creators, suggesting that harmonizing with international best practices could foster both innovation and intellectual property protection.

Keywords: Artificial Intelligence (AI), AI training data, copyright law, intellectual property, fair use, United States law, Uzbekistan law, comparative legal analysis, machine learning, legal reform, copyright protection, data mining, generative AI, U.S. Copyright Office, restrictive copyright regimes, digital innovation, AI and law, copyright exceptions, emerging technologies, legal frameworks.

Introduction

The rapid evolution of artificial intelligence (AI) has brought transformative changes to content creation, automation, and knowledge processing. At the core of many AI models—especially generative ones—lies the use of large datasets often consisting of copyrighted texts, images, music, and other works. These training datasets, scraped or collected from the internet, raise a fundamental legal question: **can copyrighted material be used to train AI models without permission?**

In the United States, this question is being explored within the legal doctrine of **fair use**, which allows limited unauthorized use of copyrighted works under specific conditions. By contrast, **Uzbekistan’s copyright regime**, rooted in civil law traditions and post-Soviet legal structure, provides **narrower exceptions**, often requiring explicit authorization or strict conditions for lawful use. This divergence offers a unique opportunity to analyze how two legal systems—one flexible and precedent-driven, the other more formal and codified—are addressing the legal and ethical implications of training AI on copyrighted works.

This article aims to provide a **comparative legal analysis** of the U.S. and Uzbek approaches to AI

training using copyrighted data. The first half explores the U.S. legal concept of fair use and how it is being interpreted in ongoing cases involving AI developers. The second half will examine Uzbekistan's copyright law, highlighting the limits of its exceptions and the challenges it faces in addressing AI-related issues.

Research and Methodology

This study employs a comparative legal research methodology to examine the concept of "fair use" in the context of artificial intelligence (AI) training, with a specific focus on the legal frameworks of the United States and Uzbekistan. The research is structured around doctrinal analysis, statutory interpretation, and case law evaluation, supplemented by secondary sources such as academic commentaries, policy papers, and scholarly articles.

The U.S. portion of the study involves an in-depth analysis of Section 107 of the U.S. Copyright Act, judicial interpretations by U.S. courts—particularly landmark cases such as *Authors Guild v. Google* and *Andy Warhol Foundation v. Goldsmith*—and recent litigation involving AI technologies (e.g., *The New York Times v. OpenAI*). It also considers how the flexible, four-factor test of fair use is being applied to emerging technologies and generative AI tools.

For Uzbekistan, the research draws upon national copyright legislation, including the Law on Copyright and Related Rights, as well as government policy documents and official statements. Due to the relatively nascent stage of AI-related copyright regulation in Uzbekistan, the study also incorporates interviews with legal scholars and policy experts, and it explores the potential for legal reform based on global trends.

By comparing these two jurisdictions, the research aims to identify both commonalities and divergences in how fair use and copyright limitations are interpreted in the context of machine learning. The findings are intended to inform legal scholars, policymakers, and technologists about the challenges and opportunities of AI training under different legal systems.

I. Understanding Fair Use in U.S. Copyright Law

The U.S. Copyright Act of 1976 codifies **fair use** under **17 U.S.C. § 107**, allowing the use of copyrighted works without permission for purposes such as criticism, comment, news reporting, teaching, scholarship, or research. The statute identifies four key factors for determining whether a specific use qualifies as fair:

1. **The purpose and character of the use**, including whether it is of a commercial nature or for nonprofit educational purposes.
2. **The nature of the copyrighted work**.
3. **The amount and substantiality** of the portion used in relation to the copyrighted work as a whole.
4. **The effect of the use upon the potential market** for or value of the copyrighted work.¹

U.S. courts have emphasized that these factors must be considered holistically and not in isolation. In recent years, courts have increasingly focused on whether the new use is "**transformative**"—i.e., whether it adds new meaning, expression, or function to the original work.

II. Fair Use in the Context of AI Training

Generative AI tools like **ChatGPT**, **DALL-E**, **Stable Diffusion**, and **Google's Gemini** are trained on massive datasets containing books, websites, images, code, and other materials, much of which is protected by copyright. The use of such data raises the critical question: does training an AI model constitute a "transformative" and therefore permissible use under the fair use doctrine?

There is **no settled case law** yet definitively addressing this question in the U.S., but several lawsuits

¹ U.S. Copyright Act of 1976, 17 U.S.C. § 107.

are currently underway. For example:

- In **Andersen v. Stability AI** and **Doe v. GitHub Copilot**, artists and programmers have sued AI companies for using their copyrighted content without authorization.
- Plaintiffs argue that mass scraping and training on their content violates their exclusive rights.
- Defendants argue that the AI systems do not reproduce the works themselves, but only **learn from the data** in a transformative and non-expressive manner, akin to reading.

Legal scholars and tech companies supporting fair use argue that the **training process is analogous to human learning**: humans read books and learn from them; they do not infringe copyright unless they copy the expression. According to this view, if an AI merely “learns” patterns from a dataset and generates new, unique content, then the original works are not being unlawfully used.

The **Authors Guild v. Google** (2015) case is often cited in support of this argument.² In that case, Google’s digitization and indexing of millions of books for search and research purposes was held to be fair use, partly because the use was transformative and non-commercially exploitative of the original works. Some legal analysts believe this sets a precedent that AI training could also be fair use.

However, the **commercial nature** of many AI products raises questions about whether this analogy holds. Tools like ChatGPT or Midjourney are monetized and sometimes produce outputs that **closely resemble** the input data. This could potentially undermine the fair use defense under the fourth factor—market harm to the original creator.

III. Copyright Exceptions in Uzbekistan: A More Restrictive Model

In contrast, **Uzbekistan’s copyright framework**, governed primarily by the **Law on Copyright and Related Rights** (1996, as amended), follows the **civil law tradition**, where **exceptions and limitations are strictly defined by statute**. Uzbekistan does **not recognize a general doctrine of fair use**. Instead, it relies on a **closed list** of permissible uses (similar to the “three-step test” under international law), such as:

- ✓ Use for personal purposes;
- ✓ Use for education and research within strict limits;
- ✓ Quotations with proper attribution;
- ✓ Library and archival reproduction;
- ✓ Parody and news reporting (to a limited extent).

The **absence of a flexible fair use doctrine** means that training AI models on copyrighted data **without explicit authorization** is likely to be considered unlawful under Uzbek law. Moreover, Uzbekistan’s IP system lacks developed case law, and judicial interpretation of new technologies is still limited.

As of 2025, there is **no formal policy or legislation** in Uzbekistan specifically addressing the legality of using copyrighted materials for training AI models. This legal vacuum creates uncertainty for tech companies, researchers, and rights holders, particularly as Uzbekistan tries to modernize its digital economy and foster innovation through government-led AI initiatives.

The contrast between the U.S. and Uzbekistan reveals two fundamentally different approaches to handling the legal challenges posed by AI training data. The U.S. system allows for a flexible interpretation of fair use, which could potentially accommodate innovative uses like AI training, provided certain conditions are met. Uzbekistan, on the other hand, remains bound by a rigid statutory framework that does not yet recognize the legal or technological nuance of machine learning and generative AI.

² Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

In the next section, we will explore the **implications** of these differences for policymakers, content creators, and tech companies in both countries. We will also propose **recommendations for harmonizing legal frameworks**, improving AI governance, and ensuring a balance between innovation and copyright protection.

IV. Legal and Technological Implications of Divergent Copyright Doctrines

The divergence between U.S. and Uzbek copyright laws has significant implications for the **development of AI technologies, international cooperation, and the digital economy.**

1. Innovation and Technological Growth

In the U.S., the flexible doctrine of fair use enables technology companies and startups to experiment with AI applications without seeking prior permission for every piece of data used in training. This approach **reduces legal friction** and fosters an innovation-friendly environment. Companies such as OpenAI, Google, and Meta have capitalized on this framework to develop world-leading AI systems.

In contrast, Uzbekistan's restrictive copyright system **limits the scope for experimentation.** The absence of legal exceptions for AI training creates **legal uncertainty** and potential liability for developers. Without explicit provisions recognizing the legitimacy of AI training on data corpora, researchers and tech companies risk copyright infringement lawsuits or regulatory pushback. This may stifle innovation and discourage foreign investment in Uzbekistan's tech sector.

2. Data Availability and Licensing Models

The U.S. system encourages the development of **open datasets** and licensing schemes such as Creative Commons, which allow AI developers to train their models on legally permissible content. Even when content is copyrighted, companies often invoke fair use as a shield against liability.³

In Uzbekistan, due to a more cautious legal climate, data availability for AI training is often **limited to government-sanctioned or public domain content**, unless licenses are explicitly granted. The lack of local licensing models or collective management organizations that offer access to copyrighted works for AI purposes further restricts AI development.

3. International Cooperation and Legal Fragmentation

Global AI development depends on **cross-border data flows**, but legal inconsistency creates barriers. AI models trained under U.S. fair use principles may not be legally deployable in countries like Uzbekistan, where similar training could be considered illegal. This fragmentation complicates the international rollout of AI technologies, especially when developers must **customize their legal strategies** for each jurisdiction.

For instance, a company that trains an AI under U.S. law and seeks to commercialize the product in Uzbekistan may face claims of copyright infringement if outputs or training processes involve protected Uzbek content. Without harmonization, this legal conflict can hinder technological exchange and cooperation between countries.

V. Challenges Specific to Uzbekistan

Despite aspirations for digital modernization, Uzbekistan faces several **domestic challenges** in integrating AI into its legal framework:

Uzbekistan's copyright law was initially drafted in the mid-1990s and has undergone only minor amendments. It does not account for modern technological developments like data mining, algorithmic generation, or machine learning. The law defines "reproduction" and "use" in ways that can be **broadly interpreted to prohibit AI training**, even when no direct copying of output occurs.

Limited Judicial and Regulatory Capacity courts and regulatory bodies in Uzbekistan have limited experience with emerging technologies. Judges may lack the technical understanding required to

³ U.S. Copyright Office, *Copyright and Artificial Intelligence*, August 2023, <https://www.copyright.gov/ai/>.

assess complex AI-related claims, especially those involving indirect use of copyrighted content in training datasets.

Moreover, the **Intellectual Property Agency of Uzbekistan** has not yet issued guidelines or policies on AI training, further contributing to uncertainty.⁴

Lack of Public and Institutional Awareness Public understanding of how AI interacts with copyright is low, and legal professionals, creators, and developers often have limited awareness of their rights and obligations. There is also a **lack of institutional coordination** between digital development agencies and IP regulators.

VI. Recommendations for Harmonization and Policy Reform

To bridge the gap between legal protection and innovation, especially in the context of AI, several steps can be taken:

1. Introduce Flexible Exceptions for AI Training in Uzbek Law

Uzbekistan should consider amending its copyright legislation to include **technology-neutral exceptions** for the use of copyrighted content in AI training. One possible model is the **EU's exception for text and data mining (TDM)** under the DSM Directive, which allows AI researchers to process data unless rights holders explicitly opt out.

2. Establish Clear Regulatory Guidelines

The Intellectual Property Agency of Uzbekistan, in coordination with the Ministry of Digital Technologies, should publish **guidelines on the lawful use of data for AI training**, clarifying what constitutes infringement and what does not. This would reduce ambiguity and provide a roadmap for innovators.

3. Develop Licensing Frameworks for AI Use

Uzbekistan could also benefit from creating **collective licensing models** where rights holders can voluntarily allow their content to be used for AI training in exchange for fair compensation. This balances the interests of creators and developers.

4. Encourage International Collaboration

To harmonize with global norms, Uzbekistan should participate in **WIPO-led discussions on AI and IP**, and study the legal systems of countries like the U.S., U.K., and Germany to adopt **hybrid models** that protect authors while enabling AI growth.

5. Invest in Capacity Building and Education

Training judges, policymakers, lawyers, and developers on AI and IP issues is essential for developing a responsive legal ecosystem. Workshops, legal clinics, and joint university programs can help bridge the knowledge gap.

Results

The comparative analysis between the United States and Uzbekistan reveals significant legal and structural differences in the treatment of fair use in the context of AI training.

In the **United States**, the fair use doctrine remains flexible and adaptable to technological change. The courts rely on a four-factor balancing test that allows AI developers to argue that the use of copyrighted works for machine learning constitutes transformative use. Several U.S. cases, such as *Authors Guild v. Google*, have established that mass copying for non-expressive purposes like indexing or training algorithms may fall within the scope of fair use. Moreover, recent lawsuits against AI companies are actively shaping the contours of how fair use applies to generative models.

In contrast, **Uzbekistan's copyright law** does not explicitly recognize a doctrine equivalent to fair use.

⁴ Uzbekistan Law "On Copyright and Related Rights," No. 42-I of July 20, 1996 (as amended through 2020).

Instead, it operates through a system of narrowly defined exceptions and limitations. These are primarily limited to educational, personal, or informational use and do not account for the complexity of AI training processes. There is currently no case law or formal policy guidance addressing the use of copyrighted works for AI-related purposes. As a result, developers and researchers in Uzbekistan face considerable legal uncertainty, and there is a risk that unlicensed data use may be treated as infringement, regardless of purpose.

Despite these differences, the research identified a shared need for legal modernization. In both jurisdictions, stakeholders are calling for clearer regulations on how copyrighted data may be used to train AI systems. In Uzbekistan, legal scholars and policymakers express cautious openness to adopting international best practices, particularly from jurisdictions like the U.S. and the EU.

Overall, the findings underscore a key challenge: while the U.S. offers a relatively favorable legal climate for AI training under fair use, Uzbekistan's legal system remains underdeveloped in this area, necessitating urgent reforms to support innovation while protecting copyright holders' rights.

Conclusion

The comparison between the United States and Uzbekistan reveals how different legal traditions approach the intersection of copyright and AI training. While the U.S. leverages its fair use doctrine to support AI development and innovation, Uzbekistan's stricter legal framework places greater emphasis on author protection, often at the expense of flexibility.

As AI continues to revolutionize how content is created, understood, and consumed, countries like Uzbekistan must **rethink their copyright regimes**. Without adapting to the realities of AI, Uzbekistan risks falling behind in the global digital economy. However, with targeted legal reform, cross-sector dialogue, and international cooperation, it is possible to build a copyright system that both **respects authors' rights** and **fosters technological progress**.

References

1. U.S. Copyright Act of 1976, 17 U.S.C. § 107.
2. Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).
3. U.S. Copyright Office, *Copyright and Artificial Intelligence*, August 2023, <https://www.copyright.gov/ai/>.
4. Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith, 598 U.S. ___ (2023).
5. Thaler v. Perlmutter, No. 22-1564 (D.D.C. 2023).
6. Uzbekistan Law "On Copyright and Related Rights," No. 42-I of July 20, 1996 (as amended through 2020).
7. World Intellectual Property Organization (WIPO), *Uzbekistan's IP Legal Framework*, https://www.wipo.int/members/en/details.jsp?country_id=195.
8. European Parliament and Council Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market.
9. Michael Geist, "Fair Use in the Age of AI," *Journal of Intellectual Property Law & Practice*, Vol. 18, No. 1, 2023, pp. 25-34.
10. J. Litman, *Digital Copyright*, Prometheus Books, 2006.
11. Ministry of Digital Technologies of the Republic of Uzbekistan, "Strategy for Digital Uzbekistan 2030," 2021.
12. WIPO Conversation on Intellectual Property and Artificial Intelligence, 2023, https://www.wipo.int/meetings/en/details.jsp?meeting_id=73167.n

13. S. Dusenbury, "AI Training and Copyright Reform: An International Overview," *Harvard Journal of Law & Technology*, Vol. 37, 2024.
14. Creative Commons, "Licensing for Machine Learning," <https://creativecommons.org/>.
15. UNESCO, "Ethics of Artificial Intelligence: Guidelines for Policymakers," 2021.