

THE LEGALITY OF ARTIFICIAL INTELLIGENCE'S UNAUTHORIZED USE OF COPYRIGHTED MATERIALS UNDER CHINA AND U.S. LAW

YUDONG CHEN*•

ABSTRACT

This article is the first to discuss the legal regulation of artificial intelligence on the fair use of unauthorized materials from the perspective of China and the United States. Artificial intelligence is an important part of the future industrial revolution, but the problem behind it is the various lawsuits caused by its large import of unauthorized materials. The lack of reasonable regulation in this regard in the copyright laws of China and the United States has seriously hindered the development of artificial intelligence. In this article, after comparing the different legal provisions on fair use in China and the United States, the author finds that the enumeration

* IP Fellow, IIT Chicago-Kent College of Law. For their comments and views, I would like to thank Edward Lee, Graeme B. Dinwoodie, Mark. A. Lemley, Peter K. Yu, Benjamin L. W. Sobel, Patrick Goold, Xiaoben Xu, Yinan Yang, An Li, Mauritz Kop, and LLM classmates from various countries who have given me many suggestions and criticisms. I would especially like to thank my paper advisor, Edward Lee, for his advice on writing my dissertation time and time again, giving countless novel revisions, and finding and printing many related papers for my reference. At the same time, I would also like to thank my domestic master supervisor, Zhang Jian, a professor in China, who helped me review the Chinese legal studies part of this paper and gave me a lot of support and comments.

• IDEA Vol. 63 thanks Yudong Chen for participating in IDEA's Fall 2022 Symposium "IP in the Metaverse: Protecting IP Rights in the Virtual World."

exceptions method in China’s fair use does not include the regulation of artificial intelligence. Moreover, under U.S. law, when the general fair use four factors apply to the unauthorized use of copyright materials by artificial intelligence, some of the factors are obviously very unfavorable and unfair to artificial intelligence. Therefore, the author proposes a synthetic “fair use” model to be enacted by WIPO, which combines the “Four-Factor Standard” and “AI Exceptions” for artificial intelligence’s unauthorized use of copyrighted materials.

Introduction.....	244
I. Whether Artificial Intelligence’s Use of Copyrighted Materials Is Fair Use or Permissible under the Copyright Laws of China and the U.S.	247
A. The Practice of Artificial Intelligence Unauthorized Use of Copyrighted Materials.....	247
1. Replication Is a Prerequisite for Machine Learning	248
2. Creating Independent Works Is the Ultimate Meaning of AI.....	251
B. Comparing the Copyright Laws of China and the U.S.	254
1. The U.S. Copyright Law: Four-Factor Standard of Fair Use.....	254
2. China’s Copyright Law: Thirteen Enumerated Copyright Exceptions.....	258
C. The Need to Clarify the Status of Artificial Intelligence’s Unauthorized Use.....	261
II. A Synthetic “Fair Use” Model in WIPO for Artificial Intelligence’s Unauthorized Use of Copyrighted Materials	262

The Legality of Artificial Intelligence's Unauthorized Use of Copyrighted Materials under China and U.S. Law **243**

- A. Proposal of the Synthetic “Fair Use” Model: Four-Factor Standard + AI Exceptions..... 263
 - 1. Draft of the Synthetic “Fair Use” Model 263
 - 2. A Combination of Enumeration and Four-Factor Standard 264
- B. Similarities and Differences between Synthetic “Fair Use” Model and Fair Learning 271
- C. Reasons for Adopting Proposal 272
 - 1. Promote the Development of AI..... 273
 - 2. Avoid Unfair Monopoly and Competition of Data Materials 273
 - 3. Harmonization 274
- III. Criticisms on A Synthetic “Fair Use” Model for Artificial Intelligence 275
 - A. Intellectual Property Rights of Creators of Copyrighted Materials 275
 - B. What if the AI Use Fails the Enumerated Exceptions but Satisfies Fair Use?..... 277
 - C. Disturbs the Original Order of Copyright Law... 277
- Conclusion 278

INTRODUCTION

Suppose that you are a famous economist in the United States.¹ One day, you discover that an artificial intelligence “X” from a tech company has published a series of articles on economic analysis using your data, and the logic of its economic analysis is very similar to yours. Later, you are told that the tech company admitted to copying your writings and using them in machine learning to make the artificial intelligence “X” more economically analytical. You try to assert your intellectual property rights through litigation, but in the U.S. you find no clear legal basis or precedent for artificial intelligence’s unauthorized use of copyrighted materials to rely on.² If the same case were to come to China, there would still be no way to resolve the issue under China’s copyright law on fair use, since China’s unique enumeration method does not include a clause on artificial intelligence machine learning.³ Thus, the following problem arises: “[c]opyright law forces artificial intelligence into a binary: it is either a mystical author or a dumb machine.”⁴

Artificial intelligence (“AI”) is an activity that makes machines intelligent, and the key to that is machine

¹ See Anjuli Kannan et al., *Smart Reply: Automated Response Suggestion for Email*, in KKD ‘16: PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING 955–64 (2016), <https://dl.acm.org/doi/pdf/10.1145/2939672.2939801> [<https://perma.cc/W8LG-KE8W>], for the research forming the basis of this hypothetical.

² Benjamin L. W. Sobel, *Artificial Intelligence’s Fair Use Crisis*, 41 COLUM. J.L. & ARTS 45, 48–49 (2017).

³ See Copyright Law of the People’s Republic of China (promulgated by the Standing Comm. Nat’l People’s Cong., Sept. 7, 1990, rev’d Nov. 11, 2020, effective June 1, 2021) art. 24, 2020 P.R.C. Laws (China) [hereinafter Article 24], <http://www.npc.gov.cn/englishnpc/c23934/202109/ae0f0804894b4f71949016957eec45a3.shtm> 1 [<https://perma.cc/5TVW-KM7V>] (last visited Jan. 24, 2023).

⁴ Sobel, *supra* note 2, at 49.

learning. Machine learning is when machines “learn” to emulate human work product from data and “training data” by inputting and analyzing huge amounts of information from data models.⁵ In this stage, information, such as copyrighted works, is completely copied and even changed countless times as input data.⁶ This action may infringe the right of reproduction or compilation of the exploited works.⁷ The question of whether AI analyzing “training data” necessarily constitutes copyright infringement or whether the action can be covered by fair use regimes concerns the future development of AI technology.⁸ More clearly, “[g]iven the doctrinal uncertainty and the rapid development of [machine learning] technology, it is unclear whether machine copying will continue to be treated as fair use.”⁹ Too strict copyright protection can hinder the development of AI technology. Lax copyright protection may change the pattern of copyright interests and inhibit the creative motivation of human authors.

In China, the answer to this issue is unclear, and there is no specific law to provide for it. In China, AI’s machine learning is not covered by the types of fair use behavior stipulated in copyright law and regulations on the protection of information network transmission.¹⁰ In the

⁵ Xiaoben Xu (徐小奔) & Yinan Yang (杨依楠), *Lùn Réngōng Zhīnéng Shēndù Xuéxí Zhōng Zhùzuòquán de Héilǐ Shǐyòng* (论人工智能深度学习中 著作权的合理使用) [On the Rational Use of Copyright in Artificial Intelligence Deep Learning], 3 S.J.T.U. L. REV. 32, 33 (2019) (China); Sobel, *supra* note 2, at 48.

⁶ Sobel, *supra* note 2, at 61.

⁷ *Id.*

⁸ See Peter K. Yu, *Artificial Intelligence, the Law-Machine Interface, and Fair Use Automation*, 72 ALA. L. REV. 187, 193–94 (2020).

⁹ Mark. A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 746 (2021).

¹⁰ Xu & Yang, *supra* note 5, at 40; Yang Gao et al., *Research on the Dilemma and Improvement of the Copyright Fair Use Doctrine*

U.S., although Google Images and Google Books are excused under the fair use doctrine as services supported by unauthorized reproduction of protected expressions,¹¹ machine learning is a new legal issue. Thus, the legality of AI's unauthorized use of copyrighted materials is not clear under China and the U.S. law.

This Article aims to analyze and clarify the legality of AI's unauthorized use of copyrighted materials under the current different copyright laws of the two countries. This Article proposes that the use of AI on unauthorized materials should be included in the category of fair use. This is in line with copyright law's goal to stimulate innovation and is also conducive to the healthy development of the artificial intelligence industry.

Part I of the article provides an overview of AI's unauthorized use of copyrighted materials and real cases in China and the U.S. Part I presents and analyzes the practice of AI's unauthorized use of copyrighted materials, compares the copyright laws of China and the U.S. and finally explains the need to clarify the status of AI's unauthorized use. Part II provides a synthetic fair use model to combine the two different characteristics of fair use regulations in China and the United States and proposes that an amendment be added by WIPO to facilitate the resolution of international disputes related to the use of unauthorized materials by AI. Part III responds to potential criticism of this proposal.

Related to Machine Learning in China, 22 U.I.C. REV. INTELL. PROP. L. 1, 17 (2022).

¹¹ Sobel, *supra* note 2, at 48.

I. WHETHER ARTIFICIAL INTELLIGENCE'S USE OF COPYRIGHTED MATERIALS IS FAIR USE OR PERMISSIBLE UNDER THE COPYRIGHT LAWS OF CHINA AND THE U.S.

AI may face the dilemma of copyright infringement in the process of analyzing data.¹² In particular, all machine learning requires the aggregation and input of data beforehand and the formation of digital copies of that data in an AI system.¹³ If the input data contains a large number of works not authorized by the copyright owner, such behavior may constitute infringement of the copyright owner's right to copy.¹⁴ For AI industry players, who often input tens of millions of pieces of data, the practice of obtaining prior permission obviously faces high transaction costs. Once input data is identified as infringement, damages for thousands of works may even bankrupt the AI industry.¹⁵ This Part compares the current copyright laws and judicial practices in China and the U.S. regarding AI learning from copyrighted materials and the need for greater clarification to address this issue under both countries' laws.

A. *The Practice of Artificial Intelligence Unauthorized Use of Copyrighted Materials*

A brief overview of AI's unauthorized use of the specific meaning and processes of copyrighted materials is helpful to analyze and present the possible conflicts with

¹² *Id.*

¹³ Patrick Goold, *Artificial Authors: Case Studies of Copyright in Works of Machine Learning*, 67 J. COPYRIGHT SOC'Y U.S.A. 427, 432–33 (2020).

¹⁴ Mauritz Kop, *The Right to Process Data for Machine Learning Purposes in the EU*, 34 HARV. J.L. & TECH. 1, 5 (2021).

¹⁵ Xu & Yang, *supra* note 5, at 40; Sobel, *supra* note 2, at 80.

copyright law in machine learning, i.e., the data input process. The practical significance of AI to help human beings lies in understanding the inner ideas of works and forming new ideas to realize the updating of knowledge.¹⁶ This occurs through two major processes—data input and data output.¹⁷ However, the use of unauthorized materials by AI mentioned in this paper is limited to the process of data input.¹⁸

1. Replication Is a Prerequisite for Machine Learning

Massive reproduction or interpretation of other works is the premise of AI editing technology development and application, so the replication and input of data is the first step for AI to create independent works.¹⁹ AI got its start at the Dartmouth Workshop in 1956.²⁰ The core of the new generation of AI technology is machine learning characterized by data-training algorithms.²¹ The raw

¹⁶ See Sobel, *supra* note 2, at 57–58.

¹⁷ See *id.* at 65.

¹⁸ See *id.* (“Legal concerns may persist after the development of a model. After all, protectable input data are commonly used to train models to generate similar output. If that similarity is ‘substantial,’ then that output may infringe copyright in the pre-existing work or works to which it is similar—or, at least, it could be found infringing if it were rendered by a human . . . while machine learning is developing quickly, its outputs have not yet supplanted works of human authorship.”). Therefore, this paper will not take the use of unauthorized materials by artificial intelligence as the main angle of research.

¹⁹ Kop, *supra* note 14, at 9.

²⁰ *Dartmouth Workshop*, WIKIPEDIA, https://en.wikipedia.org/wiki/Dartmouth_workshop [https://perma.cc/2MMN-HKMU] (last visited Nov. 16, 2021).

²¹ Li An (李安), *Jīqì Xuéxí Zuòpǐn Zhùzùdàquán Fǎ Fēnxī*——*Bù Shìyòng, Héli Shìyòng Hé Qīnquán Shìyòng (机器学习作品的著作权法分析—非作品性使用、合理使用与侵权使用)* [Copyright Law Analysis of Machine Learning Works–Nonuse, Fair Use and

material for machine learning is data; AI can analyze or execute commands autonomously based on the composition of internal data.²² AI is a subversive technology, and its text editing function opens up a new trend for the development of cultural industry in the future.²³

Further speaking, AI editing transcends the human path from information to knowledge and then to creation based on will and creates a technical path from data to knowledge and then to editing based on algorithm.²⁴ To be more specific, AI first converts the knowledge acquired by human eyes into a data model, thus breaking away from the learning and analysis mode of human mind in the first step, and then digesting the data model or superposing different data models to make it the internal knowledge of machine learning.²⁵ Then, AI analyzes the habituation rules set by human beings to edit and produce new products needed by human beings.²⁶

However, AI does not have neurons like human beings, so there is no innovation consciousness and brain activity.²⁷ Therefore, if AI is not given a certain amount of

Infringing Use], 6 Diànzǐ Zhīshì Chǎnquán (电子知识产权·) [ELEC. INTELL. PROP.] 60, 61 (2020) (China); Yuji Roh et al., *A Survey on Data Collection for Machine Learning: A Big Data- AI Integration Perspective*, 33 IEEE TRANSACTIONS ON KNOWLEDGE & DATA ENG'G 1328, 1328 (2021).

²² See generally Daniel Faggella, *What is Artificial Intelligence? An Informed Definition*, EMERJ (Dec. 21, 2018), <https://emerj.com/ai-glossary-terms/what-is-artificial-intelligence-an-informed-definition/> [<https://perma.cc/GHD9-X7AY>].

²³ See generally Goold, *supra* note 13, at 469.

²⁴ Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, 53 U.C. DAVIS L. REV. 893, 959 (2020); Dan L. Burk, *Algorithmic Fair Use*, 86 U. CHI. L. REV. 283, 297 (2019).

²⁵ Carroll, *supra* note 24.

²⁶ See Burk, *supra* note 24, at 295–97.

²⁷ See James Vincent, *How three French students used borrowed code to put the first AI portrait in Christie's*, THE VERGE

original data or basic data, AI is just an ordinary machine with simple technology and without the characteristics of “intelligence.”²⁸ Consequently, in the process of obtaining original data, there will be a large number of copies of existing works in reality.²⁹ Not only that, but if you want AI to produce innovative work, one or two pieces of input will not suffice.³⁰ Therefore, the information input of an AI machine will not be satisfied with the filling of a small amount of data information. Instead, the AI machine will try to fill in a large amount of data information—in reality as much as possible—so that the works produced by the AI have higher independence.³¹ In the process, the modules that make AI gather and analyze knowledge information become more uniform because the AI can automatically acquire information and optimize its own output.³² For example, in 2015, Google added a feature called Smart Reply to its inbox email service.³³ Inbox mail messages automatically generated up to three responses which the user could select.³⁴ In the first iteration, the intelligent reply algorithm was “trained on a corpus of 238 million email messages, presumably sampled from Gmail accounts,” and Google even added the input of novel data in later developments to make it more conversational.³⁵

(Oct. 23, 2018, 9:34 AM), <https://www.theverge.com/2018/10/23/18013190/ai-art-portrait-auction-christies-belamy-obvious-robb-ie-barratgans> [<https://perma.cc/5HRV-BVEN>].

²⁸ James Grimmelman, *Copyright for Literate Robots*, 101 IOWA L. REV. 657, 658 (2016).

²⁹ Sobel, *supra* note 2, at 62.

³⁰ *Id.*

³¹ See Yu, *supra* note 8, at 213.

³² See *id.*

³³ Sobel, *supra* note 2, at 68.

³⁴ Kannan et al., *supra* note 1, at 963.

³⁵ Sobel, *supra* note 2, at 68.

2. Creating Independent Works Is the Ultimate Meaning of AI

The knowledge increment generated by machine learning of AI is the process of discovering new value beyond the original value of a work.³⁶ If the final product of AI is not completely independent, or if the final product is basically identical to the original product, then AI will still lose its characteristics of intelligence.³⁷ From the perspective of copyright law, the final product may fall into the category of infringement; detailed analysis will be presented later.³⁸

There are two kinds of knowledge increment in machine learning. One is when the type of works created by AI belong to the same type of original works input, and the other is when the data value is created by non-expressive use of works.³⁹ Both types are closely related to the data entry process, but the former is more relevant to the use of unauthorized materials; thus, this article discusses only the former.

For the former type, to be specific, AI can produce similar works of the same type after the AI performs a massive learning and analysis of specific works, but only

³⁶ Xu & Yang, *supra* note 5, at 36; *see also* Carroll, *supra* note 24, at 902–04.

³⁷ *See* Darrell M. West & John R. Allen, *How artificial intelligence is transforming the world*, BROOKINGS (Apr. 24, 2018), <https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/> [<https://perma.cc/W4P9-HW5Q>] (“Artificial intelligence algorithms are designed to make decisions, often using real-time data. They are unlike passive machines that are capable only of mechanical or predetermined responses . . . they combine information from a variety of different sources, analyze the material instantly, and act on the insights derived from those data.”).

³⁸ *See infra* Part II.

³⁹ Xu & Yang, *supra* note 5, at 36; Sobel, *supra* note 2, at 51–57, 61–67.

with works with similar types and independent contents.⁴⁰ For example, an AI product developed by Tencent in recent years is called Dreamwriter.⁴¹ Tencent first inputs a large amount of data of news articles into the machine, then makes Dreamwriter analyze various data in the database to generate corresponding writing techniques.⁴² After analyzing, Dreamwriter can write reports of news events associated with the database.⁴³ Although the news reports written by Dreamwriter belong to the same category as the original input works, they only use the writing method and analysis method of the original work, so the new work remains completely independent.⁴⁴ Dreamwriter also has been involved in lawsuits as a plaintiff against another company that copied an article written by Dreamwriter.⁴⁵ However, in the final judgment of the Chinese court, it is clearly mentioned that

The external performance of the article created by Dreamwriter meets the formal requirements of written works. The article's content reflects the selection, analysis and judgment of the relevant stock market information and data on the morning of that day. The article structure is proper and the logic of

⁴⁰ Xu & Yang, *supra* note 5, at 36; Sobel, *supra* note 2, at 51–57, 61–67.

⁴¹ See Alexander J. Wurzer, *Tencent is getting copyright for AI-dreamwriter article*, IP BUS. ACAD. (Mar. 25, 2021), <https://ipbusinessacademy.org/tencent-is-getting-copyright-for-ai-dreamwriter-article> [<https://perma.cc/B6TX-JF2A>].

⁴² Kan He, *Another Decision on AI-generated Work in China: Is it a Work of Legal Entities?*, IPKAT (Jan. 29, 2020), <https://ipkitten.blogspot.com/2020/01/another-decision-on-ai-generated-work.html> [<https://perma.cc/2C2R-KCK2>].

⁴³ Sarah Xuan, *Intellectual Property Protection for AI Creations in China*, HG.ORG, <https://www.hg.org/legal-articles/intellectual-property-protection-for-ai-creations-in-china-56196> [<https://perma.cc/498U-ZABV>] (last visited Jan. 26, 2023).

⁴⁴ He, *supra* note 42.

⁴⁵ *Id.*

The Legality of Artificial Intelligence's Unauthorized Use of Copyrighted Materials under China and U.S. Law 253

expression is clear, thus the involved article possesses a certain originality.⁴⁶

The court opinion is an indirect acknowledgment that independent articles created by AI machines are copyrighted.⁴⁷

Nevertheless, there are still many scholars who have expressed different opinions on this case, and who think that the AI process of creating a new article by learning another is unauthorized use of the article.⁴⁸ Presently, there is no law that stipulates this, and the court also did not use the fair use rules to analyze the AI behavior itself.⁴⁹ Thus, the question about the legality of AI's unauthorized use of copyrighted materials should be carried on with deep discussion.

⁴⁶ Shēnzhèn Shì Téngxùn Jìsuànjī Xìtǒng Yǒuxiàn Gōngsī Sù Shànghǎi Yíng Chuàng Kējì Yǒuxiàn Gōngsī (深圳市腾讯计算机系统有限公司诉上海盈创科技有限公司) [Shenzhen Tencent Computer System Co. LTD v. Shanghai Winsun Technology Co. LTD], 2019 Guangdong 0305 Minchu No. 14010, at 6–7 (Nanshan Dist. People's Ct. 2019) (China), translated in “Tencent Dreamwriter”: *Decision of the People's Court of Danshan (District of Shenzhen) 24 December 2019—Case No. (2019) Yue 0305 Min Chu No. 14010*, 51 INT'L REV. INTELL. PROP. & COMPETITION L. 652, 656 (Xiaoshuai Ren trans., 2020); see also *Chinese Court Backs Copyrights for AI-Created Works*, CHINA JUSTICE OBSERVER (Oct. 30, 2020), <https://www.chinajusticeobserver.com/a/chinese-court-backs-copyrights-for-ai-created-works> [<https://perma.cc/MWM4-XH4A>].

⁴⁷ He, *supra* note 42.

⁴⁸ Yīxiū Zhīshì Chǎnquán (一休知识产权) [Yixiu Intellectual Property], Quánguó Shǒu Lì AI Jīqìrén Zuòpǐn Zhùzuòquán Àn Xuānpàn, Téngxùn Shèngsù Huò Péi (全国首例AI机器人作品著作权案宣判·腾讯胜诉获赔) [*China's First AI Robot Work Copyright Case Was Pronounced, and Tencent Won the Lawsuit and Won the Compensation!*], ZHIHU (Jan. 5, 2020), <https://zhuanlan.zhihu.com/p/101224903> [<https://perma.cc/4J6M-RQYJ>].

⁴⁹ *Id.*

B. Comparing the Copyright Laws of China and the U.S.

By comparing the different legal systems for fair use in China and the U.S., the legal defects in the provision of AI's unauthorized use of copyrighted materials in both countries will become clearer. American copyright law uses four factors to stipulate the fair use system, while China uses the enumeration method to list thirteen cases of fair use in detail and acquiesces to the four-factor identification method to some extent.

1. The U.S. Copyright Law: Four-Factor Standard of Fair Use

The four fair use factors of § 107 of the United States Copyright Act of 1976 are stated as follows:

- (1) the purpose and character of the use . . . ;
- (2) the nature of the copyrighted work;
- (3) the amount and substantiality of the portion used . . . ; and
- (4) the effect of the use upon the potential market for or value of the copyrighted work.⁵⁰

When it comes to unauthorized use of copyrighted materials, do the U.S. fair use rules continue to be effective? It is clear from the previous description of the working process of AI that copying data is indispensable as the first step of machine learning.⁵¹ However, in the process of copying a large amount of data, unauthorized data must be included.⁵² Can this be exempted from fair use?

More specifically, the first fair use factor can be used to investigate the purpose and character of AI's

⁵⁰ 17 U.S.C. § 107.

⁵¹ See *supra* notes 24–31 and accompanying text.

⁵² *Id.*

massive input of unauthorized data.⁵³ In other words, the first factor is used to investigate whether the output works of AI are transformative.⁵⁴ Of course, a machine that merely copies the original without a transformative innovation would not be called AI as described in this article. This is because the purpose of AI machine learning is to be able to output a transformative work compared to the original data.⁵⁵ Similarly, in *Authors Guild v. Google, Inc.*, the Court held that “Google’s unauthorized digitizing of copyright-protected works, creation of a search functionality, and display of snippets from those works were non-infringing fair uses” because “[t]he purpose of the copying [was] highly transformative.”⁵⁶ Perhaps the copyright holders of the raw materials could argue that copying by AI machine learning is completely immutable copying. In *A.V. v. iParadigms, L.L.C.*, the plaintiff similarly objected that “iParadigms’ use of their works cannot be transformative because the archiving process does not *add* anything to the work.”⁵⁷ However, the court held that the difference in function and purpose between the defendant’s work and the plaintiff’s work, notwithstanding the fact that the defendant’s reproduction of the original work was not substantially altered, could not conceal its transformative nature.⁵⁸

The second factor is to examine the nature of the copying of data by AI machines.⁵⁹ As expressed in *Authors Guild*, “[o]ne cannot assess whether the copying work has

⁵³ 17 U.S.C. § 107(1).

⁵⁴ *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994).

⁵⁵ *See West & Allen, supra* note 37.

⁵⁶ *Authors Guild v. Google, Inc.*, 804 F.3d 202, 229 (2d Cir. 2015).

⁵⁷ *A.V. ex rel. Vanderhye v. iParadigms, L.L.C.*, 562 F.3d 630, 639 (4th Cir. 2009).

⁵⁸ *Id.*

⁵⁹ 17 U.S.C. § 107(2).

an objective that differs from the original without considering both works, and their respective objectives.”⁶⁰ In *Am. Geophysical Union v. Texaco Inc.*, the court stated in the second factor analysis of the fair use rule, “though we have previously recognized the importance of strong copyright protection to provide sufficient incentives for the creation of scientific works . . . nearly every category of copyrightable works could plausibly assert that broad copyright protection was essential to the continued vitality of that category of works.”⁶¹ Also, in *Harper & Row Publishers, Inc. v. Nation Enterprises*, the Supreme Court stated “[t]he law generally recognizes a greater need to disseminate factual works than works of fiction or fantasy.”⁶² Thus, the second factor is highly likely to favor AI fair use as long as it can produce factual works.

The third fair use factor should be considered the least friendly to AI machine learning because the factor to review includes the number of replicates, and AI machine learning requires plenty of unmodified replication data to create new works.⁶³ As Mark Lemley and Bryan Casey once wrote, “ML systems involve copying the entire work without alteration. That directly affects statutory factor number three, which weighs the fact that the entire work is taken against a finding of fair use.”⁶⁴ However, that factor doesn’t necessarily work against AI machine learning. For example, in *Google L.L.C. v. Oracle Am., Inc.*, “Google copied the declaring code for 37 packages of the Sun Java API, totaling approximately 11,500 lines of code.”⁶⁵ The

⁶⁰ 804 F.3d at 220.

⁶¹ *Am. Geophysical Union v. Texaco Inc.* 60 F.3d 913, 925 (2d Cir. 1994).

⁶² *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 563 (1985).

⁶³ 17 U.S.C. § 107(3); Sobel, *supra* note 2, at 51–57, 61–67.

⁶⁴ Lemley & Casey, *supra* note 9, at 763.

⁶⁵ *Google L.L.C. v. Oracle Am., Inc.* 141 S. Ct. 1183, 1204 (2021).

court held that copying the data was key to creating or improving new systems and said that without copying the data, it would have been difficult for those programmers to build Android smartphone systems.⁶⁶ Similarly, AI machine learning is inseparable from the replication of original data.⁶⁷ Without the replication of data, the creativity of artificial intelligence will die completely. Therefore, from this perspective, this factor may be beneficial to artificial intelligence.

The fourth fair use factor is likely to be highly controversial.⁶⁸ Administrators of AI machine learning might argue that AI's use of copyrighted work is confined to machines, with little connection to the public domain.⁶⁹ Therefore, "the system's use doesn't cut into the ordinary market for the copyrighted works in question."⁷⁰ But for copyright holders, an AI machine can be seen as a consumer who takes raw materials or raw data to produce profitable new work, regardless of the question of innovation, and should thus pay for the materials it takes.⁷¹

Mark Lemley and Bryan Casey argue that machine learning should be recognized as fair use.⁷² For the first factor, they think that if machine learning only "access[es], learn[s], and use[s] the unprotectable parts of the work," the use should be fair.⁷³ As for the second and third factors that have the greatest negative impact on fair learning, they believe that artificial intelligence should have the same learning rights as human beings by being able to learn

⁶⁶ *Id.* at 1205.

⁶⁷ Sobel, *supra* note 2, at 51–57, 61–67.

⁶⁸ 17 U.S.C. § 107(4); see Mark A. Lemley, *Property, Intellectual Property, and Free Riding*, 83 TEX L. REV. 1031, 1031–32 (2005).

⁶⁹ Lemley & Casey, *supra* note 9, at 765.

⁷⁰ *Id.*

⁷¹ *Id.*

⁷² *Id.* at 777.

⁷³ *Id.* at 776.

naturally by reviewing entire works.⁷⁴ As for the fourth element, they argue that machine learning's use of the original work does not interfere with the copyright owner's core market, and copyright owners do not create works for the purpose of selling them to AI.⁷⁵ Thus, this factor should generally not prevent fair use of machine learning.⁷⁶

Although America's four factors and transformative new rules—to a certain extent—may place AI's independent output of new products into the category of fair use, the legal system itself has a strong flexibility. As such, future courts may think AI machine learning is not transformative fair use because, in the process of data input, there exists the complete copying of works and production of similar works.⁷⁷

2. China's Copyright Law: Thirteen Enumerated Copyright Exceptions

The legislative method of China's copyright system is an enumeration.⁷⁸ Of course, part of this has to do with China's extensive adoption of the civil law system.⁷⁹ Civil law countries take legislation as the main body of law and strictly require judges to judge cases according to written legislation.⁸⁰ Article 24 of the Copyright Law stipulates 13 types of copyright exceptions but does not have a general

⁷⁴ *Id.* at 777, 785.

⁷⁵ Lemley & Casey, *supra* note 9, at 776–77.

⁷⁶ *Id.*

⁷⁷ *See* Yu, *supra* note 8, at 201–02.

⁷⁸ *See* Lian Yunze & Xiangjing Luo, *China adapts copyright law to reflect global business trends*, MANAGINGIP (Apr. 14, 2021), <https://www.managingip.com/article/2a5cyvdl53v4r4cjxboxes/china-adapts-copyright-law-to-reflect-global-business-trends> [https://perma.cc/KVS4-CZNE].

⁷⁹ *Does China Have Common Law?—China Law in One Minute*, CHINA JUSTICE OBSERVER (Nov. 9, 2020), <https://www.chinajusticeobserver.com/a/does-china-have-common-law> [https://perma.cc/R7VS-YCSX].

⁸⁰ *Id.*

exception for fair use.⁸¹ In practice, Chinese courts have one basic criterion for the determination of fair use behavior: it must be recognized as one of the 13 behavior modes in accordance with the provisions of Article 24 of the Copyright Law.⁸²

Obviously, in the 13 exceptions listed, it can be found that Article 24 gives judges a certain degree of discretion; thus, fair use in China is not completely rigid.⁸³ The Supreme People's Court of China also issued the Opinions on Several Issues concerning Giving Full play to the Role of Intellectual Property Adjudications to Promote the Great Development and Prosperity of Socialist Culture and The Independent and Coordinated Development of Economy (the "Opinions").⁸⁴ As stipulated in the Opinions, when considering the reasonable use of a copyright, courts should consider whether the use affects normal use of the works or damages the authors' legitimate interests, "as related to the nature and utili[z]ation of [the] works, nature of works in use, number and quality of works in use, and impact of use on potential market or value of works."⁸⁵ The standard of review has been relaxed to some extent and is no longer limited to the 13 cases mentioned above.⁸⁶

⁸¹ See Article 24, *supra* note 3.

⁸² Because China follows a civil law system where judges are required to decide cases according to written law, it follows that any fair use exception that is not listed in China's fair use legislation will not be considered fair use. See *id.*; see also *Does China Have Common Law?—China Law in One Minute*, *supra* note 79.

⁸³ See Article 24, *supra* note 3.

⁸⁴ *Full Review of the Opinions from the Supreme People's Court of China on Giving Full Play to the Functional Role of Intellectual Property Trials (Issued Dec 2011)*, HOGAN LOVELLS (Apr. 2012) [hereinafter *Opinions*], <https://www.lexology.com/library/detail.aspx?g=9d5c1150-442f-4d3a-b2a4-15225e3bd74d> [<https://perma.cc/CGH3-DHR5>].

⁸⁵ *Id.*

⁸⁶ See *id.*; see also Article 24, *supra* note 3.

If we analyze the above Opinions issued by the Supreme People's Court of China, we can find that its content is similar to the four factors of the Fair Use Rule in the United States. Therefore, there is no need to make too much interpretation of the Opinions. More specifically, the Opinions give courts the power to decide fair use like the U.S., not just the thirteen exceptions mentioned above.

In short, China's so-called flexible fair use regulations are actually a combination of the enumeration method and the U.S. four-factor method.⁸⁷ But instead of the expected result of one plus one equals two, judgments may become more inconsistent than expected. The superposition of the two different systems makes it possible for Chinese judges to apply different legal systems and reach different results when facing similar cases. Especially in high-tech, more complex cases like artificial intelligence, the judge may be worried about the difficulty of using the four factors to make an accurate decision. Judges may flee to enumeration, citing the fact that the case type is not one of the 13 exceptions. Since AI use of copyrighted materials is unlikely to be considered fair use, this directly blocks the development of AI in China.

To summarize, the law on the legality of AI's unauthorized use of copyrighted materials is unclear in both China and the United States. If the courts in the United States rely solely on the four factors of the fair use rule to review this issue, the final decision may be uncertain. In particular, the third and fourth factors are unfavorable for AI machine learning. The Opinions of the Supreme People's Court of China are broadly consistent with the four factors of the U.S. fair use statute, and the 13 additional exceptions do not cover AI issues. Therefore,

⁸⁷ Xu & Yang, *supra* note 5, at 40; *see also Opinions, supra* note 84.

the laws of both countries need to be further clarified and amended on this issue.

C. *The Need to Clarify the Status of Artificial Intelligence's Unauthorized Use*

It is necessary to include AI's unauthorized use of copyrighted materials in the scope of fair use as much as possible. First, AI technology is one of the most important directions in the future development of science and technology.⁸⁸ As mentioned above, copying and inputting existing data is an essential first step for AI to create new products.⁸⁹ Specifically, if the fair use system does not include the use of unauthorized materials by AI, the research and development of artificial intelligence will be directly stalled. The use of copyrighted materials by AI will be considered infringement, thus greatly reducing the efficiency of the development of AI technology. Recent studies have shown that “much of the most important current research in AI and machine learning is subject to litigation once it becomes commercially prominent” and that much of the raw data sets currently have licensing restrictions.⁹⁰ Any future breakthrough in AI machine learning would require either an exemption from local law for AI use of unlicensed material or permission from the copyright owner of the dataset. This is a blow to the prospects of AI machine learning.

Second, the number of enterprises that develop and apply AI technology has reached a high level, so direct

⁸⁸ Mauritz Kop, *AI & Intellectual Property: Towards an Articulated Public Domain*, 28 TEX. INTELL. PROP. L.J. 297, 300 (2020).

⁸⁹ Sobel, *supra* note 2, at 62.

⁹⁰ Martin Anderson, *Current AI Practices Could Be Enabling a New Generation of Copyright Trolls*, UNITE.AI (Nov. 5, 2021), <https://www.unite.ai/current-ai-practices-could-be-enabling-a-new-generation-of-patent-trolls/> [https://perma.cc/KY2R-CFHN].

copying of many unauthorized materials is a common phenomenon.⁹¹ If the fair use system cannot effectively cover unauthorized data input, it can cause these enterprises to go bankrupt, resulting in a large number of technical personnel becoming unemployed, and many people will be left without science and technology resources.⁹² Thus, social unrest and social problems may be inevitable.

Finally, the unlimited input of massive information by artificial intelligence can allow scientific and technological personnel to develop more AI technology, and the generation of massive amounts of works also promotes cultural prosperity.⁹³

However, it would be unrealistic to directly and indiscriminately exempt all artificial intelligence from liability for using unauthorized materials at the legal level. Therefore, we should create a new type of suitable legal framework for such a fast-developing novel technology based on the fair use norms with different characteristics of China and U.S. law. This framework should be implemented by WIPO to resolve similar disputes in the international arena.

II. A SYNTHETIC “FAIR USE” MODEL IN WIPO FOR ARTIFICIAL INTELLIGENCE’S UNAUTHORIZED USE OF COPYRIGHTED MATERIALS

In order to clear the obstacles to the development of artificial intelligence in the legal field, most scholars have also tried to show that the use of unauthorized materials by

⁹¹ *See id.*

⁹² *See First AI Art Generator Lawsuits Threaten Future of Emerging Tech*, BLOOMBERG LAW (Jan. 20, 2023, 5:05 AM), <https://news.bloomberglaw.com/ip-law/first-ai-art-generator-lawsuits-threaten-future-of-emerging-tech> [<https://perma.cc/M79C-T3AR>].

⁹³ *See* West & Allen, *supra* note 37.

artificial intelligence is fair use.⁹⁴ However, this is only a theoretical or conceptual waste of effort. In response to such issues, the copyright laws of China and the United States are very vague, but each has its own merits. I propose to combine the different provisions and characteristic understandings of fair use in both countries to form a synthetic “fair use” model. I propose that WIPO adopt a synthetic fair use model for AI, meaning that if the work involves the use of unauthorized materials by artificial intelligence, the synthetic “fair use” model can be considered.

A. *Proposal of the Synthetic “Fair Use” Model: Four-Factor Standard + AI Exceptions*

To ensure that the laws of both China and the United States can more effectively clear the obstacles for the use of unauthorized materials by AI, the legal provisions of China and the United States on fair use can be effectively combined. The four-factor standard of fair use in the United States is obviously more flexible than China’s 13 exceptions legal system, but the latter has more legal certainty than the former. Therefore, the synthetic “fair use” model may be able to integrate the advantages of both and achieve better results.

1. Draft of the Synthetic “Fair Use” Model

a. Enumerated Exception for Artificial Intelligence’s Use of Copyrighted Works

⁹⁴ See Lemley & Casey, *supra* note 9, at 764.

The uses of copyrighted works by artificial intelligence do not constitute infringement if the following three conditions are met:

- (1) Artificial intelligence used copyrighted works presented on public platforms that fall within the scope that the general public can query, observe or obtain;
- (2) Any commercial use of works created by artificial intelligence does not seize the core market of the copyright owner of the raw material; and
- (3) The means by which artificial intelligence obtains unauthorized material and the method of its use must not violate other laws.

b. Fair Use Exception

In addition, the use of copyrighted works by artificial intelligence may constitute fair use. In determining whether the use made of a work, in any case, is fair use, the factors to be considered shall include—

- (1) the purpose and character of the use;
- (2) the nature of the copyrighted work;
- (3) the amount and substantiality of the portion used;
- (4) the effect of the use upon the potential market for or value of the copyrighted work.

This regulation only pertains to AI use of unauthorized materials and has no significance for other issues involving copyright.

2. A Combination of Enumeration and Four-Factor Standard

As stated above, the four-factor standard of fair use is original in the United States,⁹⁵ and it is highly reasonable

⁹⁵ 17 U.S.C. § 107.

and practical for examining whether the use of an unauthorized work in a new work can be judged as fair use. The enumeration method, to stipulate all the situations that can be judged as fair use, was created by China.⁹⁶ Although it seems to be relatively rigid, it still plays a great role in China, a country with a civil law system.⁹⁷ Moreover, in the current judicial practice, China also cites the four-factor standard of the United States to a certain extent to determine fair use cases.⁹⁸ In general, however, the enumeration method takes precedence over the four-factor standard since the four-factor standard is not written in China's fair use law, and only the written law is binding under China's civil law system.⁹⁹ To make the fair use regulations for the use of unauthorized materials by AI, there needs to be a certain international uniformity in WIPO. The combination of fair use laws from two of the biggest countries in the AI industry makes the synthetic "fair use" model more practical and can more effectively remove obstacles for the development of artificial intelligence.

a. AI Exceptions Are the Core

The first part of the proposal is that the enumerated exception should be prioritized in cases where AI uses unauthorized material. In the synthetic "fair use" model, the three necessary conditions are the core for AI use of unauthorized materials to be directly identified as fair use. Expressed in enumerated form, the three conditions are clearer and more convenient for judicial application. Unlike the enumeration of fair use provisions in China's

⁹⁶ Article 24, *supra* note 3.

⁹⁷ *Does China Have Common Law?—China Law in One Minute*, *supra* note 79.

⁹⁸ *See Opinions*, *supra* note 84.

⁹⁹ *See* Article 24, *supra* note 3; *Does China Have Common Law?—China Law in One Minute*, *supra* note 79.

copyright law, only some of which may need to be met for a finding of fair use, the proposed enumerated exceptions require AI to satisfy all three conditions simultaneously in the process of using unauthorized material. Violations of one or more of these enumerated factors should not be considered fair use and should be further reviewed under the four-factor standard.

In the first proposed exception, the main emphasis is that the unauthorized materials used by artificial intelligence should have been presented on public platforms, which may include search websites, social platforms, etc. It is worth mentioning that the concept of a public platform should be strictly distinguished from the public domain. As mentioned in the second half of the first proposed exception, the scope of the public platform mentioned here is anywhere the general public can inquire, search or obtain. However, these materials or data presented on public platforms are not fully in the public domain and are still protected by copyright. After all, research on a subject would be meaningless if it was material that had already entered the public domain.

More specifically, why does this article stipulate that unauthorized material used by AI should have been accessible on a public platform? On the one hand, this is to prevent the creator or operator of the AI from installing a hacker-like system in the AI's system. If the AI's hacking system enters the copyright owner's private platform and steals the material before the copyright owner has disclosed the material on the public platform, this will not fall under this exception. The real mission of AI should be to save human labor, benefit humans, and facilitate the work of human beings instead of stealing human wisdom and private intellectual property through advanced technology.

For example, Singapore Prime Minister Lee Hsien Loong once disclosed on Facebook that case information on himself and several Singapore government officials had

been leaked by hackers by breaking into the hospital system.¹⁰⁰ The Washington Post analyzed data from the Department of Health and Human Services and found that in March 2015 alone, the U.S. medical system suffered more than 1,100 hacking attacks, damaging the interests of more than 120 million people.¹⁰¹ From these cases, we can see that obtaining so-called unauthorized materials on public platforms through the use of technology can be harmful to others. Thus, the first proposed exemption cannot include platforms such as hospitals and other special institutions that have the right to hold personal information.

For the second proposed exception, to a certain extent, reference is made to the fourth factor of the U.S. fair use standard, namely “the effect of the use upon the potential market for or value of the copyrighted work.”¹⁰² However, the biggest difference between the two is the identification of the market scope faced by the original copyright material. Although some scholars once argued that AI’s use of the original work does not interfere with the copyright owner’s core market, and copyright owners do not create works for the purpose of selling them to AI,¹⁰³ it still cannot be ruled out that the works produced by AI may further damage the main commercial value of the original copyright material. After all, the output of AI is indeed more efficient than that of humans. That is to say, AI works created by using unauthorized materials may not

¹⁰⁰ Wu Yingqiu (吴盈秋), *Xīnjiāpō 150 Wàn Bìng Huàn Shùjù Bèi Dào Zǒnglǐ Lǐxiǎnlóng Kāi Yào Jìlù Zāo Xièlòu* (新加坡150万病患数据被盗 总理李显龙开药记录遭泄露) [*Singapore’s 1.5 Million Patient Data Stolen, Prime Minister Lee Hsien Loong’s Prescription Records Leaked*], ZHÈJIĀNG ZÀIXIÀN (浙江在线) [ZHEJIANG ONLINE] (July 21, 2018), http://china.zjol.com.cn/ktx/201807/t20180721_7832180.shtml [<https://perma.cc/NVU8-C3GQ>].

¹⁰¹ *Id.*

¹⁰² 17 U.S.C. § 107(4).

¹⁰³ *See, e.g., Lemley & Casey, supra note 9, at 777.*

encroach on the core market of the original copyrighted materials if the core market is defined as the audience group that the original copyright material mainly faces.¹⁰⁴ The core market can also be the main source of commercial value. However, this definition of core market is too harsh for the characteristics of AI itself.¹⁰⁵ We cannot rule out that AI-created works may have overlapping or similar uses with works derived from the raw data as well as the data itself.

In any case, if the use of unauthorized material by AI is to be conditionally legalized, more consideration needs to be given to how to minimize harm to the copyright owner of the raw data. The focus of the second proposed exception is to better balance the development of AI and the protection of personal intellectual property rights. Simply put, expanding the consideration of the “potential market” in the “core market” may solve this problem. After all, if the market preemption problem faced by works created by AI is not considered at all, then the use of unauthorized materials by AI should also be resolved through the consent of the copyright owner of the raw material or by purchase as consideration for the transaction.

The third proposed exception serves to control how the AI obtains and uses unauthorized material. In the current data age, almost everyone’s information is recorded on the internet to a greater or lesser extent. Thus, it is indispensable to prevent AI from using illegal means to

¹⁰⁴ See HARV. UNIV. OFF. OF THE GEN. COUNS., COPYRIGHT AND FAIR USE: A GUIDE FOR THE HARVARD COMMUNITY 11 (2016), https://ogc.harvard.edu/sites/hwpi.harvard.edu/files/ogc/files/ogc_copyright_and_fair_use_guide_5-31-16.pdf?m=1464875856 [<https://perma.cc/599T-EH3D>].

¹⁰⁵ See *id.* (“This inquiry is not confined to the market for the original, but also takes into account derivative markets.”). Because the basis of AI created works is the use of copied data as raw materials, the possibility of touching the market of works derived from the data is very high.

obtain materials, the same illegal situation that the first proposed exception is trying to prevent.

There are also hidden dangers in how AI can illegally use data. This presents a new question: what is the specific use of AI here? Maybe it is machine learning. More generally speaking, it refers to the process of AI recording and learning. But if AI is used to copy this data and resell it illegally, such use will not be allowed. In other words, the third proposed exemption serves to prevent the creators and operators of AI from stealing private information in the name of fair use of unauthorized materials, which would pose a threat to the property safety of others.

These three exceptions are complementary and indispensable to each other, and AI's use of unauthorized material that violates any of the above-listed points will not be directly eligible for fair use. Of course, as mentioned above, the three proposed exceptions are only pre-procedures created in the case of AI's use of unauthorized materials. If a case fails to pass the review of these three exceptions, it should not directly be considered illegal and should then be transferred to the normal four-factor standard review process. It is worth mentioning that such a synthetic "fair use" model is proprietary and has no reference to cases not involving the use of unauthorized materials in AI machine learning.

b. The Four-Factor Standard is a Secondary Choice

The next part of the proposal is that the four fair use factors can be used as a secondary means to ensure that a case can be heard normally when the AI exceptions are not met. In past practice, the four-factor standard has been used in copyright cases in US courts.¹⁰⁶ For China,

¹⁰⁶ See, e.g., *Harper & Row Publishers, Inc. v. Nation Enter.*, 471 U.S. 539, 560–61 (1985).

although the four-factor standard has been stipulated in opinions of the Supreme People's Court,¹⁰⁷ the relevant judicial practice is not mature. It is possible that the existing enumerated exceptions can better meet the needs of the Chinese judiciary when it comes to fair use cases. However, for emerging technology industries and high-tech fields such as AI, relying on the enumeration method cannot cover all the legal issues that may arise in the future. In contrast, the more flexible four-factor standard is a long-term effective and deterministic regulation. Therefore, the four-factor standard should still be used as the foundation for solving copyright issues related to artificial intelligence in China.

It is worth mentioning that at this stage, the learning process of artificial intelligence still relies on the copying of a large amount of data or materials.¹⁰⁸ Therefore, if the use of data by artificial intelligence is strictly required according to the four-factor standard, the development of artificial intelligence will have a huge cost that may be difficult to achieve. For example, the third factor, "the amount and substantiality of the portion used,"¹⁰⁹ is most unfavorable to the development of artificial intelligence because the proportion of materials used by AI is close to 100%. However, it's almost unfair to completely block the development of artificial intelligence. In a way, we should think of AI as having the same right to learn as humans.¹¹⁰ Perhaps there should even be more tolerance for the nascent industry of artificial intelligence.

Accordingly, the real significance of the four-factor standard as the secondary choice is not the rationality of its review of AI's use of unauthorized materials, but the ability to ensure that the normal progress of the case can be

¹⁰⁷ See *Opinions*, *supra* note 84.

¹⁰⁸ Sobel, *supra* note 2, at 48.

¹⁰⁹ 17 U.S.C. § 107(3).

¹¹⁰ Lemley & Casey, *supra* note 9, at 782.

guaranteed under the circumstances that the three enumerated exceptions are not met. It is not so much that the synthetic “fair use” model has made a big change to the regulations of fair use, but this model optimizes these regulations from the perspective of AI development.

B. Similarities and Differences between Synthetic “Fair Use” Model and Fair Learning

As mentioned previously, the fair learning theory contends that AI should have the same right to learn as humans.¹¹¹ The introduction of fair learning theory and standards stems from the analysis of the obstacles generated in the development of machine learning and the weakness of the four factors of fair use in dealing with related AI problems.¹¹² From this perspective, it may be seen that the synthetic “fair use” model and the fair learning standard share certain similarities in theoretical bases and origins.

Furthermore, there is still a certain degree of similarity in the structure and content of the two standards. For example, the core market mentioned in the second proposed exception in the synthetic “fair use” model comes from the further optimization of the fourth factor of the four factors of fair use.¹¹³ Similarly, the fair learning standard is also a refinement of the four factors of fair use after distinguishing different situations of machine learning.¹¹⁴

However, there is also a major difference between the two. Fair learning “propose[s] a standard that courts and technologies can use to get out of the hole that recent copyright precedent and increasingly strident public

¹¹¹ *Id.*

¹¹² *Id.* at 760–61.

¹¹³ *See supra* notes 102–105 and accompanying text.

¹¹⁴ Lemley & Casey, *supra* note 9, at 777.

opinion have dug for them.”¹¹⁵ To some extent, fair learning is a new standard based on the four factors of fair use. Fair learning has not completely removed the “coat” of the four factors of fair use, but it has adjusted, refined, and clarified the specific concept underlying the four fair use factors. For example, the fair learning theory provides that “[o]nly if the use directly interferes with the plaintiffs core market should the fourth factor outweigh a finding of fair learning under the first factor.”¹¹⁶ Thus, we can think of the fair learning standard as a standard that makes it easier for courts to make judgments when hearing the machine learning cases.

In contrast, the synthetic “fair use” model is an innovation that is completely independent of the four factors of fair use. Although the four factors of fair use are included in the entire model, the enumerated exceptions contained in the pattern are completely separate from the four factors of fair use and are prioritized as the core of the synthetic “fair use” model. For example, a judging criterion such as the core market also appears in the criterion of fair learning, but there it is only used as a factor.¹¹⁷ In the synthetic “fair use” model, the core market can become a determinative factor in whether AI use of unauthorized materials is fair use.

C. *Reasons for Adopting Proposal*

The proposal combines the different legal provisions on fair use in copyright law in China and the United States. In the face of AI, a new era of high-tech industry development, the synthetic “fair use” model helps to clear obstacles for AI development. At the same time, the adoption of a synthetic “fair use” model by WIPO could

¹¹⁵ *Id.* at 760.

¹¹⁶ *Id.* at 777.

¹¹⁷ *Id.*

help reduce international disputes related to AI's unauthorized use of copyrighted materials.

1. Promote the Development of AI

The first reason to adopt the proposal is that the synthetic “fair use” model is more flexible than the previous four factors of fair use, and it also removes a lot of obstacles to the development of AI. One of the biggest bottlenecks in the innovation of AI is the input of data. AI can create new works because it can absorb and learn large amounts of material and data in a short period of time, which may need to be input in thousands at a time.¹¹⁸ If each input requires the creator or operator of the AI to obtain the authorization and consent of the copyright owner of the raw material one by one, then the innovation cost of AI will be huge. This obstacle to the development of AI is unimaginable.

Therefore, proprietary legislation on the use of unauthorized materials by AI is necessary. Compared to the original legal systems, the cases of AI using unauthorized materials will be treated leniently under the synthetic “fair use” model so that AI creators and operators can obtain a more just legal basis, stimulating their enthusiasm for the AI business and helping to promote healthy development of the AI industry.

2. Avoid Unfair Monopoly and Competition of Data Materials

The next reason to adopt the proposal is that the synthetic “fair use” model is fairer, and the rigorous combination of enumeration and four factors can more effectively prevent the occurrence of monopoly. As we all know, AI creation is highly dependent on the input of raw materials or data.¹¹⁹ When AI creates works of the same

¹¹⁸ Sobel, *supra* note 2, at 48.

¹¹⁹ *Id.*

type or works with a strong correlation, there is more than one source of raw materials. For example, in *Authors Guild*, Google was able to digitize a large amount of raw materials and data, create a search function, and display these works.¹²⁰ This is because Google, as an early stage of AI, can absorb a large amount of data or material resources from different sources rather than inputting the resources of a person or a company.¹²¹ This is the nature of AI itself.

Therefore, once someone attempts to integrate, monopolize, and sell the materials or data resources that a certain type of AI must input at high prices, it will be a devastating disaster for the AI industry. To prevent the occurrence of this problem, or to completely solve this problem at the source, from a legal point of view, conditional open AI use of unauthorized materials is a feasible method.

3. Harmonization

The last reason to adopt the proposal is that the synthetic “fair use” model combines two different styles of fair use regulations, effectively promoting international harmonization in cases of unauthorized use of materials by AI. Because the development of AI is not mature enough, there are not many cases of unauthorized use of AI in China and the United States. As far as current laws in China and the United States are concerned, when faced with two similar cases of AI using unauthorized materials, it is very likely that different courts will make different judgments according to the law in the future.

Therefore, it stands to reason that there should be a specific regulation for AI’s use of unauthorized material. When the problem arises internationally, it is not entirely appropriate to use any of the characteristic styles. From the

¹²⁰ *Authors Guild v. Google, Inc.*, 804 F.3d 202, 229 (2d Cir. 2015).

¹²¹ *Id.*

perspective of the synthetic “fair use” model, the three additional exception conditions improve the certainty of the law to a certain extent. Overall, the enactment of this model by WIPO can effectively promote international harmonization.

III. CRITICISMS ON A SYNTHETIC “FAIR USE” MODEL FOR ARTIFICIAL INTELLIGENCE

As mentioned in the proposal, the synthetic “fair use” model breaks through the original four factors of fair use from the U.S. Although the model is only applicable to the field where AI uses unauthorized materials, there are still objections that the application of the model may disrupt the original order of copyright law and be too tolerant of the AI field, which may lead to a decline in the innovation enthusiasm of creators in other fields. Perhaps there will be various questions and objections about the synthetic “fair use” model itself, which will be presented and answered one by one below.

A. Intellectual Property Rights of Creators of Copyrighted Materials

The first objection to consider is that synthetic “fair use” model may, to some extent, serve as an excuse for developers or commercial organizations in the AI field to infringe on the intellectual property rights of creators of other copyrighted materials. Because the synthetic model is designed to promote the development of the AI field, the synthetic “fair use” model is more pertinent to AI than the four fair use factors and thus allows for vitality in the research and development of the AI field. This may be accompanied by a weakening of the protection of copyrighted material in other areas. For example, after a large amount of unauthorized material is input into an AI machine, if the case is reviewed according to the synthetic

“fair use” model, it is more likely to be considered non-infringing than under the four factors of fair use. This is very important for copyright material owners and creators. Copyright owners and creators may consider this to be unfair, which may weaken the creative passion and innovation of creators in other fields.

There may be some truth to this objection, but it is not entirely correct. The purpose of copyright, since the beginning of its birth, has been to protect the private property of creators.¹²² The synthetic “fair use” model does not violate this purpose and concept. The model is based on existing copyright law and is still committed to protecting the copyrights of creators in other fields. Compared with the four factors of fair use stipulated in U.S. copyright law, the innovation of the synthetic “fair use” model is that it only applies fair use flexibly in the context of AI data input and work output. This kind of flexible fair use model does not encroach too much on the rights of the creators of raw materials. For example, in the synthetic “fair use” model, the second necessary condition of the AI exception refers to the concept of the “core market,” which is proposed to ensure that the main economic interests of creators are not overly affected.

After all, “[c]opyright functions as a private property right that enables artists to live and make money as entrepreneurs in a free market.”¹²³ This not only makes sense, but also accounts for a large proportion of creative passion. According to this logic, I am afraid that the provisions on fair use in the current copyright law are insufficient.

¹²² Matthew Barblan, *Copyright as a Platform for Artistic and Creative Freedom*, 23 *GEO. MASON L. REV.* 793, 798 (2016).

¹²³ *Id.* at 797.

B. *What if the AI Use Fails the Enumerated Exceptions but Satisfies Fair Use?*

A second objection that might be raised is that there might be a situation where AI uses copyrighted works that fail the proposed enumerated exceptions but satisfy the four fair use factors. The possibility of this happening is very small but still exists. For this synthetic “fair use” model, by forcibly prioritizing one of the regulatory models over the other, the former should be more rigorous and should satisfy cases in almost all relevant fields.

However, this objection is not entirely accurate, nor is it necessary. As mentioned above, the synthetic “fair use” model is only an optimization and transformation of general fair use in the field of AI, rather than a replacement. As such, the synthetic model can be looser than the four factors of fair use. However, the three proposed exceptions are essentially derived from the four factors of fair use. Thus, there is almost no case that fails the proposed enumerated exceptions but satisfies fair use. Although there may be a case that fails the enumerated exceptions but satisfies fair use, it wouldn’t matter because, in the synthetic system, “if a case fails to pass the review of these three exceptions, it should not directly be considered illegal and should then be transferred to the normal four-factor standard review process.”¹²⁴ Thus, after failing the three exceptions, AI is not directly eliminated but needs to be reviewed by fair use. The result is still passable and fair.

C. *Disturbs the Original Order of Copyright Law*

The third objection might argue that the proposed synthetic “fair use” model disturbs existing copyright law to some extent. Under existing copyright law, the four

¹²⁴ See *supra* Part II.

factors of fair use can address most of the controversial cases. Some may argue that incorporating synthetic “fair use” models into fair use statutes as exceptions for the AI space could undermine the integrity of fair use statutes. In the future, there may be people in other fields who hope that their fields can break away from the four factors of fair use, like the AI field, to form their own synthetic model.

This third objection is unfounded and illogical. The proposed synthetic “fair use” model is ultimately a special model that handles situations in the AI field flexibly, and this flexibility is moderate rather than ambiguous and confusing. In the proposed synthetic “fair use” model, the application procedures of the three exceptions and the four factors of fair use are strictly distinguished. The former needs to be satisfied first, and the latter is a secondary choice when the former cannot be complied with. The general four factors of fair use are sufficient to solve non-AI cases on their own, and the synthetic “fair use” model has no reference to non-AI cases.

CONCLUSION

“AI is going to change the world more than anything in the history of mankind. More than electricity.”¹²⁵ Now is the stage of rapid development of AI, and it is understandable to want to eliminate unreasonable legal hurdles as much as possible. When the general four factors of fair use apply to the unauthorized use of copyright materials by AI, some of the factors are

¹²⁵ Catherine Clifford, *The ‘Oracle of A.I.’: These 4 Kinds of Jobs Won’t Be Replaced by Robots*, CNBC (Jan. 15, 2019, 3:17 AM), <https://www.cnbc.com/2019/01/14/the-oracle-of-ai-these-kinds-of-jobs-will-not-be-replaced-by-robots-.html#:~:text=%E2%80%9CI%20believe%20%5BAI%5D%20is,particular%20concern%20for%20many%20workers> [https://perma.cc/SB64-AQ7Y].

The Legality of Artificial Intelligence's Unauthorized Use of Copyrighted Materials under China and U.S. Law **279**

very unfavorable and unfair to AI. The synthetic “fair use” model is an optimization of fair use to a certain extent that will solve the issue of the legality of AI use of copyrighted materials.