

# UIC REVIEW OF INTELLECTUAL PROPERTY LAW



## RESEARCH ON THE DILEMMA AND IMPROVEMENT OF THE COPYRIGHT FAIR USE DOCTRINE RELATED TO MACHINE LEARNING IN CHINA

YANG GAO, PAUL KOSSOF, & YAN DONG

### ABSTRACT

While artificial intelligence has brought convenience to human life, it has also profoundly affected copyright law. Mechanical learning in the past did not infringe the exclusive rights of the copyright owner, and fair use doctrine could be applied to ensure the development of technology. As machine learning is the core technology of artificial intelligence, if the fair use doctrine is applied then it will infringe the exclusive rights of copyright owners. Based on the working principle of machine learning technology, this paper discusses the copyright infringement risk of machine learning technology. Furthermore, this paper analyzes the challenge of machine learning technology to the basic principle of copyright fair use doctrine and the challenge of machine learning technology to "transformative use" doctrine from the perspective of rule application. To resolve the conflict between the exclusive rights of the copyright owner and the development of artificial intelligence technology, a legal license can be applied to artificial intelligence technology companies as a solution.



*Cite as* Yang Gao, Paul Kossof, & Yan Dong, *Research on the Dilemma and Improvement of the Copyright Fair Use Doctrine Related to Machine Learning in China*, 22 UIC REV. INTELL. PROP. L. 1 (2022).

# RESEARCH ON THE DILEMMA AND IMPROVEMENT OF THE COPYRIGHT FAIR USE DOCTRINE RELATED TO MACHINE LEARNING IN CHINA

YANG GAO, PAUL KOSSOF, YAN DONG

I. INTRODUCTION .....	1
II. POTENTIAL COPYRIGHT INFRINGEMENT RISKS IN THE PROCESS OF “LEARNING” BY MACHINE LEARNING.....	3
A. The Learning Process of Machine Learning .....	3
B. Risk of copyright infringement of machine learning.....	4
1. Whether machine learning infringes on reproduction rights .....	4
2. Whether machine learning violates the right of translation.....	5
3. Whether machine learning infringes on the right of adaptation .....	7
III. MACHINE LEARNING APPLIES FAIR USE DILEMMA ANALYSIS.....	8
A. Challenges of Machine Learning to the Fundamentals of fair use doctrine .....	8
B. Machine Learning Challenges for Transformative Use .....	10
1. History of the Development of Transformative Use .....	10
2. Whether Machine Learning Constitutes Transformative Use.....	12
IV. MACHINE LEARNING IS THE WAY OUT OF THE FAIR USE DILEMMA .....	15
A. Analysis of the Necessity of Setting Statutory Licenses .....	15
B. Institutional advantages of statutory licensing.....	17
C. Realization of statutory licenses .....	19
V. CONCLUSION.....	21

## RESEARCH ON THE DILEMMA AND IMPROVEMENT OF COPYRIGHT FAIR USE DOCTRINE RELATED TO MACHINE LEARNING IN CHINA

YANG GAO, PAUL KOSSOF, YAN DONG\*

### I. INTRODUCTION

Artificial intelligence ("AI") has a profound impact on the social life of human beings. While AI facilitates human life, it also brings great challenges to the existing copyright system.<sup>1</sup> At present, academics mostly focus on the copyrightability and rights attribution of AI products, and there is little research on copyright infringement in the process of AI learning.<sup>2</sup> In general, "learning materials" of AI include two kinds: (1) data not protected by copyright law, and (2) works protected by copyright law.<sup>3</sup>

In the former case, it is controversial whether AI can learn from data for free. If it is argued that a "data property right" should be created over the data, AI should obtain authorization from the data owner before learning from the data.<sup>4</sup> Conversely, if it is argued that no new property rights should be created for the data, then AI can learn from the data for free.<sup>5</sup> Research by the Max Planck Institute in Germany suggests that property rights should not be created for data for three reasons.<sup>6</sup> First, the properties of data make data transactions more complex, and creating property rights for data does not facilitate data transactions; second, creating property rights

---

\* © 2022 Yang Gao, Paul Kossof, Yan Dong.

This article was funded by the National Social Science Foundation of China under the project "Research on the Boundary of Enterprise Data Rational Use and its Institutional Construction," (22CFX028).

Yang Gao, ORCID: 0000-0003-2777-042X, is a postdoctoral fellow at East China University of Political Science and Law lecturer at Shanghai University of International Business and Economics. Primary research interests: intellectual property law; data law.

Paul Kossof (JD, LL.M), ORCID: 0000-0002-8462-6179, is Senior Vice President (Legal) at DAC Management LLC. He advises on a wide range of international investment/legal matters such as entity formation & administration, labor, tax, compliance, IP protection, distressed debt & collections, and infrastructure & construction.

Yan Dong, ORCID: 0000-0002-5098-0110, is a postgraduate law student at Shanghai University of International Business and Economics. Primary research interest: intellectual property.

<sup>1</sup> See generally Guan Yuying, *A Study on New Issues of Intellectual Property Arising by Artificial Intelligence*, 5 J. OF GUIZHOU PROVINCIAL PARTY SCH. 5 (2018).

<sup>2</sup> *Id.*

<sup>3</sup> *Id.*

<sup>4</sup> Ji Leilei, *Progress and Analysis of Data Ownership Research in Big Data Era*, 2 LIBR. 27-32 (2019).

<sup>5</sup> *Id.*

<sup>6</sup> See Reto M. Hilty & Heiko Richter, *On the Urge to Regulate Freedom*, MAXPLANCKRESEARCH (2019) 10, [https://www.mpg.de/14148813/W001\\_Viewpoint\\_010-015.pdf](https://www.mpg.de/14148813/W001_Viewpoint_010-015.pdf).

for data may strengthen the monopoly of data holders over data and exclude third parties; and third, creating property rights for data may create legal uncertainty.<sup>7</sup>

This paper agrees with the view that AI can copy and learn from the data without obtaining the data holder's permission. In the case of copyrighted works, works may originate from two sources.<sup>8</sup> First, AI technology companies use works after obtaining authorization from the copyright holders, and second, AI technology companies use works without authorization from the copyright holders.<sup>9</sup>

In the first source, the use of a work is not illegal.<sup>10</sup> However, in the second source, the use of copyrighted works by AI technology companies may fall within the control of the content of copyright rights and constitute infringement.<sup>11</sup> This article namely explores whether the copyright fair use system is applicable to machine learning under the second case.

The development of artificial intelligence is characterized by a progression from mechanical learning to machine learning. Mechanical learning is the process of acquiring works and further using them by relying on technical equipment.<sup>12</sup> It in itself exhibits a strong tool attribute, requiring the user to plan the operation process in advance and determine the learning object and characteristics.<sup>13</sup> Machine learning focuses on how computers can simulate human learning behaviors as the main research content.<sup>14</sup> Machine learning can autonomously identify learning objects, construct their features, and step outside the limitations of preset commands for additional operations.<sup>15</sup>

From the perspective of the "idea/expression dichotomy," mechanical learning can only obtain experience from the thought of the work, while machine learning can discover value from the expression of the work.<sup>16</sup> For example, Arriba's thumbnail search engine, Google Image Search, and Google Books all belong to the category of mechanical learning, which cannot "learn" from the original expression of works.<sup>17</sup> However, Google's application, "Smart Reply," belongs to the category of machine learning as it can "learn" the expressions of a large number of works in the network, making smart response sentences no different from those of people.<sup>18</sup> Whether it is machine learning, mechanical learning, for the purpose of scientific research, or for the

---

<sup>7</sup> *Id.*

<sup>8</sup> Jiao Heping, *Copyright Risks of Data Acquisition and Utilization in Artificial Intelligence Creation and the Path to Resolution*, 4 CONTEMP. L. REV. 128, 137 (2022).

<sup>9</sup> *Id.*

<sup>10</sup> Ma Zhongfa, & Xiao Yulu, *On Fair Use of Artificial Intelligence's Learning to Create*, 5 J. OF SHANDONG U. OF SCI. AND TECH. (SOC. SCI.) 33, 37 (2020).

<sup>11</sup> *Id.*

<sup>12</sup> Sun Yang, *Argumentation on Rational Usage of Artificial Intelligence*, 3 CROSS-STRAIT LEGAL SCI. 46, 48 (2018).

<sup>13</sup> *Id.*

<sup>14</sup> Ou Huajie, *An Overview of Machine Learning Algorithms in the Context of Big Data*, 4 INFORMATIZATION-RES. 50, 50 (2019).

<sup>15</sup> *Id.*

<sup>16</sup> Zhang Huibing, *Copyright Dilemma and Solution of Sports News in Artificial Intelligence Era—Taking Tencent's Robot Reporter as an Object*, 2 J. OF SHANGHAI U. OF SPORT 33, 38 (2019).

<sup>17</sup> Yao Hehui, *Research on the Fair Use of Copyright in Google Digital Library*, 1 LIBR. WORK AND STUDY 44, 44-48 (2015).

<sup>18</sup> Li An, *Copyright Law Analysis of Machine Learning Works—Nonuse, Fair Use and Infringing Use*, 6 ELEC. INTELL. PROP. 60, 63 (2020).

collection of cultural heritage institutions, the provisions of Article 22 (6) and (8) of the Copyright Act can be applied and fall under the scope of fair use.<sup>19</sup>

However, courts may disagree on whether fair use can apply to mechanical learning and machine learning when used for commercial purposes.<sup>20</sup> The reason why machine learning can still apply fair use and use works without permission is that fair use can reconcile the conflicting interests between technological development and exclusive right.<sup>21</sup> As for machine learning, it is worth studying whether the fair use system can still be used as a tool for balancing interests. Taking the working principle of machine learning as a starting point, this paper analyzes the basic principles and criteria of fair use; argues that the application of fair use to machine learning for commercial purposes is caught in a dilemma; and proposes a solution to how to get out of the dilemma and balance the conflict between the author's exclusive rights and technological development.

## II. POTENTIAL COPYRIGHT INFRINGEMENT RISKS IN THE PROCESS OF “LEARNING” BY MACHINE LEARNING

### A. *The Learning Process of Machine Learning*

To analyze the potential copyright infringement risk of machine learning, it is necessary to understand the learning process of machine learning. Taking Natural Language Processing (NLP) as an example, the workflow of machine learning consists of the following five steps.

Step 1: Data is collected and databased from works, dictionaries, and web texts. Step 2: Pre-processing of the database converts the human language in the database into computer language. Step 3: Database annotation, according to different classification criteria such as grammar, part of speech, morphology, etc. The data is labeled, and the initial model is established. Step 4: Model training. The initial model extracts a set of labeled data and learns the correlation between the data and the classification criteria and derives the rule between things. Step 5: Model fixation. The model after step 4 is created as a permanent file to get the final model of machine learning. Based on the data automatically collected by machine learning in the network or input by the user, the final model produces the output, the AI generator.

---

<sup>19</sup> (中华人民共和国著作权法) [Copyright Law of the PRC] (promulgated by P.R.C. LAWS, effective Oct. 27, 2001) art. 22:

(6) translation, or reproduction in a small quantity of copies, of a published work for use by teachers or scientific researchers, in classroom teaching or scientific research, provided that the translation or reproduction shall not be published or distributed; (8) reproduction of a work in its collections by a library, archive, memorial hall, museum, art gallery or any similar institution, for the purposes of the display, or preservation of a copy, of the work.

<sup>20</sup> Yang, *supra* note 12, at 48.

<sup>21</sup> *Id.*

Machine learning can be divided into three categories according to the aforementioned step three: supervised learning, unsupervised learning, and reinforcement learning.<sup>22</sup> In supervised learning, step 3 still classifies and labels the data, and then trains the model with the labeled data.<sup>23</sup> Unsupervised learning does not include the process of classification and labeling.<sup>24</sup> Because of this, machine learning is directly based on unlabeled data for training, and the laws between things are derived autonomously.<sup>25</sup> Reinforcement learning is in between supervised and unsupervised learning, using both labeled and unlabeled data for model training.<sup>26</sup> In either type of machine learning, a database of learning data needs to be created.

### *B. Risk of copyright infringement of machine learning*

Artificial intelligence is based on the digital abstraction of the human brain by a computer, and machine learning is even more similar to the human learning process, which can be said to be the reproduction of the human brain in a computer. When humans learn a new language, they usually transform the training information, such as text, into electrochemical traces stored in the area of the brain dedicated to language, and this is the process of human learning to memorize. The reason why human memorization of other people's works does not infringe on other people's copyrights is that, on the one hand, the economic interests of copyright holders are not affected, and, on the other hand, the regulation of memorization is not realistically operative, so traditional copyright law theory considers such activities to be beyond the scope of copyright law.<sup>27</sup>

However, machine learning faces the risk of copyright infringement from the collection of data to the generation of products. As to which content of copyright rights may be infringed by machine learning, academic discussions mainly focus on reproduction rights, translation rights, adaptation rights, and compilation rights, which are divided into the following paragraphs.

#### *1. Whether machine learning infringes on reproduction rights*

There is no disagreement in the academic community on whether the reproduction right is infringed.<sup>28</sup> Most scholars argue that the "learning" process of machine

---

<sup>22</sup> Xu Hong-xue, et al., *Survey on the Classic Machine Learning Algorithms and Their Applications*, 33 *COMPUTER KNOWLEDGE & TECH.* 17, 18 (2020).

<sup>23</sup> *Id.*

<sup>24</sup> *Id.*

<sup>25</sup> *Id.*

<sup>26</sup> *Id.*

<sup>27</sup> Thomas Margoni, *Artificial Intelligence, Machine learning and EU copyright law: Who owns AI?*, 12 *CREATE WORKING PAPER* 1, 2 (2018).

<sup>28</sup> Liu Youhua & Wei Yuansha, *The Copyright Infringement Issue of Machine Learning and its Solution*, 2 *ECUPL J.*, 22, 73-75 (2019); Hua Jie, *The Dilemma and Solution Concerning Application of Copyright Exceptions to Artificial Intelligence's Creation*, 4 *ELEC. INTELL. PROP.* 26, 30-32 (2019); Ma Zhongfa & Xiao Yulu, *Infringement Dilemma and Outlet of Artificial Intelligence Learning Creation*, 5 *J. OF WULING*, 66, 67 (2019).

learning violates the right to copy.<sup>29</sup> To ensure the accuracy of the data obtained by machine learning and to eliminate errors in the output, machine learning requires the inclusion of as much known data as possible in the database.<sup>30</sup> Therefore, Step 1 involves a large number of unauthorized copying, which is precisely the kind of behavior that the copying right in the copyright property right is intended to regulate.<sup>31</sup> Step 4 is the initial learning process of the model, which is often accompanied by the generation of temporary copies. Since China does not include "temporary copies" in the scope of "reproduction", the issue of temporary copies of machine learning does not need to be adjusted in China.<sup>32</sup>

## 2. Whether machine learning violates the right of translation

In Step 2, machine learning converts human language in the database into computer language.<sup>33</sup> There are some scholars that believe such behavior violates the translation right.<sup>34</sup> The author holds a different view on this, and such behavior is still copying, not translation.

First, the act of translation mainly refers to the translation of this language into another language.<sup>35</sup> If two works meet the infringement criteria of "access and substantial similarity," the latter infringes the translation right of the former.<sup>36</sup> The "work" with computer language as the element does not belong to the category of work. Throughout the types of works stipulated in Article 3 of the Copyright Act, either they can be appreciated by human beings, such as written works and musical works, or they can be used by human beings, such as engineering design drawings and product design drawings.<sup>37</sup>

---

<sup>29</sup> Joshua P. Meltzer, *The Impact of Artificial Intelligence on International Trade*, BROOKINGS (Dec. 13, 2018), <https://www.brookings.edu/research/the-impact-of-artificial-intelligence-on-international-trade/> at 2.

<sup>30</sup> *Id.*

<sup>31</sup> Jiao Heping, *Data Acquisition in Artificial Intelligence Creation Copyright Risks and Mitigation Paths for Data Acquisition and Utilization*, 4 CONTEMP. L. REV. 128, 136 (2022).

<sup>32</sup> Liu Haoyang, *On the Legal Nature of Temporary Reproduction*, Z1 ELEC. INTELL. PROP. 106, 107 (2013).

<sup>33</sup> Youhua & Yuansha, *supra* note 28, at 73-74.

<sup>34</sup> *Id.*

<sup>35</sup> Yin Fenglin & Zhao Yixin, *Challenges and Solutions: The Development of Machine Translation and the Future of Translation Right*, 3 SCI. TECH. & L. (Chinese-English Version) 45, 49 (2021).

<sup>36</sup> *Id.*

<sup>37</sup> (中华人民共和国著作权法) [Copyright Law of the PRC] (promulgated by P.R.C. LAWS, effective Oct. 27, 2001) art. 3:

For the purposes of this Law, the term "works" includes works of literature, art, natural science, social science, engineering technology and the like which are expressed in the following forms: (1) written works; (2) oral works; (3) musical, dramatic, quyi', choreographic and acrobatic works; (4) works of fine art and architecture; (5) photographic works; (6) cinematographic works and works created by virtue of an analogous method of film production; (7) drawings of engineering designs, and product designs; maps, sketches and other graphic works and model works; (8) computer software; (9) other works as provided for in laws and administrative regulations.

However, computer language cannot be understood or used by humans, and can only be used indirectly through computers. The conversion of human language into computer language does not fall into the category of "computer software."<sup>38</sup> Article 2 of the Regulations on the Protection of Computer Software stipulates that computer software includes computer programs and related documents, and Article 3 further explains the meaning of "documents," while the "works" composed of computer language do not belong to the category of "documents."<sup>39</sup> Since "works" composed of computer language do not belong to the category of works in the sense of copyright law, it is impossible to infringe the translation right of the original copyright owner.

Second, even if the "work" constituted by computer language is a work within the meaning of copyright law, the conversion of human language into computer language is not an act controlled by the right of translation.<sup>40</sup> Rather, it is controlled by the right of reproduction.<sup>41</sup> According to Article 10 of the Copyright Act, the right of translation refers to the right to transform the original work into another language.<sup>42</sup> However, whatever the language, it should be a language that human beings can understand, and computer languages, such as C and Java, are computer languages, not "human languages."<sup>43</sup> Moreover, the translated work should have originality. If there is a one-to-one correspondence between two languages, the conversion from one language to another is not original and is not an act controlled by the right to translate.<sup>44</sup> For example, if a work is converted from traditional characters to simplified characters, it is an act of copying rather than an act of translation, because simplified characters

---

<sup>38</sup> (计算机软件保护条例) [Computer Software Protection Regulations of the PRC] (promulgated by P.R.C. LAWS, Dec. 20, 2001, effective Jan. 1, 2002) art. 2. "Computer software referred to in these Regulations (hereinafter referred to as software), refers to computer programs and their related documentation."

<sup>39</sup> *Id.* "For the purposes of these regulations computer software (hereinafter referred to as software) refers to computer programs and related documentation." *Id.* at art. 3:

Meanings of the following words used in these regulations are: (1) Computer programs refers to coded instructional sequences-or those symbol is instructional sequences or numeric language sequences which can be automatically converted into coded instructional sequences-which are for the purpose of obtaining a certain result and which are operated on information processing equipment such as computers. Computer programs include source code programs and object code programs. The source code text of a piece of software and its object code text should be seen as one work. (2) Documentation: refers to written materials and diagrams, using natural language or formal language, which are used to describe the contents organization, design functions and specifications development circumstances, testing results and method of use of the program, for example: program design explanations flow charts, user manuals, etc.

<sup>40</sup> Fenglin & Yixin, *supra* note 35, at 3.

<sup>41</sup> *Id.*

<sup>42</sup> (中华人民共和国著作权法) [Copyright Law of the PRC] (promulgated by P.R.C. LAWS, effective Oct. 27, 2001) art. 10. "(15) the right of translation, that is, the right to translate a work in one language into one in another language."

<sup>43</sup> Pan Jiaqi, *Development and Characteristics of Computer Languages*, in CHINESE & FOREIGN ENTREPRENEURS 29 (2019).

<sup>44</sup> Fenglin & Yixin, *supra* note 35, at 3.

and traditional characters are just different forms of writing Chinese characters.<sup>45</sup> Similarly, the conversion from human language to computer language is done according to the rules set by the programmer, and there is a one-to-one correspondence between the two, so this conversion process is not a translation, but only an act of copyright control.<sup>46</sup>

### *3. Whether machine learning infringes on the right of adaptation*

Regarding the right of adaptation, judicial practice typifies the act of adaptation into two categories, both premised on not changing the basic content of the original work.<sup>47</sup> The first is the transformation of a work of one genre into a work of another genre, such as the adaptation of a novel into a screenplay.<sup>48</sup> The second is the creation of a new work without changing the genre of the original work, such as adapting a famous book into a children's book.<sup>49</sup> Both behavioral approaches may appear in the working process of mechanical learning, but not in the working process of machine learning.<sup>50</sup>

Machine learning is autonomous and does not require human intervention; it "learns" the input data and creates the work according to its own algorithm, without the intention of adapting it.<sup>51</sup> It is hard to imagine that an artificial intelligence would have a sudden idea to adapt a work one day. But mechanical learning may adapt a work under the user's instruction because it is the user's will that dominates mechanical learning.

As for the subject of tort liability, there are differences in academic circles, such as "AI developer," "AI owner," "AI operator," and "AI investor."<sup>52</sup> In the author's view, the "artificial intelligence investor" should be advocated as it aligns with the legislative purpose of the current law and has the "incentive theory" as the support of legal philosophy. Therefore, the subject of machine learning tort liability should be identified as the artificial intelligence technology company.

Whether it is supervised learning, unsupervised learning, or reinforcement learning, in the process of collecting works in step one, AI technology companies may face the risk of infringement of copying rights. How to characterize the copying behavior of machine learning for copyrighted works and whether it belongs to the

---

<sup>45</sup> Zhu Jinlin, *The Rationality of Authorization of Chinese Simplified and Traditional Chinese from Translation Rights*, 4 EDITORIAL FRIEND 104, 106 (2013).

<sup>46</sup> *Id.*

<sup>47</sup> Zhang Yao, *Research on the Identification of Infringement of the Adaptation Right*, 41 J. OF HENAN INST. OF EDU. (PHIL. AND SOC. SCI. ED.) 33, 35 (2022).

<sup>48</sup> *Id.*

<sup>49</sup> *Id.*

<sup>50</sup> An, *supra* note 18, at 6.

<sup>51</sup> *Id.*

<sup>52</sup> Wu Handong, *Institutional Arrangements and Legal Regulation in Age of Artificial Intelligence*, 5 J. OF NORTHWEST UNIV. OF POL. SCI. & L. 35, 133 (2017); Xiong Qi, *Copyright Recognition of Artificial Intelligence Generated Content*, 3 ELEC. INTELL. PROP. 3, 8 (2017); Evan H. Farr, *Copyrightability of Computer-Created Work*, 15 RUTGERS COMPUTER & TECH. L.J. 1, 63-66 (1989); Li Xiaoyu, *Discussion on the Copyrightability and Right Allocation of Artificial Intelligence Products*, 6 ELEC. INTEL. PROP. 31, 33 (2018).

category of fair use are the problems faced by the copyright laws of various countries today, and they also challenge the copyright law of China.

### III. MACHINE LEARNING APPLIES FAIR USE DILEMMA ANALYSIS

#### A. Challenges of Machine Learning to the Fundamentals of fair use doctrine

Copyright law has a dual purpose. On the one hand, it aims to protect the rights and interests of authors, and on the other hand, it aims to protect the public interest.<sup>53</sup> Copyright law is a balancing act between public interest and exclusive rights. The use of copyrighted works by machine learning also has the problem of balancing public interest and exclusive rights.<sup>54</sup>

On the one hand, if copyright law adopts strong protection of exclusive rights, machine learning needs to obtain permission for using works one by one.<sup>55</sup> Using works without permission would also expose AI technology companies to infringement charges. Due to the large number of works contained in the input data, large damages would overwhelm AI technology companies and discourage the development of machine learning technology. This would lead to a flow of technological innovations within the field of AI to jurisdictions with more lenient copyright protection.<sup>56</sup> On the other hand, by adopting weak protection of exclusive rights and considering the use of copyrighted works by machine learning as fair use, copyright law enables the rapid development of machine learning technology, while ignoring the rights of authors.<sup>57</sup>

As an institutional tool to balance public interest and proprietary rights, fair use has worked well in some specific fields, including the field of machine learning.<sup>58</sup> However, in the issue of balancing interests in machine learning, fair use cannot balance the contradiction between the exclusive rights of copyright owners and the development of machine learning technology. The reason for the failure of the system is the issue of balance of interests in machine learning runs contrary to the basic principle of fair use.<sup>59</sup>

Fairness and justice are the bases of the fair use system.<sup>60</sup> According to Japanese scholar Katsumoto Masaakira, the purpose of restricting copyright is to prevent "abuse of rights" and allowing others to use works properly is "fair use of rights."<sup>61</sup> On the level of institutional purpose, fair use limits the scope of copyright, prevents the copyright owner's control from intruding into the domain that should be dominated by

---

<sup>53</sup> FENG XIAOQING, THE THEORY OF BALANCE OF INTERESTS IN INTELLECTUAL PROPERTY LAW 93-97 (2006).

<sup>54</sup> Gao Jiajia, *The Fair Use Analysis of Machine Learning from the Perspective of Typology*, 5 ELEC. INTELL. PROP. 18, 20 (2021).

<sup>55</sup> *Id.*

<sup>56</sup> See Benjamin L. W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM. J. OF L. & THE ARTS 45, 33 (2017).

<sup>57</sup> Jiajia, *supra* note 54, at 5.

<sup>58</sup> *Id.*

<sup>59</sup> *Id.*

<sup>60</sup> WU HANDONG, RESEARCH ON FAIR USE SYSTEM OF COPYRIGHT 31 (2013).

<sup>61</sup> Lu Mei, Liu Zhigang, *A Multidimensional Perspective on the Fair Use of Copyright Regime*, in RES. ON LIBR. SCI. 3 (2009).

the public, and destroys the delicate balance between maintaining incentives and safeguarding public freedom in copyright law.<sup>62</sup> From a utilitarian perspective, the purpose of the law is to bring the principles of the individual and society into a state of equilibrium.<sup>63</sup> By allowing subsequent authors to recreate without harming the profits or prospects of the original work, fair use allows human intelligence to progress and the cultural wealth of society to increase, and the individual and society are in a state of equilibrium.<sup>64</sup>

However, in machine learning, it is not an abuse of rights for most web users to prohibit machine learning from copying their works.<sup>65</sup> The value that machine learning derives from the expression of the work affects the profitability and prospects of the original work, which is not "in the domain of the public" and hardly constitutes "fair use."<sup>66</sup> But it is difficult to constitute "fair use." The application of fair use to machine learning will break the balance of copyright law's inherent interests.<sup>67</sup> The reason for this is the shift in the identity of the right holder and the user has led to a shift in the direction of the value of the work under the fair use system.<sup>68</sup>

In the past, network users were users of other people's works, and authors or publishers were the rights holders of copyright.<sup>69</sup> The user was in an economically disadvantaged position compared to the right holder, while the publisher and other companies were in an advantageous position.<sup>70</sup> Fair use is justified on the grounds of public interest and is appropriately restricted on the basis of protecting the right holder, so the value of the work fluctuates from the author or publisher to the public.

In the age of artificial intelligence, artificial intelligence technology companies have become users of other people's works, and Internet users are the rights holders of their own works. If the fair use system is applied to machine learning, the value in the work will flow from the vast number of network users to the AI technology company. The company, which is in the advantageous position of capital, will enjoy legal advantages and promote the expansion of "hegemony." The expansion of the "hegemony" will make the fair use system a tool for corporate services, thus diminishing the status of authors, which is against the principle of public interest priority and the spirit of balance. This issue directly conflicts with the legislative purpose of copyright law.<sup>71</sup>

---

<sup>62</sup> Jiang Ke, *Conduct in Fair Use Stage*, 186 L. REV. 6, 33 (2015).

<sup>63</sup> EDGAR BODENHEIMER, JURISPRUDENCE 229 (2015).

<sup>64</sup> Jane C. Ginsburg, *Fair Use for Free, or Permitted-but-Paid?*, 29 BERKELEY TECH. L. J. 1383, 1390 (2014).

<sup>65</sup> Jiajia, *supra* note 54, at 5.

<sup>66</sup> *Id.*

<sup>67</sup> *Id.*

<sup>68</sup> *Id.*

<sup>69</sup> Sobel, *supra* note 56, at 24. (Professor Benjamin likens this concept to "hegemony.")

<sup>70</sup> *Id.*

<sup>71</sup> Xiaoqing, *supra* note 53, at 93-97.

## *B. Machine Learning Challenges for Transformative Use*

### *1. History of the Development of Transformative Use*

Both mechanical learning and machine learning involve the unauthorized use of copyrighted works.<sup>72</sup> The copyright limitation system, with fair use at its core, provides the possibility of defenses against such actions.<sup>73</sup>

Section 107 of the U.S. Copyright Act establishes a four-factor standard for determining fair use, but there is no clear standard for the order in which the four factors should be considered, and how much weight each factor should carry in the determination of fair use.<sup>74</sup> To address the lack of interpretative power of the law, Judge Leval first proposed "transformative fair use" in his article, "Toward a Fair Use Standard."<sup>75</sup> He argued:

the use in fair use should be productive and should be done in a manner or for a purpose different from that of the original work. A mere repackaging or redistribution of the original work cannot constitute fair use. In other words, the secondary use should give new value to the original work, such as new information, new aesthetics, and new insights, for the use to enrich the cultural wealth of society and be protected by the fair use system.<sup>76</sup>

The earliest case on transformative use is *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 580 (1994). The plaintiff was the copyright owner of the popular song "Oh, Pretty Woman," which rap group 2 Live Crew parodied the rap song "Pretty Woman" while retaining the basic melody of the original song.<sup>77</sup> The Supreme Court held that 2 Live Crew's parody constituted fair use.<sup>78</sup> In our view, even if the work in question reproduces the basic melody of the original song, the parody may still constitute fair use. The act of parody is to use the content of the original work to create a derivative work, and it is very reminiscent of the original work when people are exposed to the work of parody. Furthermore, parody is "transformative" in that it adds new expressions, meanings, and messages to the original work.<sup>79</sup> Because the audience is different, parody does not replace the original work's audience.<sup>80</sup> Although the parody may cause damage to the market of the original work, such damage is not an aspect of

---

<sup>72</sup> Jiajia, *supra* note 54, at 5.

<sup>73</sup> *Id.*

<sup>74</sup> 17 U.S.C. §107 (2022) (sets forth the factors to be considered in determining fair use: (1) the purpose and nature of the use, including whether the use is commercial in nature or has a nonprofit educational purpose; (2) the nature of the work; (3) the amount and substantiality of the portion used in relation to the work as a whole; and (4) the effect of the use on the potential market for and value of the work.).

<sup>75</sup> Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 5, 1005 (1990).

<sup>76</sup> *Id.*

<sup>77</sup> See *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 577 (1994).

<sup>78</sup> *Id.*

<sup>79</sup> Xie Yuan, *Parody's Legitimacy and Limitation*, 10 J. OF CHONGQING UNIV. OF TECH. (SOC. SCI.) 65, 70 (2014).

<sup>80</sup> *Id.*

copyright law. If a harshly worded literary criticism dampens consumer demand for the original work, the decrease in sales of the original work is not caused by the literary criticism replacing the original work.

The guiding significance of *Campbell* is it resolved the ambiguity of the criteria for determining fair use under Section 107 of the U.S. Copyright Act.<sup>81</sup> Although there are still different controversies over the criteria for applying "transformative," it has become a consensus that transformative use can be used to determine whether secondary use constitutes fair use.<sup>82</sup> Since copyright law aims to encourage the creation and dissemination of works, and since any creative act cannot be completely independent of the borrowing of existing works, transformative use should be recognized as a legitimate basis for encouraging "secondary creation."<sup>83</sup> After *Campbell*, in determining whether a new work constitutes fair use, the more transformative the content, the less important other factors, such as whether it is for commercial purposes, and excessive consideration of other factors may hinder a proper determination of fair use.<sup>84</sup>

As information technology has evolved and computers have become more powerful in terms of their ability to reproduce, store, and analyze, they have also attracted an increasing number of copyright lawsuits, and courts have overcome the copyright risks of computer technology by applying the transformative use rule.<sup>85</sup> In *Kelly v. Arriba Soft Corp.*, the defendant was a search engine company that fed back thumbnail search results based on user searches.<sup>86</sup> Defendant's "web crawler" downloads the original images of these thumbnails collected by a "web crawler. It loads full-size images when it accesses image information, scales the images to thumb size, deletes the full-size copies, and saves only the thumb-size thumbnail images. When a user performs an image search, it provides thumbnails displayed to the user.<sup>87</sup>

The Ninth Circuit held that the plaintiff's images were "aesthetic works that provide[d] readers with an artistic experience," and that the defendant's scaling, a tool to help users index and engage with the images which did not involve the expression of a work or of the images to thumb size, was merely instrumental.<sup>88</sup> The low resolution of the defendants' thumbnails prevented any reader from using them for artistic appreciation.<sup>89</sup> The artists could only use the images for informational purposes, and Arriba's thumbnails were intended for transformative purposes and to provide guidance to users, as opposed to the original images' purpose of providing an aesthetic experience.<sup>90</sup>

---

<sup>81</sup> Xiang Jing, *Development, Disputes and Deliberation on the Fair Use System in US Copyright Law since the Campbell Case*, 12 ELEC. INTELL. PROP. 82, 83 (2016).

<sup>82</sup> *Id.*

<sup>83</sup> Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1109-10 (1990).

<sup>84</sup> See *Acuff-Rose Music, Inc.*, 510 U.S. at 569.

<sup>85</sup> Li Gang, *Research on Transformative Use: Based on the Judicial Practice of Judgment on Fair Use of Copyright*, 41 ZHONGNAN J. OF ECON. & L. 41, 138 (2017).

<sup>86</sup> See *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 907 (9th Cir. 2003). An automated computer program that navigates between web pages and indexes the pages it visits.

<sup>87</sup> *Id.*

<sup>88</sup> *Id.*

<sup>89</sup> *Id.*

<sup>90</sup> *Id.*

At the same time, the provision of the thumbnails did not harm the potential market for the plaintiff's work, but rather directed users to the plaintiff's website.<sup>91</sup> The defendant's actions did not stifle the plaintiff's artistic creativity and did not displace the plaintiff's work, therefore falling within the scope of transformative use.<sup>92</sup>

## 2. Whether Machine Learning Constitutes Transformative Use

Machine learning constitutes transformative use because the act of use stimulates the accumulation of public knowledge without weakening the incentives for innovation, and therefore is considered legitimate.<sup>93</sup> However, machine learning is a new technology different from mechanical learning, and machine learning breaks the theoretical foundation of the rules of transformative use in many ways, challenging the criteria for determining transformative use.<sup>94</sup>

First, machine learning is no longer "transformative" for the copying of copyrighted works. Transformative use includes both conversion of content and of purpose.<sup>95</sup> The use act of transforming content is a typical transformative use, and such use is mainly to make changes to the original work, such as criticism, commentary, and comic parody.<sup>96</sup> Transformative use refers to the use of a work for a different purpose without changing the expression of the original work, as in the case of *Kelly v. Arriba*. In this case, the defendant used the plaintiff's image in the form of a thumbnail, but its use did not constitute a substantial substitution for the original work, so it constituted a transformative fair use.

Machine learning does not alter the original work, but rather creates its own work at the request of the user, based on the previous "learning" content.<sup>97</sup> Therefore, the reproduction of the work by machine learning is not transformative in content.<sup>98</sup> As far as the use for the purpose of transformation is concerned, U.S. courts only recognize that the purpose of the original work is to make the public appreciate the expression of the work, and the purpose of subsequent use is different from the original work in the following three situations:

- (1) adjusting the scale and clarity of the original work so that the public cannot appreciate the work even if it is available, such as using pictures to report news, thumbnail search, etc.;
- (2) making substantial adjustments to the original work, although the purpose of use also includes allowing the public to read and appreciate, but the main purpose is to evaluate the original work;
- (3) reproducing the original work in its entirety, but providing only fragments or key information for public retrieval, so that the public

---

<sup>91</sup> *Arriba Soft Corp.*, 336 F.3d at 934

<sup>92</sup> *See id.*

<sup>93</sup> Yang, *supra* note 12, at 48.

<sup>94</sup> *Id.*

<sup>95</sup> Xiong Qi, *Transformative Use Interpretation in China Copyright Law*, 2 THE JURIST 124, 129 (2019).

<sup>96</sup> *See Acuff-Rose Music, Inc.*, 510 U.S. at 580. Campbell is an example of this.

<sup>97</sup> *Youhua & Yuansha*, *supra* note 28, at 73-75.

<sup>98</sup> *Id.*

cannot access the full text of the work, such as the construction of digital libraries. In contrast, the act of copying and creating autonomously by machine learning does not have the above three purposes, and therefore, the use of the work by machine learning is also not transformative in purpose.<sup>99</sup>

In 2015, Google launched an application called "Smart Reply" in Gmail, which automatically generates three replies by processing emails received by users, from which users can choose a reply without having to write it themselves.<sup>100</sup> The "Smart Reply" database contains many human-written sentences, and the team used a combination of data analysis and human review to train the machine to learn to use different words to convey the same meaning, while avoiding ambiguity.<sup>101</sup> This approach did not make the "Smart Reply" write wonderful prose, so Google Built a database with fiction as data input to improve the "Smart Reply" to make it more rhetorical and similar to a human tone.<sup>102</sup> However, the authors of the novels were not aware that their work was being used and did not receive any royalties.<sup>103</sup>

Confronted with allegations of copyright infringement by the authors of those novels used in the "Smart Reply", a Google spokesperson argued that the use of "Smart Reply" was not detrimental to the author's interests, and that the purpose of its use was entirely different from the author's purpose and fell within the scope of fair use.<sup>104</sup> However, the spokesperson did not specify why the purpose of such use was different from the author's purpose.<sup>105</sup> In fact, the "Smart Reply" uses the expression of the novel, which is the object copyright law is intended to protect. The use of the novel's database by "Smart Reply" largely misappropriated the author's original expression and exceeds the limits of fair use.

Second, machine learning can pose a threat to the potential market for the original work in copyright law. In a copyright dispute, *Tencent, Inc. v. Yingxun Technology, Inc.*, the plaintiff, Tencent, used its self-developed artificial intelligence application, Dreamwrite, to create a financial reporting article.<sup>106</sup> The defendant, Shanghai Yingxun Technology, copied the article in question and disseminated it to the public through its website "Home of Online Loan."<sup>107</sup>

---

<sup>99</sup> Hua Jie, *Transformative Use of Copyrighted Works*, 4 SCI. TECH. & L. 26, 28 (2019).

<sup>100</sup> Li An, *Copyright Law Analysis of Machine Learning Works*, 6 ELEC. INTELL. PROP. 60, 65 (2020).

<sup>101</sup> See Anjuli Kannan et al., *Smart Reply: Automated Response Suggestion for Email*, KNOWLEDGE DISCOVERY AND DATA MINING (Aug. 2016), <https://dl.acm.org/doi/10.1145/2939672.2939801>.

<sup>102</sup> See Alex Kantrowitz, *Google is Feeding Romance Novels to its Artificial Intelligence Engine to make its Products more Conversational*, BUZZFEED NEWS (May 4, 2016), <https://www.buzzfeednews.com/article/alexkantrowitz/googles-artificial-intelligence-engine-reads-romance-novels>.

<sup>103</sup> See Richard Lea, *Google Swallows 11000 Novels to Improve AI's Conversation*, THE GUARDIAN (Sept. 28, 2016), <https://www.theguardian.com/books/2016/sep/28/google-swallows-11000-novels-to-improve-ais-conversation>.

<sup>104</sup> *Id.*

<sup>105</sup> *Id.*

<sup>106</sup> Chen Xin, *Tencent v. Shanghai Yingxun wins first instance copyright infringement case - AI original creation also has copyright*, CHINA INTELL. PROP. NEWS (Jan. 15, 2020), [http://www.iprchn.com/cipnews/news\\_content.aspx?newsId=120713](http://www.iprchn.com/cipnews/news_content.aspx?newsId=120713).

<sup>107</sup> *Id.*

As to whether the article in question constituted a work, the court held that “The article was generated by the plaintiff’s creative team using Dreamwrite software, and its external performance met the formal requirements of a written work, and its performance content reflected the selection, analysis, and judgment of the relevant stock market information and data in the morning of that day, with a reasonable structure and clear expression logic, and with a certain degree of originality.”<sup>108</sup> Therefore, it is deemed to be a “work” in the sense of copyright law.<sup>109</sup>

There is still a controversy in academic circles as to whether artificial intelligence-generated works belong to “works” in the sense of copyright law.

On one hand, Wang Qian believes that the content generated by machine learning is not original and cannot be protected by copyright because it is the result of programming.<sup>110</sup> On the other hand, Xiong Qi believes that if the content generated by machine learning meets the requirements of copyright originality standard, the owner of machine learning can enjoy the copyright of the work.<sup>111</sup> However, Liu Ying believes that the intelligence of machine learning has already reached the level of children at the age of 6, and the content generated by machine learning under its own consciousness is not protected by copyright.<sup>112</sup> Moreover, Liu Ying believes that the intelligence of machine learning has reached the level of an 18-year-old child, and the content generated under its own consciousness should be given copyright protection according to Locke’s theory of natural rights and incentive theory.<sup>113</sup>

However, in judicial practice, even if the copyrightability of AI-generated materials is not recognized, the protection of AI-generated materials will be affirmed.<sup>114</sup> Compared to works created by humans, AI-generated works are from human works in terms of external expression and are far more productive than human works.<sup>115</sup> The threat posed by machine learning to the market of human works is self-evident, as machine learning “learns” from the reproduction of human works and forms the productions to compete with human works.<sup>116</sup> The ultimate basis for evaluating behavior in copyright law is the effect of the behavior on the incentive effect.<sup>117</sup> Machine learning weakens the economic incentive for authors and defeats the legislative purpose of copyright law.

---

<sup>108</sup> *Id.*

<sup>109</sup> *Id.*

<sup>110</sup> Wang Qian, *Qualitative research on the Content of Artificial Intelligence in Copyright Law*, 5 SCI. OF L. 148, 150 (2017).

<sup>111</sup> Qi, *supra* note 52, at 7.

<sup>112</sup> See generally Liu Ying, *A Preliminary Study on the Protection of Copyright Law for Artificial Intelligence Products*, 9 INTELL. PROP. 44 (2017).

<sup>113</sup> *Id.*

<sup>114</sup> See *Beijing Film Law Firm v. Beijing Baidu Netcom Science & Tech. Co., Ltd.*, Beijing 0491 Civil Judgment No. 239 (Jing 0491 Min Chu No. 239) (Beijing Internet Court 2018) (China). The court held that “although the analysis report does not constitute a work, it does not mean that it enters the public domain and can be freely used by the public.”

<sup>115</sup> Gao Jiajia, *The Fair Use Analysis of Machine Learning from the Perspective of Typology*, 5 ELEC. INTELL. PROP. 18, 25 (2021).

<sup>116</sup> *Id.*

<sup>117</sup> Ke, *supra* note 62, at 190.

Transformative use does not exist in China's fair use provisions, but courts sometimes invoke transformative use to adjudicate cases in copyright cases.<sup>118</sup> The court introduced transformative use by including it in the fair use statutory provision "appropriately quoting from the published works of others in a work for the purpose of introducing or commenting on a work or illustrating a certain issue" and using "illustrating a certain issue" as a synonym for the term "to illustrate a problem" is synonymous with "transformative use".<sup>119</sup>

In other words, the original work is used as an argument or material for the new work. However, machine learning use of others' works does not fall into this statutory category but is intended to learn, imitate, or draw on the expression of the work.<sup>120</sup> Therefore, machine learning's copying of others' works cannot be explained by transformative use in our country.

Meanwhile, China, as a member of the Berne Convention, the WIPO Copyright Treaty, TRIPs and other international conventions or treaties, should fulfill the international obligations of the "three-step test" established by these international conventions or treaties.<sup>121</sup>

The "three-step test" means that member states may make exceptions to copyright restrictions, but "only in specific circumstances and not in conflict with the normal exploitation of the work, and without reasonably prejudicing the legitimate rights and interests of the copyright owner."<sup>122</sup> It should be noted that all three of the above should be satisfied, and all of them are necessary.

#### IV. MACHINE LEARNING IS THE WAY OUT OF THE FAIR USE DILEMMA

##### *A. Analysis of the Necessity of Setting Statutory Licenses*

Copyright law takes the protection of the rights and interests of authors as the core legislative principle. The author's right is the first right – the native right.<sup>123</sup> The exclusive rights of authors should not be ignored. At the same time, machine learning technology has unique value for the convenience of human life and the progress of society and culture. As mentioned earlier, fair use is not desirable for machine learning, but it is also undesirable to require machine learning technologies to seek permission from the author whenever they use a protected work.<sup>124</sup> In the digital network era, there are "massive works" on the internet.<sup>125</sup> Behind them are "massive

---

<sup>118</sup> See Wang Xin v. Guxiang Info. Tech. Co., Ltd., Yizhong Min Chu No. 1321 Civil Judgment (Beijing No. 1 Intermediate People's Court 2011) (China); Shanghai Fine Arts Film Studio v. New Film Era Culture Communication Co., Ltd., Shanghai Intellectual Property Court 2015) Hu Zhi Min Zhong No. 730 Civil Judgment Shanghai Intellectual Property Court 2015) (China).

<sup>119</sup> Qi, *supra* note 95, at 126.

<sup>120</sup> *Id.*

<sup>121</sup> Zhang Chenguo, *Misreading and Re-interpretation of "Three-Step Test" and "Fair-use Doctrine" in Chinese Copyright Law*, 5 GLOBAL L. REV. 5, 8 (2016).

<sup>122</sup> *Id.*

<sup>123</sup> Handong, *supra* note 60, at 68.

<sup>124</sup> Sobel, *supra* note 56, at 33.

<sup>125</sup> Heping, *supra* note 31, at 139.

copyright holders” that will require AI technology companies to negotiate with each copyright holder.<sup>126</sup> This undoubtedly adds huge negotiation costs to the companies, and these costs will eventually become a constraint to the development of machine learning technology.<sup>127</sup>

The development of innovation requires the synergistic evolution of new technology and the legal system.<sup>128</sup> Since 1976, the U.S. Congress has passed amendments to ensure that the legislative purpose of copyright law is not undermined by new technological developments.<sup>129</sup> For example, the 1995 amendments to the Copyright Act added a right to digital audio transmissions; the 1998 amendments to the Copyright Act added provisions prohibiting circumvention of technological measures.<sup>130</sup> These copyright law amendments clearly demonstrate the intent of the U.S. Congress to prevent legal loopholes in copyright law due to the creation of new technologies.<sup>131</sup>

Article 22 of China's Copyright Law provides for 12 circumstances that can constitute "fair use."<sup>132</sup> Article 6 of the Regulations on the Protection of the Right to Information Network Dissemination provides for eight circumstances that can constitute "fair use" in the network environment within this scope.<sup>133</sup> However, these close-ended lists cannot solve the imbalance problem of conflicting interests in the era of artificial intelligence, and the lagging nature of the law has failed to meet the development needs of new technologies.<sup>134</sup>

In the author's opinion, the solution to this problem still depends on the provisions of copyright restrictions and exceptions. The key lies in clarifying to what extent machine learning can use copyrighted works on the internet. In other words, the way to solve the problem is to find a balance of interests between strong and weak copyright protections.

Because the fair use system without payment and license cannot be applied, and the license and payment for authorized use are required, it will make the negotiation cost of AI technology companies too high. Then, the statutory licensing system becomes the way out of the problem of whether machine learning can "copy" works under the framework of copyright system.

Statutory license means that the law explicitly stipulates that the performance of an act originally controlled by exclusive rights does not require the permission of the copyright owner but should be paid to the copyright owner.<sup>135</sup>

---

<sup>126</sup> *Id.*

<sup>127</sup> *Id.*

<sup>128</sup> Liu Ying & Xiao Jigang, *Japan's Big Data Legislation Adds a "Limited Provision of Data" Clause and its Implications for China*, 4 INTELL. PROP. 88, 89 (2019).

<sup>129</sup> Wan Yong, *Copyright Regulations for Deep Links on the Internet*, 6 STUD. IN L. AND BUS. 167, 169 (2018).

<sup>130</sup> *Id.*

<sup>131</sup> *Id.*

<sup>132</sup> (中华人民共和国著作权法) [Copyright Law of the PRC] (promulgated by P.R.C. LAWS, effective Oct. 27, 2001) art. 22.

<sup>133</sup> (信息网络传播权保护条例) [Regulations on the Protection of the Right to Information Network Dissemination] (promulgated by the State Council May 10, 2006, effective July 2006), art. 6.

<sup>134</sup> Li Mingxuan, *The Dilemma and the way out of the Rigidity of the Fair Use Rules of Copyright Law*, 11 VIEW ON PUBL'G 11, 52 (2022).

<sup>135</sup> WANG QIAN, COPYRIGHT LAW 370 (2015).

Statutory license reduces certain absolute rights in copyright to the right to legal remuneration.<sup>136</sup> This fuses the deprivation of the copyright owner's right to license with the protection of the copyright owner's right to remuneration, further promoting the dissemination of works based on the economic incentive mechanism of copyright law.<sup>137</sup> Statutory license, thus constitutes an important part of copyright restrictions and exceptions.<sup>138</sup>

At present, China's Copyright Law provides four statutory licenses in Article 23, Article 33(2), Article 40(3), Article 43(2), and Article 44, and a quasi-statutory license in Article 9 of the Regulations on the Protection of the Right to Information Network Dissemination.<sup>139</sup> However, these provisions are not yet sufficient to meet the challenges posed by machine learning to copyright in the digital era.

Therefore, from the perspective of legislative theory, an additional statutory license can be provided for the copying behavior of machine learning. Since the subject of liability for infringement of machine learning is the AI technology company, the statutory license fee should be paid by the AI technology company. In this way, the contradiction between the exclusive right and technological development can be balanced, and the interests of the company and the copyright owner can be considered.

On the one hand, by limiting the exclusive right of the author, the copyright owner can avoid refusing others to use their work because the consideration cannot be satisfied, and the development of machine learning technology is guaranteed. On the other hand, stipulating that the AI technology company pays a certain remuneration to the copyright owner ensures the economic incentive to the copyright owner, and reflects the respect for the exclusive right of the author.

### *B. Institutional advantages of statutory licensing*

The application of statutory licensing to machine learning can accommodate the values of justice and efficiency at the same time.<sup>140</sup>

First, the statutory license system can adjust for the justice of social relations concerning machine learning.<sup>141</sup> Under the licensing model, the prerequisite for users to legally use other people's works is to obtain the permission of the copyright owner.<sup>142</sup>

In the face of massive works, potential users use works without permission to pursue efficiency and reduce costs, damaging the exclusive rights of copyright owners.<sup>143</sup> The fair use system ignores the exclusive rights of copyright owners.<sup>144</sup> The

---

<sup>136</sup> HE MING, RESEARCH ON COPYRIGHT STATUTORY LICENSING SYSTEM 6 (2017).

<sup>137</sup> *Id.*

<sup>138</sup> *Id.*

<sup>139</sup> (信息网络传播权保护条例) [Regulations on the Protection of the Right to Information Network Dissemination] (promulgated by the State Council May 10, 2006, effective July 2006), art. 9.; (中华人民共和国著作权法) [Copyright Law of the PRC] (promulgated by P.R.C. LAWS, effective Oct. 27, 2001) art. 23, 33(2), 40(3), 43(2), 44.

<sup>140</sup> Youhua & Yuansha, *supra* note 28, at 73-74.

<sup>141</sup> *Id.*

<sup>142</sup> *Id.*

<sup>143</sup> *Id.*

<sup>144</sup> *Id.*

statutory license balances the conflict between the exclusive rights of the copyright owner and the user's right to use, ensuring that machine learning can use the work on a large scale while fully respecting the interests of the copyright owner, which is in line with the dual value of "protection plus restriction" of copyright law.<sup>145</sup>

Secondly, from the perspective of efficiency, the statutory licensing system allows machine learning to be used and paid for later.<sup>146</sup> This skips the negotiation between the copyright owner and the user, improving the efficiency of the dissemination and utilization of works.

Under the strict protection of copyright, machine learning needs to negotiate with and obtain permission from the copyright owner before using the work, which leads to inefficiency or ineffectiveness if the negotiation fails or is reversed after negotiation. Where machine learning applies fair use, there is no sufficient economic incentive for the author, which may lead to inefficiency in the output of the work. Therefore, the application of statutory licensing to machine learning allows copyright owners and machine learning to avoid these inefficiencies or ineffectiveness, achieving a win-win situation.

Statutory licensing is an alternative to the copyright market.<sup>147</sup> It eliminates the obstacles to the flow of rights caused by market failure by constructing an artificial government-regulated market.<sup>148</sup> Today, neither the market itself nor other systems of copyright can reconcile the contradiction between the development of machine learning technology and the exclusive rights of many authors.<sup>149</sup> The statutory license, which combines "limitation of rights" and "compensation," can both mitigate the "rigidity" of licensing and compensate for the "softness" of "fair use" in the game of interests between technological development and proprietary rights, and it provides a solution for the imbalance of interests.<sup>150</sup>

Copyright law in the Western world often solves the legal problems raised by new and recent technologies through the tax system.<sup>151</sup> The tax system has been a feature of European copyright law since the Federal Republic of Germany passed a law in 1965 to compensate copyright owners for the private copying of their products.<sup>152</sup> In the draft civil law report of the European Parliament Committee on Robotics, rapporteur Mady stressed that

for tax and social security contribution purposes, it may be necessary to introduce corporate reporting requirements, which would specify the extent and proportion of the contribution of robots and artificial intelligence to the economic efficiency of a company.<sup>153</sup>

---

<sup>145</sup> Youhua & Yuansha, *supra* note 28, at 73-74.

<sup>146</sup> *Id.*

<sup>147</sup> Xiong Qi, *Reflections on the Origin and Transplantation of the Copyright Law Licensing System*, 5 L. SCI. 72, 77 (2015).

<sup>148</sup> *Id.*

<sup>149</sup> Ming, *supra* note 136, at 6.

<sup>150</sup> *Id.*

<sup>151</sup> PAUL GOSTEEN, *THE TAO OF COPYRIGHT: FROM GUTENBERG TO DIGITAL ON-DEMAND* 135 (2008).

<sup>152</sup> *Id.*

<sup>153</sup> *Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics*, Eur. Parliament Doc. A8-0005/2017 (Jan. 1, 2017).

*Sony Corp. of Am. v. Univ. City Studios, Inc.*, the famous Digital Tape Recorder case also illustrates the preference of U.S. copyright law for a tax system to address technological issues.<sup>154</sup> Taxing artificial intelligence technology can indeed safeguard the future of this technology by avoiding infringement lawsuits that companies get into because of developing machine learning.<sup>155</sup> However, in the interests of copyright owners, this incentive is not targeted to effectively protect the rights of copyright owners. It is also too expensive in terms of legislative costs.<sup>156</sup> Therefore, compared to the tax system, it is more institutionally advantageous to apply statutory licenses to machine learning.

### C. Realization of statutory licenses

The statutory license may face multiple problems in the process of realization. AI technology companies need many works, which are scattered throughout the internet and even belong to unidentified orphan works<sup>157</sup> This forces AI technology companies to incur high search costs to find the authors.<sup>158</sup> Secondly, multiple works mean many copyright owners.<sup>159</sup> Because of this, AI technology companies need to negotiate with copyright owners individually to determine the royalty rate of each work, which undoubtedly brings high negotiation costs to the companies.

After works are published, it is often the AI technology companies that take the initiative to issue offers to copyright owners for transactions.<sup>160</sup> However, compared with AI technology companies, copyright owners are in a weak negotiation position and are unable to negotiate the terms of use with users on an equal footing.<sup>161</sup> It is doubtful whether the royalty rate finally determined can meet the needs of copyright owners' interests and thus guarantee the realization of the incentive purpose of copyright law.

Lastly, considering that it is difficult for copyright owners to keep track of the utilization of their works, even if there is infringement, it cannot be detected in time.<sup>162</sup> After the evaluation of transaction cost and risk, the AI technology company may consider the transaction cost to be higher than the risk and use the work without

---

<sup>154</sup> See *Sony Corp. of Am. v. Univ. City Studios, Inc.*, 464 U.S. 417, 430 (1984). The result of the case was the Audio Home Recording Act of 1992, which was passed by Congress. First, it required the addition of a serial copy management system (Serial Copy Management System) to these digital audio devices. Second, it requires manufacturers of blank digital audio tapes and recording devices to pay a statutory levy of 3% of the sale price for tapes and 2% for recording devices. These levies are divided by the Copyright Office into two funds, 2/3 of which is the phonogram fund and 1/3 is the musical composition fund.

<sup>155</sup> Jie, *supra* note 99, at 27.

<sup>156</sup> *Id.*

<sup>157</sup> See generally Zhang Xiaoqia & Zhao Yuan, *On the Fair Use of "Orphan Works" in China*, 10 INFO. RES. 25 (2019).

<sup>158</sup> *Id.*

<sup>159</sup> Zhongfa & Yulu, *supra* note 28.

<sup>160</sup> Wei Zhi, *On the Right of Management of Copyright Collective Management Organization - Reflections on the Revised Draft of Copyright Law*, 3 STUD. IN L. AND BUS. 75, 76 (1999).

<sup>161</sup> *Id.*

<sup>162</sup> Zhongfa & Yulu, *supra* note 28, at 70.

authorization.<sup>163</sup> Waiting for the copyright owner to discover the fact of infringement, file a lawsuit and make compensation will make the statutory licensing system null and void.<sup>164</sup>

The author believes that the copyright collective management system can better solve the above problems and help the realization of statutory license. The purpose of the copyright collective management system is to reduce the search cost, transaction cost, and legal risk of the users of works.<sup>165</sup> At the same time, the copyright collective management supplements the deficiency of the copyright owners in terms of mastering the utilization of works and the ability to defend their rights to promote smooth transactions.<sup>166</sup>

On one hand, the copyright collective management organization, after being authorized by the copyright owner, can negotiate with the AI technology company on the issue of royalty rate and sign a license agreement in its own name.<sup>167</sup> Therefore, AI technology companies do not need to search for copyright owners and negotiate with them individually, greatly reducing the search and negotiation cost.<sup>168</sup> The lower the cost of fulfilling obligations, the higher the probability that AI technology companies will fulfill their obligations and will not take the legal risk of bearing infringement liability, which is conducive to ensuring the true realization of copyright.<sup>169</sup>

On the other hand, for the copyright owners, the copyright collective management organization will collect the utilization of works on their behalf, enhance the bargaining power of the copyright owners in the negotiation, and help them to defend their rights when necessary.<sup>170</sup> It can be argued that the design of the statutory licensing system reflects the "two-pronged" approach of reducing transaction costs and safeguarding the rights and interests of authors.<sup>171</sup>

The Amendment to the Copyright Law (Draft) draws on the fee arbitration mechanism in the U.S. and states that if the negotiation of royalties fails, one can apply to the competent copyright authority for arbitration or file a lawsuit in court, which is worthy of recognition.<sup>172</sup> However, on the issue of applying statutory licenses to AI technology companies, the system of copyright collective management organizations should be further improved in order to better realize the system's function of balancing interests.

How to improve the system of copyright collective management organizations. In our view, first, the digitization of works should be strengthened. After works are digitized, it is convenient for the copyright collective management organization to preserve and manage works in a unified manner, and for the AI technology companies to use works for machine learning.<sup>173</sup> Second, improve the registration of personal

---

<sup>163</sup> *Id.*

<sup>164</sup> *Id.*

<sup>165</sup> Xiang Bo, *Copyright Collective Management Organizations: Market Functions, Role Arrangements and Pricing Issues*, 7 INTEL. PROP. 68, 69 (2018).

<sup>166</sup> *Id.*

<sup>167</sup> LI CHEN, KEYWORDS OF INTELLECTUAL PROPERTY LAW 192 (2006).

<sup>168</sup> *Id.*

<sup>169</sup> *Id.*

<sup>170</sup> Liu Xiaochun, *Technological Innovation may Promote Legal Licensing of Online Copyright*, 5 CHINA TRADE J. 9, 10 (2014).

<sup>171</sup> *Id.*

<sup>172</sup> LI MINGDE, U.S. INTELLECTUAL PROPERTY LAW 189 (2003).

<sup>173</sup> Zhongfa & Yulu, *supra* note 28, at 69.

information of copyright owners so the copyright collective management organizations can contact the copyright owners and transfer the royalties in a timely manner to guarantee the realization of the incentive function of the copyright law.<sup>174</sup> For orphan works, the copyright collective management organization should publish the information of works to search for the copyright owners of works. If the copyright owners cannot be identified, even after a diligent search, they should be registered. Again, the autonomy of intent is respected in the negotiation of the determination of royalty rates.

The fees for statutory licenses in China are generally determined by administrative organs. Because of the need to maintain the stability of administrative regulations, the fee rates are not compatible with the need for flexible pricing of works in market transactions and lag behind the actual requirements of copyright transactions.<sup>175</sup>

Therefore, both parties to the transaction should be allowed to negotiate on an equal footing and determine the royalty rates for works based on the actual situation of the market. When both parties cannot agree, the royalty rate is determined through arbitration or litigation.

Finally, after the transaction agreement is signed, technical means are used to control the scope and duration of use of the work to guarantee that the work is used reasonably. For example, to prevent AI companies from using the work in ways other than machine learning, the work can be encrypted using "Federated Learning" technology to ensure that AI companies can only use the work for the training of machine learning.<sup>176</sup>

## V. CONCLUSION

The emergence of machine learning technology is known as the "last invention," based on the widespread use of the internet, and based on the application of widely distributed sensing technology, big data storage, and communication technology.<sup>177</sup> Based on the exponential growth of data scale, the world has entered the era of big data.<sup>178</sup> The operation of programs in computers involves many unauthorized copyrighted works on the internet.<sup>179</sup> Mechanical learning, due to its strong instrumental properties and its inherent lack of self-improvement, can only derive value from the ideas or facts of the work and can often be included in the category of fair use.<sup>180</sup> Machine learning, on the other hand, is a digital abstraction of the human

---

<sup>174</sup> *Id.*

<sup>175</sup> Huang Xianrong & Chang Jialing, *Research on the Development Dilemma and Countermeasures of China's Copyright Collective Management Mechanism*, 1 CHINESE EDITORS J. 44, 46 (2017).

<sup>176</sup> *Id.* Federated learning is an emerging artificial intelligence technology that unfolds efficient machine learning across multiple participants and multiple computational nodes by securing information and privacy during data exchange through encryption.

<sup>177</sup> Handong, *supra* note 60, at 128-29.

<sup>178</sup> *Id.*

<sup>179</sup> Zhongfa & Yulu, *supra* note 28, at 68.

<sup>180</sup> Jie, *supra* note 99, at 28.

brain that exhibits autonomy and anthropomorphism and can follow up after imitating the original expression of a work, which may infringe the author's copyright.<sup>181</sup>

Neither strong nor weak copyright protection can balance the contradiction between technological development and proprietary rights, and between users of works and rights holders of works.<sup>182</sup> It is difficult for machine learning to meet the basic principles of fair use, and it is also impossible to apply the criteria for determining transformative use.<sup>183</sup>

The application of the fair use doctrine to machine learning faces a dilemma. The solution to this problem lies in seeking a balance between strong and weak protection of copyright. This balance can be achieved by applying statutory licensing to AI technology enterprises and, at the same time, improving the collective copyright management system with a view to resolving the contradiction between the development of AI technology and the protection of copyright owners' rights.<sup>184</sup>

---

<sup>181</sup> *Id.*

<sup>182</sup> Zhongfa & Yulu, *supra* note 28, at 75.

<sup>183</sup> *Id.*

<sup>184</sup> Meng Lei, *Collective Management of Copyright in the Age of Intelligence: Challenges, Reflections and Reconstructions*, in PUBL'G RES. 1 (2020).