ISSN 2616-5783 Vol.5, Issue 18: 112-117, DOI: 10.25236/AJHSS.2022.051819

The Copyright Dilemma and Solution for "Data-Driven Creation"

Jiehua Lu

School of Law and Intellectual Property, Foshan University, Foshan, China
lujiehua@fosu.edu.cn

Abstract: *"Data-driven creation" means that data plays an important role in the creation of works and distribution of content. Big data has brought new challenges to the copyright legal regime while driving innovation in the copyright industry. First, the data-driven creation of works has shifted from "supply-oriented" to "demand-oriented", and the main body of creation has shifted from "author-centrism" to "reader-centrism". Second, the free use of data is the key to creative quality, and the boundaries of fair use should be appropriately expanded. Finally, data-driven creation brings negative problems such as algorithmic bias, information cocoon, and data monopoly, and a collaborative governance approach should be sought to ensure the healthy development of the industry.*

Keywords: *Data-driven creation; Copyright; Reader-centrism; Fair use; Data monopoly*

In the era of Big Data, data is affecting all areas of human society in various ways, and innovation is being generated by data. At a macro level, data-driven creativity is one aspect of data-driven innovation, which was first formally introduced in the "Data-driven Innovation: Big Data for Growth and Well-being", published by the Organization for Economic Co-operation and Development (OECD) in October 2015, referring to the role of big data in driving change in various areas such as product innovation and process innovation. The change is demonstrated by the following, firstly, at the creation level, massive data constitute a database for machine learning, which can be used as material for algorithm creation, applied to news robots writing news, music creation and other scenarios; secondly, at the content distribution level, data-driven algorithms analyze user data to discover their consumption preferences and needs, which can become an important basis for copyright product customization and realize precise marketing, such as Netflix, Amazon, Google, etc. have already used big data to develop and market copyright products and expand their copyright market. These typical scenarios of data-driven creation have gradually become popular, and this process is raising some new copyright issues, which will pose new challenges to the existing copyright law. Based on this, this paper examines the impact of data-driven creation on the basic categories of copyright system, such as subject matter and fair use, and the proposes solutions for such issues. At the same time, this paper analyzes the emerging negative effects of algorithmic bias, information cocoon, and data monopoly, and then proposes corresponding adjustment solutions.

1. The creation of works and creative subjects under "data-driven creation"

An analysis of copyright in the context of data-driven creativity requires an understanding of how the basic concepts of the copyright system, such as "work", "author", "creation", have been transformed by the data-driven creation model. As the way works are created shifts from being "supply-led" by those who supply content (creators) to being "demand-led" by those who demand content (readers, viewers), the "author-centrism" of traditional copyright theory gradually shifts to "reader-centrism". The author-centrism of traditional copyright theory is gradually shifting to reader-centrism, and the right to interpret works is shifting to readers.

1.1. Creation of works: from "supply-oriented" to "demand-oriented"

Under the traditional creation model, creators generally use their own creative interests, wishes or pre-defined markets as the starting point for personalised creation, which is a supply-oriented creation. However, in a data-driven creation model, a large amount of data provides creators with material that can be mined, helping them to create more targeted work. Moreover, this data-driven creation and content distribution significantly increases the market success rate. For example, the success of some

ISSN 2616-5783 Vol.5, Issue 18: 112-117, DOI: 10.25236/AJHSS.2022.051819

companies, such as Netflix and Amazon, in the copyright market is largely due to their vast amount of subscriber data, which has become an important asset and is reshaping the copyright market.

The value of data is often reflected in technologies such as data mining. In the data-driven copyright industry, creators or content providers can use data to understand user preferences in order to create or deliver works accordingly. As one of the data mining technologies, user profiling can be used to understand user preferences and become an important and effective way for creators to create works. Generally speaking, user profiling involves several steps as collecting user data, processing and extraction modelling, such finally resulting in the generation of user tags, which are characteristic identifiers based on user data and help to provide customised products and services.[1] With the help of user profiling and other data mining technologies, creators of works are able to create works in response to user needs and content providers are able to distribute content according to user preferences.

It is clear that the trend of data-driven creation is shifting from the traditional “supply-oriented” to “demand-oriented” creation. Compared to the traditional “supply-oriented” creation model, the “demand-oriented” creation model under data-driven creation has shown its great advantages. In terms of creative efficiency, with data-driven creation, creators can use data mining technology such as user profiles to grasp user preferences and determine the direction of creation, which greatly improves creative efficiency compared to traditional creation methods. In terms of marketing, data-driven creation has changed the traditional approach of producing now and selling later, and is based on user preferences, helping to increase its market success rate. For example, in the book publishing industry, publishers have used big data to improve the accuracy of the book publishing supply chain, providing useful experience for business model innovation on the supply side of book publishing.

1.2. The subject of creation: from “author-centrism” to “reader-centrism”

The above-mentioned shift from a “supply-oriented” to a “demand-oriented” approach to the creation of works has posed new challenges to the copyright legal system. This is reflected first and foremost in the authorship and personality rights regime in copyright law. Under the data-driven creation model, works are no longer personalised creations of the author, but rather products that are customised according to data mining technology to accurately identify user needs, which has a certain impact on the author-centric concepts emphasised by traditional copyright theory.

Traditional copyright theory emphasizes that a work is an extension of the author's personality, namely the “author-centric” view.[2] Author-centrism requires that the work must reflect the personality of the author and that the work can only be derived from the author's individual creation. From the philosophical perspective, “author-centrism” is the legal embodiment of the recent philosophical paradigm of subjectivity in the context of the Enlightenment and the human rights revolution.[3] Under the influence of the philosophical paradigm of subjectivity and author-centredness, legislation has established a modern author-centred copyright system.

The opposite of “author-centredness” is “reader-centredness”, which is the result of structuralist thinking, in which the work is no longer the soul of the author, but the deeper structure itself, and the text is merely a copy of the structure, thus the author as subject is relegated to the background. The literary scholar Roland described this as the death of the author, and the philosopher Foucault suggested that “it does not matter who the author is”, “it is not the author who writes, but rather the way in which the discourse produced in a given era is shaped in the text”. All these formulations suggest that at the heart of reader-centrism lies the separation of the author from the work, the meaning of which is born in the reader, not shaped by the author.

If the impact of structuralist thinking on “author-centrism” is not yet complete, the data-driven model of authorship will certainly accelerate the impact on “author-centrism”, leading to a real shift towards “reader-centrism”. The data-driven creative model will accelerate the onslaught on author-centrism, leading to a real shift towards reader-centrism. In a data-driven creation model, the author incorporates the needs and preferences of the reader into the work, and his or her personal will and individuality are eliminated. At this point, it is clear that the work is no longer an extension of the author's personality, but a message transmitted from reader to reader, a product of the reader's collective creation. In a certain sense, the creator is more like a “producer” of the work.[4]

Therefore, it is important to rethink the position of authors themselves in the creation of works. In the gradual shift from author-centrism to reader-centrism, the way to respond to this change in reality is to remove the premise that human authors are copyrightable, and to shift the copyrightable element of

ISSN 2616-5783 Vol.5, Issue 18: 112-117, DOI: 10.25236/AJHSS.2022.051819

works to "reader-centrism". Data-driven authorship has become a reality and an unstoppable trend, and with technological advances and the information revolution, artificial intelligence may play an important role in the process of data-driven authorship, and may even be able to create works independently. The element of "human creation" will then lose its relevance, and the copyrightable element of a work should shift to "reader-centrism" in response to changes in social reality and technological development, and to the paradigm shift in data-driven creation.

2. Fair use under "data-driven creativity"

Under the data-driven creation model, whether data can be freely accessed and used is the key to the quality of creation, and since such data may include a large number of books, audio, video and other copyrighted works, the process of inputting, storing and using data at this time may face infringement difficulties. How to reasonably define the boundaries of the use of data in order to build fair use rules that fit the development of new technologies is a problem that must be solved in the development of algorithm-driven creation.

2.1. The copyright dilemma of "data-driven authoring" and the solution

Take data-driven artificial intelligence writing as an example, while traditional creative writing is basically based on rich life experience, skilled writing skills and profound ideology, the basic conditions for AI writing can be deconstructed as follows: first, a sufficient database, which serves as a reserve of writing materials, is a prerequisite for intelligent writing for subsequent extraction; second, algorithms, which are logical instructions for solving problems and represent strategies for solving problems in a systematic way; third, arithmetic, which is the formalisability of the language and expression of the laws of writing.[5] Amongst these conditions, the storage and extraction of data is the basic prerequisite. However, the storage and use of data may face the possibility of infringement, as it may constitute infringement if the database contains works that are not authorised for use by the copyright owner.

In terms of solutions to the copyright dilemma of data-driven creativity, the current copyright regime has some, albeit imperfect, solutions, such as the "opt-out" implied licence model, the fair use model and the statutory licence. Specifically, an "opt-out" implied licence requires the right owner to make an opt-out decision and is deemed to be a licence if the right owner does not declare that it cannot be used.[6] The "statutory licence" model allows for no prior permission to be obtained but remuneration to be paid afterwards. The fair use model allows for the application of fair use exceptions to copyright for data mining, for example, the EU Directive on Copyright in the Digital Single Market specifies exceptions to copyright for text and data mining purposes.[7]

For the above solutions, the fair use model is a more viable solution to the copyright dilemma of data-driven creativity than the other models. Since the "opt-out" implied licence model needs high institutional costs, and the statutory licence model means the shift of costs to the user, which does not encourage the development of industries related to data-driven creativity.In contrast, the fair use model allows users to use the work for training data freely, which helps to expand the scope and number of databases, facilitate the development of algorithmic creativity and minimise algorithmic bias.

2.2. Fair use in "data-driven creativity": an interpretative path as transformative use

The legitimacy of the fair use regime lies in overcoming the negative effects of "overly broad copyright protection that hinders scientific and technological progress".[8] With the development of technology, the traditional fair use approach is often unable to cover new types of fair use, and the courts need to interpret the factors of fair use to include some new types of fair use, among which "transformative use" is a rule of judgment developed in recent years to deal with new types of fair use.

The significance of transformative use lies in the reshaping of the value orientation and interpretative focus of the elements of fair use judgment.[9] Under the traditional approach, an important factor in determining whether a work constitutes fair use is the determination of whether the purpose of use is commercial. Unlike the traditional path of judgment, transformative use takes transformativeness as the criterion, which means that it focuses on the perception of the audience of the work, and regardless of whether it is a commercial use or not, as long as the new work has a transformative or creative change from the original work, it is a fair use. For example, in the case of the Google Library Project, the court held that as Google's purpose in copying the original work was to make it available to

ISSN 2616-5783 Vol.5, Issue 18: 112-117, DOI: 10.25236/AJHSS.2022.051819

users for searching and displaying clips, this was a function of the original work, which constituted a transformative use based on the original work and was fair use.[10]

Obviously, under the premise of satisfying other constitutive elements, the deep learning behavior of artificial intelligence under data-driven creation meets the requirements of transformative use and constitutes fair use. Firstly, the content of algorithmic creation has content transformative, through learning and analysis of a large amount of data, the generated content of algorithmic creation has formed new content, which can constitute content transformative. Secondly, the non-expressive use of algorithmic creation is functionally transformative, through data mining and restructuring, algorithmic creation has transformed from satisfying the appreciation needs of the audience of the work to providing education, search and other services, which makes the original work have a new function and constitutes functional transformativeness.

3. The negative effects of "data-driven creativity" and its regulation

As mentioned in Part I above, data-driven creation can accurately identify user preferences and needs, which is of great value to the creation and marketing of copyrighted products. However, data-driven creation has also brought about some negative effects, typically algorithmic bias and information cocooning, and moreover, the data monopoly which affects free competition in the copyright industry. These negative aspects of data-driven creativity will affect the development of the copyright industry and require appropriate legal intervention when it is difficult to address them through the market.

3.1. Overcoming the "information cocoon" dilemma: strengthening the quality control of algorithmic creation

In the era of big data, the content distribution model based on the number of clicks by users has gradually become the main way for users to access content, which seems to facilitate users and cater to their preferences, but on the other hand, this has seriously affected the quality of content distribution. For example, some service providers use "headlines" to please their readers and recommend them to read in order to attract attention.

At the same time, algorithm-led content delivery also brings about the "information cocoon" effect, which increases the cost of accessing multiple information for users and tends to lead to self-containment. On the one hand, the content pushing method based on user data reduces the cost of users' access to information in a certain field, on the other hand, it greatly increases the cost of users' access to information in other fields, and such personalized recommendations actually not only fail to reflect users' real needs, but also make users fall into the situation of "information cocoon".

Addressing the negative effects of algorithmic bias and information cocoons requires quality control by content platforms on the one hand, and the support of relevant systems on the other. For content creators or content platforms, corresponding manual intervention should be carried out to avoid algorithmic bias. In addition to self-optimisation by content platforms, the relevant authorities should establish algorithmic accountability and improve algorithmic transparency, which are important safeguards to improve the quality of copyright products and achieve cultural prosperity.

3.2. Regulating data monopolies: maintaining fair competition in the data market

With the increasing value of big data for the production and marketing of copyright products, data has become an important capital element for market competition, which in turn will lead to data becoming a barrier for competitors to enter the market, and the "winner-takes-all" monopoly position of the relevant operators will affect free competition. On the one hand, data operators use their access to user data to increase their market success and dominate the copyright market, while on the other hand, they use their dominant position to monopolise data and deny access to others, which is not conducive to the orderly and healthy development of competition in the market.

In the face of these data monopoly issues, the current thinking on anti-monopoly law cannot effectively address them. As the antitrust law mainly adopts the "turnover" standard as the threshold for filing, and the Internet economy is unique in that the turnover of some data-driven mergers and acquisitions may not meet the turnover standard at all and therefore do not need to be declared,[11] the current threshold for filing under the antitrust law does not regulate these data monopolies.

ISSN 2616-5783 Vol.5, Issue 18: 112-117, DOI: 10.25236/AJHSS.2022.051819

In the face of the threat of data monopolies to free competition in the copyright market and the limitations of current antitrust, attention should be paid to changes in the regulatory thinking of antitrust law. Specifically, firstly, in order to better regulate data-driven mergers and acquisitions, diversification of the reporting criteria for operator concentrations should be promoted. Secondly, data can be considered as an "essential facility" to promote data openness.[12] Finally, policymakers and enforcement agencies should focus on the new mission of antitrust law in the data era. At present, some superplatforms hold data monopoly power to influence the market, dominate the industry by virtue of data dominance, and even manipulate social opinion by controlling output content. Typically, some Internet mega-platforms are investing in and controlling various media platforms, and deeply influencing and even manipulating media communications. These phenomena have to be alarmed, and a concerted approach to governance should be sought to promote the healthy development of the copyright industry in the data-driven era.

4. Conclusion

The development of industrial practice requires simultaneous changes in the legal system. Data-driven creativity is driving the development of the copyright industry, and just as every information technology revolution in history has driven changes in the copyright system, the trend of data-driven creativity is driving a conceptual renewal of the copyright system and changes in the rules of the basic categories of subjects, rights and limitations of rights in the copyright system. First, the data-driven creation model has shifted the way works are created from "supply-oriented" to "demand-oriented", and the traditional copyright theory of "author-centric" creative subjects has been challenged. The position of the "author-centric" creative subject of traditional copyright theory has also been challenged, and the transformation of "reader-centrism" in the data-driven creation model has shifted the right of interpretation of works to readers. Secondly, under the data-driven creation model, the freedom of access to and use of data is the key to the quality of creation. As such data may include a large number of books, audio, video and other copyrighted works of others, the process of inputting, storing and using data may face infringement difficulties. Finally, in the face of the negative effects of data-driven creation, such as algorithmic bias, information cocoons and data monopolies, data companies are required to control the quality of algorithmic creation, while supporting systems such as algorithmic accountability can be established, and a synergistic approach to governance can be sought by updating the concept of anti-monopoly law. We are in a new era of copyright system development in the data era, and in the face of new problems brought about by data-driven creation, we should not only follow the basic concept and historical logic of the copyright system, but also update the concept and try to break through in order to realize the legislative intent of the copyright system to promote social and cultural development, and promote the orderly and healthy development of the copyright industry in the era of data creation.

References

[1] Niu Wenjia, Liu Jiqiang, Shi Chuan. User Web Behavior Pictorial - User Web Behavior Pictorial Analysis and Content Recommendation Application in Big Data [M]. Electronic Industry Press, 2016:4.

[2] Lin Xiuqin, Liu Wenxian. Author-centrism and its Crisis of Legitimacy - a Philosophical Investigation based on the Authorship System [J]. Journal of Yunnan Normal University (Philosophy and Social Science Edition), 2015(2).

[3] Liu Wenxian. From Creative Authorship to Functional Authorship: The Rise and Fall of Copyright Author-centrism in the Perspective of the Subject Paradigm [J]. Xiamen University Law Review, 2016(2).

[4] Zhang Yongqing. Authorship in the Historical Process - Four Dominant Paradigms of Western Authorship Theory [J]. Academic Monthly, 2015(11).

[5] Zhang Yonglu, Liu Weidong. Artificial Intelligence Writing: a New Landscape for Creative Writing [J]. Exploration and Controversy, 2021(3).

[6] Wang Guozhu. Analysis and Legislative Construction of the Implied License of Copyright "Opt-Out"[J]. Contemporary Jurisprudence, 2015(3).

[7] Wang Wenmig, Gao Jun. Copyright Exception Rules for Library Information Analysis in the Era of Artificial Intelligence [J]. Library Forum, 2020(9).

[8] AMY ADLER. Fair use and the future of art [J]. New York University Law Review, 2016, 91(3)

[9] Xiong Qi. The Local Law Interpretation of Transformative Use of Copyright [J]. The Jurist,

ISSN 2616-5783 Vol.5, Issue 18: 112-117, DOI: 10.25236/AJHSS.2022.051819

2019(2).
[10] Authors Guild, Inc. v. Google Inc. 954 F. Supp. 2d 282 (S.D.N.Y. 2013)
[11] Ye Jun. Research on the Legal Definition Model of Operator Concentration [J]. Chinese Jurisprudence, 2015(5).
[12] Sun Jin, Zhong Yuan. Antitrust Law Analysis of Data Constituting Essential Facilities in the Era of Big Data [J]. Electronic Intellectual Property Rights, 2018(5).