

Non-copyrightability of Chat GPT Generated Content and Risk of Copyright Infringement

Xiangfeng Gao

Anhui University of Finance and Economics, Bengbu, 233000, China.

Abstract

As a phenomenal technological application, Chat GPT not only brings new possibilities to various industries, but also challenges the existing legal system, and its generated content in natural language generation applications has triggered debates on copyrightability and rights attribution. Based on the analysis of copyright law, technical development of Chat GPT and the type of content generated, this paper argues that the content generated by Chat GPT in natural language generation application scenarios cannot constitute a "work" under the existing copyright law system, and that the production of content and the training of the model challenge the core concept of copyright and the principle of fair use, respectively. Content production and model training challenge the core concept of copyright and the principle of fair use, respectively, and the utilisation of generated content may lead to ethical and copyright infringement risks.

Keywords

Chat GPT, Generated Content, Non-Copyrightability, Copyright Risk.

1. Introductory

On November 30, 2022, Open AI released an artificial intelligence dialogue system called Chat GPT (Chat Generative Pre-training Transformer). As Chat GPT is able to provide users with exhaustive answers in multiple knowledge areas, write software code, and even produce articles similar to those written by real people through chat dialogues, it has attracted widespread attention worldwide since its launch, with monthly active users exceeding 100 million in two months' time, and a number of companies are now laying out and developing products and services related to generative AI. [1]As a phenomenal technology application, Chat GPT not only brings new possibilities to several industries, but also challenges the current legal system. "Copyright has been a son of technology from the beginning", and today, new technologies are also testing the ability of copyright law to regulate the market for information and content products. Chat GPT is essentially a specific application of Artificial Intelligence Generate Content (AIGC). Chat GPT is essentially a specific application of Artificial Intelligence Generate Content (AIGC). In recent years, academics have discussed the legal attributes of AI-generated content, but have yet to reach a consensus: a few scholars have taken a modest stance on the copyrightability of AI-generated content, while the majority of scholars believe that AI-generated content constitutes a "work" in the sense of the copyright law. The "Objectivist Criterion" and the "Instrumental Approach". Based on the existing research, the technological development of Chat GPT and the type of content generated, this paper analyses the copyrightability of Chat GPT-generated content, and discusses the risk of infringement that may arise during the operation of Chat GPT. The paper argues that Chat GPT-generated content does not constitute a "work" under the existing copyright law system, that content generation and model training challenge the core concept of copyright and the principle of fair use, respectively, and that the utilisation of generated content may give rise to ethical and copyright infringement risks.

2. Artificial Intelligence Technologies, Artificial Intelligence Generators and the Copyrightability Controversy

As an important branch of computer science, Artificial Intelligence (AI) was formally proposed in 1956 at a seminar on machine simulation of intelligence; after decades of development, it is known as one of the three cutting-edge technologies of the 21st century, together with genetic engineering and nanotechnology. At present, the academic community has not yet formed a unified view on the definition of "Artificial Intelligence", but the existing definitions basically summarise the basic idea and content of this discipline, that is, through the computer simulation of human thinking process and intelligent behaviour. [2]Accompanied by the rapid development of mobile Internet, big data and other technologies and infrastructure, artificial intelligence has made rapid progress in the 21st century. Previous AI mostly focused on the development and application of analytical capabilities, i.e., the analysis of data to discover the laws and patterns, personalised recommendation algorithms are typical application scenarios of AI technology at this stage; in recent years, AI is no longer limited to the analysis of the existing content, but is able to autonomously generate multiple forms of new content based on data training and algorithmic generation models, which achieves the goal of "creating" new content from the perception and understanding to the generation of "creation". The leap from perception and understanding to generation and "creation" has been achieved. At present, robots can already "create" music, draw pictures, write poems and novels, etc. Microsoft's AI product "Xiaobing" generated AI poetry collection "Sunshine Lost Glass Window" was officially published in May 2017. The emergence of AI-generated materials has also brought a series of new challenges to copyright law, and in recent years, academics and the industry have discussed issues such as the copyrightability of AI-generated content.

Scholars have not yet formed a unified opinion on the legal attributes of AI-generated works. Some scholars are of the view that AI-generated materials cannot constitute "works" in the sense of copyright law. A more representative viewpoint is that the formal "originality" of AI-generated content is not a sufficient condition for constituting a work, and that only the intellectual achievements of human beings can be protected as a work under the copyright law. Some scholars take the legislative purpose of the copyright system as the starting point for analysis, and believe that the pursuit of richness and diversity of works in the copyright law embodies the "disorder" orientation; at the present stage, the underlying logic of AI work is still to learn, make use of the laws and choose the optimal solution, and this process is from the "disorder" to the "diversity" orientation. [3]This process is a shift from "disorder" to "order", which deviates from the purpose and standard of the copyright legal system, and therefore, AI generated works do not have copyrightability. Most domestic scholars, on the other hand, believe that AI-generated works are copyrightable, and their main views include the "Objectivist Criteria" and the "Tool". The scholars holding the "objectivist standard" believe that the copyrightability of AI creations should not be denied because the authors of the creations are not natural persons, and the objectivist standard should be adhered to in judging whether the contents of AI creations can constitute a work, and there is no need to take into account the process of creation of a work, and the results of the creations should only be formally judged as to whether they are "creative". The "tool theory" should be endorsed. Scholars who recognise the "tool theory" acknowledge the premise that "works should be the results of creation by natural persons", and believe that AI essentially belongs to the creative tools of human beings, and its generated content is an extension of the expression of the thoughts of human authors, so as long as it meets the criteria of originality and expression in certain forms as required by copyright law, it should be considered a work of art. [4]As long as it meets the criteria of originality and being able to express itself in a certain form as required by the copyright law, it should be considered as a work. Some scholars think that the discussion on legal doctrine has

no substantive significance, and instead, from the perspective of incentive theory and market competition, they point out that recognising the copyrightability of AI creations is more effective in realising the interests of policy choices. Some scholars have further pointed out that AI is no longer a simple programme, but has become an "intelligent system with the capacity for self-learning and rational action", and that "AI can be treated as a new civil subject in the future". On the basis of recognising the copyrightability of AI generators, academics have discussed the attribution of their rights, arguing that they can be regarded as "deductive works of design copyright", "legal person's works of AI designers" or "works of civil law". "fruits in civil law".

3. Non-copyrightability of Chat GPT-generated content under the Current Copyright Law System

Chat GPT can be applied in several scenarios, but not all of its generated contents cause copyright disputes. By analysing the content generated by Chat GPT, the content generated by Chat GPT in natural language generation application scenarios is the main part of the content involved in copyright disputes. The assertion that Chat GPT-generated content is copyrightable under the objectivist standard is not substantively established, as the current copyright law only protects human achievements; the view that generative AI is regarded as a tool for human creativity confuses the concepts of "AI-generated" and "AI-assisted generation", therefore, the concepts of "AI-generated" and "AI-assisted generation" have been confused. "The Chat GPT generation is therefore not copyrightable.

3.1. Development of Chat GPT Technology and Types of Content Generated

Chat GPT is a chatbot that learns and understands human language and communicates with users according to the context; however, it is essentially a specific application of AI generative technology, capable of automatically generating specific content based on in-depth learning of human language and related domain knowledge. Chat GPT's key technological foundation is a large-scale language model developed on the basis of a series of models called Generative Pre-trained Transformer (GPT), which has been iterated for four years. The key technology of Chat GPT is a large-scale language model, which is developed based on the Generative Pre-trained Transformer (GPT) series of models, and it has been iterated over four years, and the version of the model at the time of its initial release was GPT-3.5. One of the important reasons why Chat GPT has gained a lot of attention is that it has introduced a new technology, RLHF (Reinforcement Learning with Human Feedback), which is a new technology that can automatically generate specific content based on deep learning of human language and related domain knowledge. [5] Learning with Human Feedback). This technology aligns the content generated by AI models with human common sense, perception, needs and values by manually annotating the fine-tuned models, training the rewarded models and optimising the models to generate responses based on proximal policies. Therefore, Chat GPT, compared with similar products in the previous period, the proportion of fabricated facts has decreased substantially, and the toxicity of generated content has been greatly reduced, and it still belongs to generative AI in nature. Currently, the language model of Chat GPT has evolved from GPT-3.5 to GPT-4, and more forms of content may be generated in the future.

Currently, Chat GPT can be applied to intelligent Q&A, language conversion and natural language generation application scenarios, according to which its generated content can be broadly classified into three types. In the intelligent Q&A application scenario, Chat GPT can answer common sense or scene-based questions raised by users, such as knowledge quizzes, role simulations, etc.; this scenario can also play a role as a retrieval method, in which the content generated by Chat GPT is mainly factual answers based on existing knowledge or communication content in the context of specific scenes. In the language conversion application

scenario, Chat GPT can realise the conversion between different languages, in addition to the mutual translation between natural languages, it can also realise the mutual conversion between programming languages, between programming languages and natural languages, and between natural languages and special code symbols; in this scenario, the content generated by Chat GPT mainly includes the translation of the text, the code for reading and writing, and so on. The natural language generation scenario is the core application scenario of Chat GPT's content "creation", in which Chat GPT can perform three types of tasks: one is the overview from complexity to simplicity, i.e., assisting in the rapid refinement of the main points from complex information, such as the generation of keywords, etc.; the second is the expansion from less to more, such as writing short stories based on the prompts given by the user. [6]The second is to expand from less to more, such as writing a short story based on the prompts given by the user; the third is to create from something to nothing, i.e. to assist the user in generating the content required by the dialogue scenario, e.g. writing an outline of a dissertation or a literature review. It can be seen that not all the content generated by Chat GPT is subject to copyright disputes, but mainly the content generated in natural language generation application scenarios, which is the focus of the following discussion in this paper.

3.2. The Objectivist Perspective on the Establishment of the Representation of Copyrightability of Chat GPT Generated Content

Article 3 of the Copyright Law of the People's Republic of China (hereinafter referred to as the "Copyright Law") stipulates that a work "refers to an intellectual achievement in the fields of literature, art and science that is original and can be expressed in a certain form". In determining whether the content generated by Chat GPT constitutes a work, there is no way to get around the interpretation of the "originality" criterion. As a common practice, the criterion of "originality" has not yet been defined or clearly interpreted at the legislative level in various countries, and judicial practice is also divided. As a representative of the civil law system, France's traditional view that "originality" is a reflection of the author's personality, from the author in the creative process of creative choices, in the landmark case of Pachot, the judge will be "originality" is defined as In the landmark Pachot case, the judge defined "originality" as "intellectual input", which is not protected if it is automatic or logically imposed. The British courts have interpreted the originality standard in a number of cases, ultimately establishing two basic principles: first, that the work is not a copy of another's work, and second, that the work must have been made with "personal skill, labour or judgement". The U.S. Copyright Act of 1909 set forth the requirement of originality of a work, and its judicial practice in the early days adopted the principle of "sweat of the brow", which is similar to the traditional standard in the United Kingdom, but after the Feist case required that "originality" include However, after the Feist case, "originality" is required to include both "independent creation" and "a small amount of creativity". [7]From this, it can be seen that the understanding and discussion of "originality" has centred on the meaning of the concepts of "independent creation" and "creativity".

Some scholars are of the view that the inclusion of "independent creation" in the scope of "originality" is the result of a combination of factors such as the historical background of the copyright law, the need to solve the operational problems of the copyright system, and the contingent factors of judicial practice; The term "independent creation" describes the relationship between a work and its creator, rather than the attribute that essentially distinguishes a work from other things, and the term "creativity" is a provision of the copyright law based on a certain value objective and the statutory attributes and requirements for works. "Independent creation" involves the attribution of copyright in a work, and the judgement of whether the creation can constitute a "work" should first adhere to an objectivist judgement standard, i.e., to judge whether the creation can satisfy the sufficient "creativity" required by

the copyright law in terms of the form of expression. "Creativity". Under the objectivist criterion of originality, there is no need to consider the creator and the process of creation of the content generated by Chat GPT, but only whether the generated content meets the minimum level of creativity and whether it can be expressed in a certain form. In terms of the first question, nowadays, Chat GPT can write poems, complete short stories with certain user prompts, and several academic papers have listed Chat GPT as a collaborator; thus, it is evident that Chat GPT-generated content is formally close to human works, and in the absence of explicitly labelling the source of the content, its generated content is already superficially different from human works. It is more difficult to distinguish between the generated content and human works at the superficial level without clearly indicating the source of the content. Therefore, Chat GPT-generated content can be recognised as being able to meet the minimum level of creativity. As for the question of "whether it can be expressed in a certain form", there is no doubt that the content of the text generated by Chat GPT can be fixed in a certain form. Thus, under the Objectivist standard, Chat GPT-generated content is copyrightable on a superficial level.

3.3. Substantial Failure to Establish the Copyrightability of Chat GPT Content from the Perspective of a Civil Law Subject of Rights - a Person

"The generalisation of civil law on factual acts often takes the objective consequences caused by the act as the final constituent element", the act of creation, as a factual act, the content of the creation is the result of the act, and the content of the creation itself is used to judge whether or not the generated content has originality to a certain extent with a certain degree of reasonableness. However, the validity of the theory and method should be built on the basis of the applicable conditions and specific context. The premise of the application of the objectivist judgement standard is that the generated content originates from human beings, and it is only feasible to discuss the originality based on the creation result for the creation result of natural persons. Copyright law protects the expression of ideas, but "expression" does not only exist as a "combination of symbols", but also contains the meaning of human subjects. In the past, when the impact of new technologies on copyright law was mainly reflected in the reproduction and dissemination of works, the obvious advantage of the objectivist criterion in the application of the law was the prospect of a simple method of determining whether a certain result was original or not. Nowadays, generative AI, such as Chat GPT, intervenes directly in the process of creative production of content - intuitively, one is no longer using computers to produce works in new ways, but rather letting computers produce works in new ways. Therefore, it would be inappropriate to apply objectivist criteria directly to judge whether non-natural human-generated content can constitute a work, ignoring the premise that a work is a human expression. [8]If only the distinguishability of the work itself is taken into account, without considering the subject and process of creation, not only machine-generated content is copyrightable, but also "symbol combinations" generated by animals or even in nature may constitute works, which will result in an unjustified expansion of the scope of the object of copyright, and destabilise the balance between private and public interests, and even the stability of the legal system of copyright. This will lead to an unjustified expansion of the scope of copyright objects, shaking the balance between private and public interests and even the stability of the copyright legal system.

Following the history of the emergence and development of copyright, it can be seen that human creativity has always been the centre of gravity of copyright, both as a natural right and as an economic incentive from a utilitarian perspective. Established jurisprudence has also anchored the ship of copyright in the waters of human creativity. More than a century ago, in *Sarony v. Burrow-Giles Lithographic Company*, the U.S. Supreme Court defined "author" as "the person who owns the original thing", indicating that "author" is "the person who owns the original thing", and that "author" is "the person who owns the original thing". in which the

United States Supreme Court defined "author" as "the person who possesses the original thing", indicating that the "author" must be a "natural person". In subsequent cases such as *Mazer v. Stein* and *Goldstein v. California*, the Court has also repeatedly cited *Sarony v. Burrow-Giles Lithographing Co.* as indicating that human authorship is a prerequisite for a work to be protected by copyright law. In 2018, the U.S. Copyright Office denied a copyright application for AI-generated visual content, *A Recent Road to Paradise*, emphasising that copyright law protects the "original work of authorship" (or "authorship") of a creative work. authorship" (original work of authorship) that fixes its creative matter in a tangible medium; Congress left the definition of "author of an original work of authorship" deliberately blank in the legislation in order to "avoid altering the standard of originality established by the courts under the Copyright Act". "The concept of "author of an original work of authorship" is broad, but not unlimited (U.S. Copyright Review Board, 2022). Thus, under current copyright law, "human beings" are the subject of rights, and only human intellectual works may be copyrightable.

Based on the above discussion, the key to determining whether Chat GPT-generated content constitutes a work is to clarify the relationship between the content generated and a human being, i.e., whether a human being plays a key role in the production of Chat GPT content. On 16 March 2023, the United States Copyright Office published a statement in the Federal Register explaining the examination and registration of copyrights for works generated using AI technology. According to the statement, the basis for discussing the copyrightability of a creative work is "whether the author is a human being", i.e., whether the expression, selection, or arrangement of the literary, artistic, or musical elements of the work was conceived and performed by a natural person; for creations that contain some AI-generated content, the determination of copyrightability depends on the degree of creative control exercised by a human being. [9]The determination of copyrightability for creations containing partially AI-generated content depends on the extent to which humans had creative control over the expression of the work and whether they "authored" the traditional elements of authorship in the work. When applying for a work that contains AI-generated content, the author is obliged to label the AI-generated content and provide a brief description of the human author's contribution to the work; if the machine's work exceeds the maximum limits, the creation cannot be recognised as a work. The United States Copyright Office rejected the copyright application for "A Shortcut to Heaven" on the grounds that "there is no evidence of sufficient creative input or intervention by a human author in the image".

There are arguments for the copyrightability of AI-generated content based on the view of AI as a tool for humans to perform creation. However, is Chat GPT just a tool for human creativity in the content generation process? A tool is "an instrument used by a human being in the production process to process and manufacture a product" or "used to achieve a purpose. From the definition, as far as intellectual products are concerned, "human" is the one who takes the initiative in the production process, and tools only play a supplementary role. The essence of Chat GPT is the specific application of AI content generation technology, which means that, in general terms, people use existing human works to conduct large-scale training on AI technology and generate content using the laws obtained from the training. the laws obtained from the training to generate content. Specifically, the training process is based on a given sequence of text, the model will be the previous sequence of words as input, one by one to predict the distribution probability of the next word, thus learning the relationship between words, contextual semantics and grammatical rules, etc., and ultimately trained to understand human language. Links, Common Crawl series corpus and other datasets, etc., and the learning contents are all human intellectual achievements. During the system operation, the user gives text instructions (prompt), and then Chat GPT generates certain text results (answer) according to the instructions. Even if the Chat GPT receives human (user) prompts before generating textual content, the user does not have sufficient creative control over how the Chat GPT

understands the human prompts and actually generates the textual material; in other words, it is the machine, not the user, that actually selects and organises the output text. The user's instructions merely specify the subject matter of what the human wants the machine to output, but the machine actually determines how those instructions are realised in its output text. For example, if a user instructs Chat GPT to write a seven-line poem in the style of Li Bai on the subject of homesickness, he expects the system to generate poetry that is in the genre of a seven-line poem and deals with ideas and a style similar to Li Bai's, but Chat GPT determines the rhyming pattern of the generated content and the order of the phrases and structures in each line. The protection of works under copyright law does not extend to ideas in any case, and the user's idea of a "homesick poem in the style of Li Bai" is closer to the category of "ideas", the expression of which is actually generated by the machine.

As early as 1997, in *Urantia Foundation v. Maaherra* (United States Court of Appeals, Ninth Circuit, 1997), the Court described the "human intellectual element" contained in a work as "a human being". in a work of art. In that case, the plaintiff claimed to have authored a divinely inspired book entitled *Divine Revelation*, which the defendant Maaherra had copied and distributed on computer discs. The plaintiff argued that the book should be protected by copyright law as a work created by God, and that the defendant's actions constituted copyright infringement; the defendant argued that the book, as a "work of God," did not contain human creations, and therefore could not be protected by copyright law, and that its own actions did not constitute copyright infringement. The Ninth Circuit Court of the United States indicated in its opinion that the "author" is the first person who compiles, selects, coordinates, and arranges the text, and that the intellectual labour of human beings is reflected in the selection and arrangement of the content of the book, so that the book can be protected by the copyright law as a creative achievement of human beings. In the process of generating the content of Chat GPT, it is the machine that selects, arranges and expresses the text, and the content generated by Chat GPT should belong to the category of "automatically generated by artificial intelligence" rather than "completed with the assistance of artificial intelligence". [10] According to the Revised Issues Paper on Intellectual Property Policies and Artificial Intelligence issued by the World Intellectual Property Organisation (WIPO), "AI-generated" and "AI-autonomous" are terms used interchangeably to refer to outputs generated by AI without human intervention. The terms "AI-generated" and "AI-autonomously created" are used interchangeably to refer to outputs produced by AI without human intervention; "AI-generated" should be distinguished from "AI-assisted" outputs, which require significant human intervention or guidance. The view of AI as a human creative tool confuses these two concepts: Chat GPT-generated content is "AI-generated" output, not "AI-assisted" output, and is therefore not copyrightable.

4. Possible Infringement Risks of Content Production in Chat GPT Contexts

While generative AI has attracted academic debate around deeper issues of creativity, expression and the nature of works, it has also forced the rules of copyright law to face some pressing practical challenges. Although the content automatically generated by AI is not copyrightable, Chat GPT can still challenge existing copyright laws and even create infringement risks in the process of model training and generative exploitation, specifically in the areas of text generation, large-scale machine learning and generative exploitation.

4.1. Rethinking the "Protection of Form but not of Thought" in Copyright Law

If Chat GPT learns the laws of natural language from large-scale copyright-protected works during model training, the model is likely to produce content similar to the input data. When asked if it infringes on the copyright of others by learning from large-scale written works by human authors in the process of generating text, here is how Chat GPT responded, "No, because

Chat GPT just learns these texts, not copies the exact same text, and Chat GPT just uses probabilistic models based on known textual rules to generate new Chat GPT does not directly copy existing works in the text generation process, but rather recombines them. Some scholars have found through testing that one of the functions of Chat GPT is a "smart scrubber": even if the search turns up information that is directly related to the question, Chat GPT does not "copy and paste" the text directly, but rather performs synonym replacement, i.e., applies a different word than the original content. Chat GPT does not "copy and paste" the text, but instead uses synonyms, i.e., a combination of words that differs from the original content to express the same idea. The Copyright Act does not protect human opinions or ideas, but rather the expression of opinions or ideas, so Chat GPT's use of natural language in a non-copy-and-paste mode may circumvent the Copyright Act's protection of works.

4.2. Challenges to the Fair Use Principle in Large-Scale Machine Learning

One of the key reasons for Chat GPT's high text comprehension performance is the powerful "base" of a generative large-scale language model. This language model requires large-scale training based on terabytes of text and data in advance, from which it learns the implicit laws and patterns of human language. According to Open AI's published information, there are two main sources of Chat GPT training data: first, the basic pre-training data from GPT-1 to GPT-3, including Wikipedia, books, journals, news articles, Reddit links, the Common Crawl series of corpora, and other datasets; and second, the fine-tuning data manually annotated, which are used by thousands of labellers employed by Open AI. labellers (labellers) to provide training corpus for the model in the form of handwritten text . However, Open AI does not disclose the details of the source of training data related to Chat GPT, and it is still doubtful whether all these data are authorised. [11]The Wall Street Journal and other foreign news media have publicly accused Open AI of using their articles to train Chat GPT without paying any fees.The use of existing works in the machine learning process of Chat GPT mainly involves three phases: content input, deep learning and text output. The content input phase involves converting the format of the text to a standard data format, selecting useful information and adapting the information to a specific structure in order to build a library of content that falls within the scope of one learning session. Therefore, electronic reproduction of the text and data required for training is an unavoidable part of the process; if not all of the training data for Chat GPT is derived from authorised textual works or non-copyrighted public works, copy right infringement may be involved. The deep learning stage requires analysing and learning the laws of natural language from the data training set, and then using the acquired laws to process the data and optimise the model by manually sorting the generated responses and rewarding model evaluation, etc. This stage may involve temporary reproduction and imitation of established works, but temporary reproduction does not constitute copyright infringement in China. In the text output stage, if the content output by Chat GPT is substantially similar to the original work, it may infringe the reproduction right of the original work, etc.

In order to meet the needs of society for knowledge and information, copyright systems in all countries provide for limitations and exceptions to copyright, of which the fair use system is the most important form of copyright limitation. "Fair use" refers to the use of a work without the permission of, or payment to, the copyright holder under certain conditions. If the large-scale use of a work by a generative AI developer, such as Chat GPT, falls into the category of fair use, the behaviour does not constitute copyright infringement. If the use of works at different stages is consistent in terms of objectives and effects, it can be treated as a whole and given a unified characterisation, therefore, the large-scale machine learning behaviour of Chat GPT in this paper is no longer subdivided according to different stages. China's Copyright Law does not stipulate the general principle of determining whether a particular act belongs to the "limitations and exceptions" of copyright, but lists thirteen specific situations in Article 242 to

which the rule of fair use can be applied, and the current large-scale machine learning behaviour of AI cannot be covered. On the one hand, Chat GPT and other generative AI investors are mostly corporate organisations, and do not belong to the category of "for personal study, research or enjoyment"; on the other hand, the large-scale machine learning of generative AI is often used to pave the way for the subsequent profitability of the enterprise, and the commercial purpose cannot meet the requirements of "teaching in school classrooms or for appreciation". "On the other hand, large-scale machine learning of generative AI is often used to pave the way for subsequent profit-making by enterprises, and commercial purposes cannot meet the requirements of "teaching in a school classroom or scientific research. [12]At the comparative law level, Japan and the European Union have expanded the scope of fair use in their copyright legislation, adding "text data mining" in AI training as a new appropriate use case. However, whether the principle of fair use can be applied to the large-scale reproduction of works protected by copyright law by AI is not a novel topic, and there is a case law in the United States that has responded to this issue. Through *Perfect10 v. Amazon and Authors Guild v. Google*, U.S. courts have established two basic conditions for the application of the fair use principle to large-scale reproduction by artificial intelligence: first, the reproduction of the work by the machine will not be used to motivate others to produce new works, and second, the use of the work by the machine will not have an impact on the potential market and value of the copyrighted work being used. The second is that the use of the work by the machine does not affect the potential market and value of the copyrighted work. The emergence of generative AI such as Chat GPT challenges both of these premises. On the one hand, Chat GPT's large-scale copying and learning of works is intended to generate content, which is likely to be used in the creation of users' works; on the other hand, more efficient and less expensive machine-generated content may replace some works, with a potential impact on the copyright-related market. Regardless of whether or not the fair use rule can be applied to large-scale machine learning, this behaviour of generative AI poses a challenge to the fair use rule.

The most immediate negative impact of the non-application of fair use rules to the large-scale machine learning behaviour of generative AI is that AI developers need to invest a lot of money in obtaining licences for their works or risk being liable for large amounts of compensation. The size of the training content library largely determines the AI's learning ability and usage. Currently, the size of Chat GPT's training data has reached tens of terabytes, and there may be hundreds of thousands or even millions of copyrighted works included in it. If the use of each work needs to obtain a licence and pay compensation, it will undoubtedly greatly increase the economic burden of AI developers, and even trigger a "chilling effect" at the technical level. As one of the most important cutting-edge technologies of the 21st century, strict restrictions on the level of copyright law may hinder the progress of this key technology and even the overall development of society. Second, if large-scale machine learning behaviours cannot be included in fair use scenarios, AI developers may choose to use public domain works or limited works obtained by agreement to train algorithmic models due to considerations of high costs and potential legal risks. Language models trained on limited or even low-quality text and data are likely to develop "algorithmic bias", resulting in a generative AI that is unable to distinguish or even produce dangerous speech or advice, and the "toxicity" of replies is likely to increase significantly. The "toxicity" of the replies may be greatly increased, which is not conducive to the advancement of AI technology in the long run. In addition, the high cost of obtaining copyright licences will further widen the gap between AI research and development companies of different sizes, creating an unfair competitive environment or even an industry monopoly. Stronger enterprises are more likely to rely on resources to obtain more training data, optimise their models on this basis to provide more comprehensive and better quality services, attract more users to consolidate their market share, and form a virtuous cycle that will ultimately lead to a "winner-takes-all" situation in the industry.

However, the application of fair use rules to large-scale machine learning will challenge the copyright legal system itself. On the one hand, the fundamental value of the copyright system is to maintain the balance between individual interests and public interests, and the core of all its rules is like an intricately interwoven network of private and public interests; some scholars have compared the boundary between the two as an elusive "metaphysics", and technological changes always make the relationship between the two in a more unstable state. Some scholars have likened the boundaries between the two to an elusive "metaphysics", and technological change always makes the relationship between the two more unstable. The principle of fair use was originally established to balance the dual purposes of copyright law to protect the interests of authors and other copyright holders and to promote the wide dissemination of knowledge and information. The most intuitive consideration is that the impermissible use of other people's works will hinder the free expression and the exchange of ideas, and therefore the behaviour it is most concerned with is the use of works for non-profit-making purposes. For example, Article 24 of China's Copyright Law lists "for the purpose of news reporting", "for the purpose of classroom teaching and scientific research in schools", "for the purpose of display or preservation of editions in libraries, archives, memorials, museums, art galleries, cultural centres, etc.". for the purpose of display or preservation of editions". The establishment of the fair use principle is not intended to protect individual profits, but rather to promote the realisation of the interests of more people through the redistribution of economic benefits and the like. In the case of large-scale machine learning of generative artificial intelligence, a large number of users of the actual copyrighted works are enterprises developing artificial intelligence, and their use of the works is ultimately intended to attract more users and obtain more commercial profits, and the nature of the use is commercial rather than public. [13] The practical effect of safeguarding the economic interests of large-scale enterprises can be said to be a departure from the original purpose of the establishment of the principle of fair use, and the inclusion of large-scale machine learning behaviours in the scope of fair use will make the copyright law deviate from the track of balancing the interests of the public and the private to a certain extent. On the other hand, one of the important roles of intellectual property law is to promote knowledge innovation, but Chat GPT itself cannot create new knowledge, but rather "knowledge reorganisation" based on existing human knowledge reserves, and the novelty and authority of the content it generates deserves further consideration. The ability of Chat GPT to express learned human knowledge in a non-copy-and-paste manner also "encourages" speculative behaviours, and if this kind of "machine-fed" content, which is not novel, is widely used, it will not be conducive to the enhancement of human creativity in the long run. The fair use system is also designed to address the issue of how subsequent authors can utilise previous works in order to create new ones. If the end result of large-scale machine learning is not to generate novel content and promote human intellectual innovation, the inclusion of this behaviour in the scope of fair use would defeat the purpose of intellectual property protection.

4.3. Ethical and Copyright Risks that may Arise from the Exploitation of Generated Content

If Chat GPT learns the laws of natural language from large-scale copyrighted works during model training, the model is likely to produce content similar to the input data. Scriptwashing using Chat GPT may be morally hazardous; if a user uses Chat GPT-generated content in a creative work that constitutes a substantial similarity to the copyrighted work in the training data, the user may be infringing on the copyright of the original work. Since 2022, a number of students have been using Chat GPT instead of writing and coding their own papers; a number of journals have already stated that the use of generative AI, such as Chat GPT, to write academic papers is either completely banned or severely restricted.

Some scholars have also mentioned the indistinguishability of AI-generated content from works when discussing the legal attributes of AI-generated content, but the solution to this reality is now more urgent. The latest Public Guidance Act issued by the U.S. Copyright Office clearly requires that authors applying for copyright for visual and textual works should indicate which parts are done by AI and which parts are done by human beings; if the AI-generated part exceeds the maximum limit, it should not be put in the work for copyright application. AI-generated content is not a work protected by copyright law, and it is necessary for authors who use AI-generated content as part of their own work to indicate the source. While new technologies create new tensions for copyright law and pose new problems for legislators, they also present new opportunities. In the late 20th century, digital compression technology enabled digital music files to be shared conveniently on the Internet. In order to safeguard the copyright of musical works, the major U.S. record labels took a joint initiative with the Internet, computer, and other companies to devise a standardised technology (SDMI) for watermarking of performed musical works with a watermark to prevent others from making unauthorised use of digitally recorded musical works. Relevant regulations and historical copyright experience provide some reference points for determining the copyrightability of human creations containing AI-generated content. The law and the market will also drive the creation and application of AI-generated content identification technologies, and in the future, consideration may be given to labelling AI-generated content through the addition of electronic watermarks and other technological means. In copyright disputes involving human creations containing AI-generated content, the court may first identify the "AI-generated" content when determining whether the work meets the standard of "originality". If all or nearly all of the content was generated directly by the AI, the content is not copyrightable; if the content was created by both the AI and a human being, and if the human being made the majority of the choices and arrangements, the content may be recognised as a work.

5. Concluding remarks

In summary, Chat GPT-generated content is not copyrightable, its generation does not satisfy the basic premise that the author of a work is a natural person, and the view that it is a tool for human creativity confuses "AI-generated" with "AI-assisted" content. "AI-assisted generation". However, the widespread use of Chat GPT also brings some practical challenges to the current copyright law.^[14] Text data mining in the process of machine learning may constitute copyright infringement, but whether it can apply the rule of fair use needs to be further considered in detail; Chat GPT-generated content may infringe on the copyright of the original work when utilised, and in the future, the content created by AI can be identified through electronic watermarking and other technologies, and the originality can be judged according to the criterion of whether human beings have made the main choices and arrangements for the work.

6. Conclusion

The legal system has a lagging effect, but thinking about legal issues should be forward-looking. In recent years, the rapid development of big data and artificial intelligence technology has triggered discussions on whether artificial intelligence can serve as a suitable copyright subject. Currently, the simulation of human behaviour by generative AI such as Chat GPT is limited to content creation, and there is still a big gap with strong AI. If the future development of AI technology reaches the stage of autonomous consciousness, it is also necessary for the civil law to respond to the legal status of AI in the subject system, rather than taking the lead in making breakthrough changes in the field of copyright law. In addition, the realisation of rights and the fulfilment of obligations often go hand in hand, and the protection of AI-generated content

should not be rushed because of the apparent similarity between AI-generated content and human works, and the assumption of responsibility for infringement should be considered. If the copyright law provides protection for AI-generated content, when the generated content infringes on the rights and interests of others, the owner of the rights should also be liable for the infringement of its "creation". If the owner, developer or user of a generative AI has an interest in the machine-generated content, they will be liable if the content infringes on the rights of others to reproduce or adapt it, or if it constitutes a defamatory infringement of the personality rights of others. Considering the actual process of generating content such as Chat GPT, it is not sufficiently plausible that a natural or legal person who is not directly involved in the creation of the content would be liable for the infringement of machine-generated objects. Since the enactment of the Anna Act in 1709, technological changes have always brought new tensions to the copyright legal system. In the future, when the copyright law responds to the new problems brought about by AI-generating technology, it should not only consider the degree of change of the technology itself, but also consider the degree of its impact on the balance between public and private interests, and whether or not to change the existing legal rules, it is necessary to explore the conditions of its application in the context of new technology on the basis of the basic principles of copyright law.

References

- [1] Flora G.: Daily briefing: The pros and cons of writing a paper with Chat GPT(Nature),2023.
- [2] Antonio V, Fabiana R.: Chat GPT use in teaching social psychiatry(The International Journal of Social Psychiatry),2023.
- [3] Wang Qian:Revisiting the characterisation of AI-generated content in copyright law(Political and Legal Forum,China 2023),No.41(04),p.16-33.
- [4] Cai Zhenjiang: Che Yulu.Exploring the Legal Personality Determination and Nature of Generated Content of Chat GPT(Journal of Xinjiang Normal University (Philosophy and Social Science) Edition,China 2023),p.1-8.
- [5] Chen Chen: On the Copyrightability of Artificial Intelligence Generated Content - An Example of Chat GPT Generated Content(Technology and Publishing , China 2023),p.1-9.
- [6] Teng Haibo: Reflections on the Regulation of Generative Artificial Intelligence Represented by ChatGPT(The Internet World),Vol.06(2023)p.38-44.
- [7] Li Yi, Wang Meng:Research on the Dilemma of Intellectual Property Protection and Countermeasures under the Vision of ChatGPT(Journal of Chongqing University of Technology (Social Sciences),China 2023),p.1-10.
- [8] Sam S. Success Through Simplicity: What Other Artificial Intelligence Applications in Medicine Should Learn from History and Chat GPT(Annals of Biomedical Engineering),2023.
- [9] Wang Zongshui, Gu Yingxiao, Zhao Hong: Innovation of Content Generation and Knowledge Dissemination Model Based on Chat GPT(Library),Vol.06(2023)p.42-48.
- [10]Wang Yang: Is Chat GPT-generated content copyrighted?(Shanghai Science & Technology News), Vol.06(2023).
- [11]Deng Shasha, Li Zhenyu, Pan Yu: Chat GPT and AI-generated content: should scientific research adopt or resist?(Shanghai Management Science), Vol45(2023) No.2, p.15-20.
- [12]Wang Qian: Is the content generated by Chat GPT protected by copyright law?(Exploration and Controversy), Vol.03(2023), p.17-20.
- [13]Zhao Feng: Challenges and Responses of Copyright System under the Exploitation of Intelligently Generated Content - A Case Study of Chat GPT (Publication and Distribution Research), Vol.03 (2023), p.48-56+47.
- [14]Li Weihua : Who does Chat GPT-generated content belong to and what are the legal risks?(Shanghai Enterprises), Vol.03(2023), p.84-85.