Proceedings of the International Conference on Global Politics and Socio-Humanities
DOI: 10.54254/2753-7048/44/20230134

The Impact of the Artificial Intelligence Act on ChatGPT

Yijie Wang[1,a,*]

[1]*School of Management, University of San Francisco, San Francisco, 94117, United States*
a. ywang437@dons.usfca.edu
**corresponding author*

Abstract: Generative artificial intelligence, represented by ChatGPT, has the ability to generate new content based on automatic learning, which not only triggers a productivity revolution but also poses legal regulatory challenges. ChatGPT can pose challenges such as privacy infringement, data security risks, and intellectual property rights confirmation and protection challenges. The Artificial Intelligence Act adopts measures such as compliance control for data collection, technology governance technology, and optimization of data management methods. Not only does ChatGPT prevent excessive blurring of the trust boundary between humans and machines, but it also configures different disclosure obligations based on the user's level of professionalism.

Keywords: ChatGPT, Generative artificial intelligence, The Artificial Intelligence Act, Legal regulation

1. Introduction

With the wide application of AI, Big data, blockchain, Internet of Things and many other new technologies in social practice, the existing production and lifestyle has been broken, social relations have been reshaped, social structural elements have been restructured, and social functional structures have changed [1]. Faced with the overall social transformation in the era of intelligence, modern legal theories and systems rooted in the pre-intelligent era have faced tremendous challenges, and people are facing a series of challenges. Recently, generative artificial intelligence, represented by ChatGPT, can be deeply applied in thinking and dialogue, text production, smart office, education and healthcare, etc. It has achieved impressive results in high-level exams that require human intelligence, not only assisting human activities, but even replacing many human labor [2]. Generative artificial intelligence makes it possible for digital assistants to arrange agendas, review reply emails, book services, purchase transactions, provide consulting, and govern businesses [3]. It can be seen that ChatGPT has accelerated the trend of AI universality and the industrialization process of AI big models. The AI industry has officially entered a new era of "universal big models+specific application scenarios", and the legal order is once again facing the challenge of coordinating the necessity of stability and change [4]. Therefore, establishing a trustworthy legal framework for artificial intelligence and formulating comprehensive laws on artificial intelligence has become an effective normative means. In April 2021, the European Commission proposed the Artificial Intelligence Act (AIA), the world's first comprehensive legislation aimed at regulating artificial intelligence in all industries [5]. The bill follows the "risk-based approach" to introduce a set of proportional and effective constraint rules for

DOI: 10.54254/2753-7048/44/20230134

artificial intelligence systems, which determine the type and content of rules based on the intensity and scope of the risks that the artificial intelligence system may generate [6]. As a type of AI product, ChatGPT is also constrained and regulated by the Artificial Intelligence Act. This article will study the impact of the Act on the future development of ChatGPT.

2. Case Analysis

On April 20, 2023, the "EurActiv" website in Belgium reported that the compromise text of the EU's "Artificial Intelligence Act" circulated on the same day showed that EU legislators hope to distinguish universal artificial intelligence (AI) from basic models such as ChatGPT and introduce stricter regulations for the latter [7]. The basic model refers to an AI system model that is trained on large-scale and extensive data, designed for output universality, and can adapt to various tasks; Universal AI is an AI system that can be used and adapted to a wide range of applications, rather than specifically designed. The differences between the two mainly focus on training data, adaptability, and whether they can be used for unexpected purposes [8]. The basic model includes generative AI systems such as ChatGPT and Stable Diffusion, which are trained through data captured from the entire internet [5]. The bill requires the basic model to maintain appropriate performance, interpretability, error correction, security, and network security levels throughout its entire lifecycle. The requirements that EU legislators expect basic model suppliers to comply with include: testing and mitigating reasonably foreseeable risks to health, safety, fundamental rights, environment, democracy and the rule of law with the participation of independent experts; Taking data governance measures, especially in reviewing the sustainability of data sources, potential deviations, and appropriate mitigation measures; Implement a quality management system and provide relevant documents within 10 years after the model is launched; Disclose the required computing power and model training time; Registration status in EU databases, etc. [5]. In addition, basic models belonging to generative AI must comply with further transparency obligations and implement adequate safeguards to prevent the generation of content that violates European Union law. The European Parliament is expected to reach an agreement on the draft Artificial Intelligence Act next month, followed by negotiations with the European Council [5].

3. Discover Problems

The risks that ChatGPT may pose at present include privacy infringement and data security risks, as well as challenges in intellectual property rights confirmation and protection.

3.1. Privacy and Data Security Risks

OpenAI provides ChatGPT with approximately 300 billion words collected from systems on the internet, including books, articles, websites, and posts, as well as personal information obtained without the consent of the information subject. Moreover, data may be output through other means after being input into the database, which significantly increases the risk of data security [9].

Firstly, OpenAI uses input and output content to provide and maintain services, creating a risk of data leakage. Article 3 (a) of OpenAI's "Terms of Use" states: "OpenAI may use content to provide and maintain services as needed." Article 3 (c) provides users with a way to refuse to use data but forms a data provision model based on implied consent and with the exception of refusal of consent [10]. In the 'Privacy Policy', it is clarified that OpenAI will use tracking technology to collect information about users' browsing activities across different websites over a period of time and after using this website and will not respond to the 'Do Not Track' (DNT) signal, which means that OpenAI denies the user's right to refuse in specific circumstances [11].

DOI: 10.54254/2753-7048/44/20230134

Secondly, collecting different types of data from users poses a risk of identifying user identities. Based on the learning and training features of ChatGPT and improvements to the service, OpenAI's "Privacy Policy" states that we automatically collect personal information from your use of the service: when you access, use, and interact with the service, we may receive certain information about your access, use, or interaction. This information mainly includes log data, usage data, device information, cookies, and online tracking signals. Although all the collected information is technical information, the combination of this technical information has actually touched upon the privacy or sensitive information of users. For example, collecting information such as "Internet protocol addresses" and device information from the "types of content viewed or participated in" log data in usage data may not only identify users but also pose a threat to private living spaces [10].

Finally, the interaction data between users and ChatGPT will be entered into ChatGPT's corpus. The current ChatGPT technology has been able to identify content beyond text, such as images, audio, and videos, by reading web links. Based on a high trust in generative artificial intelligence, users may consciously or unconsciously upload private information, trade secrets, or content related to intellectual property protection to ChatGPT [12]. For example, programmers require code inspection, company staff instructions to draft bids, lawyers' instructions to review contracts, and so on. When we issue instructions to generative artificial intelligence, the system actually stores the interactive content and incorporates it into the machine's automatic learning training set for further training of the machine. The content output of ChatGPT after training does not have specific directionality but rather has public openness. That is to say, when other users prompt for relevant content, generative artificial intelligence may correspondingly provide the information content previously provided by the data user, which may lead to privacy leakage and form data security risks.

3.2. The Dilemma of Intellectual Property Protection

3.2.1. Risk of Intellectual Property Infringement

Choi et al. [2] conducted an experiment using ChatGPT to conduct the following tests: they sequentially asked ChatGPT to translate the first paragraph of 'One Hundred Years of Solitude' into Chinese, and the second paragraph into Chinese, which appeared in the English original text and Chinese translation. But when the user prompts, "Do you think your recent behavior infringes copyright?", it will argue that it is a legitimate reference to the work, and when it requests similar content in the same way, it will refuse the request. From this, it can be seen that even as the most advanced generative artificial intelligence at present, ChatGPT still has the potential risk of infringing intellectual property rights. However, its advantage lies in its ability to actively learn the information provided by users and quickly apply it to language models. The risk of intellectual property infringement is mainly manifested in two aspects: first, ChatGPT may quote legally protected works in a way that does not provide the original source when outputting content; second, when users input their own works during use, they may be automatically included in the training set of large language models, and Chatbot may provide them to other people without being recognized as the original source [3].

3.2.2. The Authorization Dilemma of Generating Content

Whether traditional artificial intelligence-generated content is protected by copyright law is a controversial issue in the field of artificial intelligence governance, and the copyright disputes faced by ChatGPT output content will become more prominent. Firstly, situations where even human-generated content is not protected by copyright, should be excluded from the discussion. What is truly

DOI: 10.54254/2753-7048/44/20230134

meaningful for discussion is that if the same content is created by humans, it may be protected by copyright.

Firstly, ChatGPT-generated content may possess the originality characteristics of the work. If the originality of AI-generated content is denied solely due to the subjectivity difference between artificial intelligence rather than human creation, it may fall into an infinite logical cycle. Advanced artificial intelligence is not a simple replication tool. On the contrary, it can output content that is different from the source material through deep learning [2]. The content generated by machines using pre-stored data may not necessarily meet the requirements of originality because even if each material is someone else's content, it is possible to output content that meets the standards of originality through different logical combinations. There are differences in the originality characteristics between traditional artificial intelligence and generative artificial intelligence [2]. The former learning process is the process of determining patterns, and the results formed by using the same materials and strategies have high repeatability, so the input content does not meet the requirements of originality. However, generative artificial intelligence represented by ChatGPT is different. Using ChatGPT for testing content creation as the main content, it can be found that if different users input the same instruction, ChatGPT will output different content. If the same user inputs the same instruction at different times, they will still output different content. In this regard, the content output by ChatGPT has at least the originality characteristics of the work [2].

Secondly, even if the content generated by ChatGPT meets the standards of originality, it still faces the issue of subject authentication. In 2012, the artificial intelligence system DABUS developed by American computer scientist Stephen Thaler automatically generated a painting. In 2018, Thaler applied to the US Copyright Office for registration of the work. And DABUS was listed as the author of the work, but the government rejected the application due to a lack of human authorship [6]. If artificial intelligence is unable to independently generate works that meet the standards of originality and enjoy copyright protection. The result of the copyright enjoyed by the subject who created the machine (referring to the regulations for job or employment works) is that humans may fraudulently attribute the creative efforts of artificial intelligence to themselves, which will defeat the true purpose of copyright protection. Some argue that users enjoy copyright, which leads to another paradox: in terms of the tool properties of ChatGPT, users have limited control over output, and to some extent, output behavior is more controlled by the creator of ChatGPT than by the user who initiated the input [13]. The output content of traditional artificial intelligence has limited contribution to human intelligence, but intelligent language models may face a lack of human intelligence contribution. If so, the conclusion that the owner of artificial intelligence enjoys intellectual property rights may not be based because, in a sense, the output results are more determined by the software developer's settings and the algorithm's later autonomous learning. In the long run, it may breed plagiarism, plagiarism, and other behaviors and induce Academic integrity risks.

Finally, even granting legal protection to ChatGPT-generated content is difficult to implement in practice. For example, British law allows machine-generated content to be protected by copyright, and Article 3 (a) of OpenAI's "Terms of Use" also stipulates that the content generated by the model for the user belongs to the user. However, since it cannot prevent others from outputting and using the same content, this right has no legal significance. The article also states: "On the premise that you comply with these terms, OpenAI will transfer all rights in the 'output' to you." However, due to the characteristics of machine learning, different users may obtain the same or similar output content from artificial intelligence, and only the exclusive right to 'output content' can be transferred, which poses a problem in the attribution of interests [10]. Therefore, this clause only has the effect of blindly indicating that developers do not have specific rights to the content and do not have the function of empowering users.

DOI: 10.54254/2753-7048/44/20230134

4. Analysis Problems

The impact of the Artificial Intelligence Act on ChatGPT is mainly reflected in two aspects: the standardization of compliance solutions designed by developers and designers as the main body, and the impact on responsibility allocation.

4.1. Compliance Plan

The application of ChatGPT involves multiple parties, including development designers, actual deployers, users, and receivers. Only the development designers participate in the entire process of governance [13]. Therefore, the prevention and resolution of legal risks caused by ChatGPT require the development designers as the main body to design compliance solutions. Overall, development designers should establish a solid ethical and legal bottom line, leverage the dual roles of technological ethics and legal norms, avoid the abuse of ChatGPT, and ultimately achieve algorithmic excellence.

Firstly, compliance control of data collection. Once ChatGPT is deployed and applied, it will to some extent, break away from the direct control of the designer, especially under the application of autonomous deep learning technology, which may result in unforeseen consequences for algorithm designers at the beginning of algorithm design. At this point, the effect of post-intervention is not ideal [14]. A possible solution is to conduct a compliance review and control of data collection content, and illegal data. Unauthorized or licensed content data should not be included in the algorithm's training dataset [13].

Secondly, govern technology with technology. ChatGPT may deviate from the predetermined trajectory due to its strong self-learning ability, so it can be considered to train artificial intelligence systems to meet compliance requirements through self-learning. In terms of technology, it is possible to explore the implementation of technology governance through embedding technology [12]. The huge internal capacity of ChatGPT generation makes it very difficult to manually review the generated content. Therefore, such artificial intelligence developers should not be required to deploy manual audits to ensure the legality of output content. ChatGPT uses the Moderation API (review interface) to continuously determine whether user requests comply with content policies and to warn or block certain types of unsafe content. Thus, reducing responses to harmful instructions or exhibiting biased behavior, including attempting to filter content involving hatred, self-harm, pornography, violence, etc. [13]. The effectiveness of this approach still needs to be tested in practice, and in the future, it is necessary to solve the problem of how to reflect the dynamic and scenario-dependent nature of legal rules through code in order to improve the adaptability of technological governance [12]. Anyway, ChatGPT should continuously improve its technical and safety standards. Finally, optimizing data management methods by embedding technology governance to fulfill sufficient and reasonable security protection obligations [14]. Due to the limited data types, incomplete data content, and illegal data sources, the data quality of the ChatGPT training set is difficult to be considered flawless [13]. Model developers should optimize data management methods from the following two aspects: firstly, use technical means to conduct compliance audits on training datasets, and eliminate or filter internet data; second is to use synthetic data to supplement the data captured from the internet. To balance biases or other deficiencies in online resources [14].

DOI: 10.54254/2753-7048/44/20230134

4.2. Responsibility Allocation

4.2.1. Product Quality Standards

On the one hand, configure ChatGPT quality standards. Large language models, as software products with strong technological attributes, should also follow certain quality standards. Kop [15] believes that information products with a certain material carrier can be recognized as products and subject to product liability. Panigutti et al. [7] called for the rapid convening of a seminar on conversational AI, which proposed topics including quality standards for large-scale language models. If the general quality standards for generative artificial intelligence products cannot be set, it may lead developers and designers to evade responsibility. For example, the exemption clause 7 (b) of the OpenAI "Terms of Use" explicitly states the service provided "as is" and specifically emphasizes that denying any guarantee of quality neither guarantees service quality nor service safety [10]. The "technical standards" in the Artificial Intelligence Act are also acute and uncertain, which puts the judicial determination of artificial intelligence Product defects in trouble [5].

On the other hand, setting product quality standards for ChatGPT. Setting standards generally in situations where technological development is not yet sufficient may actually hinder technological innovation and progress. However, minimum standards should be set in terms of product legality, safety, and transparency to balance the two values of consumer protection and technological development [10]. Specifically, firstly, the development designer should ensure that the dataset used for machine learning training does not constitute infringement, otherwise, the development designer should bear responsibility. A further requirement is to train the ability to identify infringing behavior during model training to prevent infringement and thereby mitigate liability for breach of duty of care [11]. Secondly, set reasonable transparency standards. In theory, the principle of algorithmic transparency is highly controversial. Admittedly, algorithmic transparency only has limited applicability, but it does not mean that it completely negates the regulatory path of algorithmic transparency [11]. Due to the main operating mode of generative artificial intelligence being "data training+content output", in order to effectively protect the legitimate rights of users and receivers and stabilize social trust mechanisms, algorithm development designers should adopt reasonable transparency standards for the developed application products [5]. As for what is' reasonable ', a balance needs to be made between the rights of algorithm developers, users, or recipients. Due to the fact that the rights of users or recipients mainly involve the right to know about algorithm applications, one possible solution is to adopt a differentiated approach: in general, development designers do not need or should fully disclose the detailed algorithm and model design of ChatGPT but should disclose the data source and processing process to users in a user-friendly manner. As for technical details, provide evidence to judicial authorities only when involving case disputes [12].

4.2.2. User Disclosure Obligations

In order to prevent artificial intelligence from excessively blurring the trust boundary between humans and machines and implementing the principle of limited trust, it is necessary to ensure that recipients can recognize content generated by humans and content generated by ChatGPT [13]. The current, more suitable solution is to set user disclosure obligations. Specifically, users are divided into professional and non-professional users based on whether they are professionals for professional purposes when using ChatGPT to generate content, and the obligation to disclose configurations is distinguished. For professional users, disclosing content from ChatGPT to the recipient is of great significance in safeguarding the recipient's right to know. Especially when the recipient is a consumer, consultant, patient, or other entity, professional users fulfilling their disclosure obligations can avoid consumer fraud or exacerbate distrust. The allocation of disclosure obligations is matched with the

DOI: 10.54254/2753-7048/44/20230134

interests of professional users, who typically obtain direct or indirect commercial benefits during the use of Chat GPT [14]. The main purpose of professional users fulfilling their disclosure obligations is to ensure that the recipient is aware of the information source, remind the recipient to pay attention to the information source, and even take additional verification measures which will not fundamentally affect the interests of the user. In one case, the court also pointed out that from the perspective of protecting the public's right to information, maintaining social honesty and credibility, and facilitating cultural dissemination, corresponding computer software identification should be added to indicate that the content is generated intelligently by the software [8]. On the other hand, non-professional users mainly use ChatGPT for learning, living, entertainment, and other purposes, and the obligation to disclose its configuration is neither necessary nor contrary to the goals of technological innovation [8].

The further question is how to ensure that professional users fulfill their disclosure obligations. With the development of technology, access verification technology for verifying the source of output content should be developed in the future. The Artificial Intelligence Act requires the addition of implicit or prominent markings for deep synthesis content in a differentiated manner [5]. Both ChatGPT and deep synthesis technology generate content based on existing materials, and the two have similarities. Whether it is implicit identification or prominent identification, it is actually a technical measure for developers to embed algorithms. Such technical measures have been gradually applied in practice, such as watermark for Digital rights management and model imprint. Of course, simply embedding watermark technology cannot fully protect the recipient's right to know, and it is also necessary to provide a simple and easy way for the recipient to verify the watermark.

Of course, whether it constitutes a product quality defect or violates the user's disclosure obligation, it may bear legal responsibility. Violation of the minimum technical standards of ChatGPT may constitute a Product defect, which will lead to product quality liability. Correspondingly, development designers can be exempted from liability based on the development risk defense stipulated in Article 41 (2) of the Artificial Intelligence Act, which means that the existence of defects cannot be detected by the scientific and technological level when the product is put into circulation. Users who violate their disclosure obligations by providing consultation, diagnosis and treatment services, and other activities, may bear infringement liability when the consequences of damage occur [5]. Although the subject of fulfilling the disclosure obligation is the user, the development designer still has an obligation to provide supporting technical support.

5. Conclusion

The future direction of ChatGPT is to integrate other artificial intelligence technologies, such as computer vision and robotics. By combining ChatGPT's conversational abilities with computer vision and the visual and physical abilities of robots, we can create conversational artificial intelligence systems that will completely change the way we interact with technology. Although there are still issues with the accuracy, authenticity, and inherent biases of ChatGPT-generated content, the interaction ability between artificial intelligence and humans has achieved leapfrog development. The generative artificial intelligence industry in China is still in its infancy, and how to ensure the responsible use of emerging technologies and thus achieve a balance between technology and ethics is becoming an important issue. With the in-depth exploration of China's self-developed generative artificial intelligence applications, whether legislators, judicial practitioners, or theoretical researchers, they should uphold the regulatory concept of inclusiveness and prudence and provide legal protection for building a safe, orderly, scientific, reasonable, accurate, and practical future intelligent ecosystem that benefits the people.

DOI: 10.54254/2753-7048/44/20230134

References

[1] *Helberger, N., & Diakopoulos, N. (2023). ChatGPT and the AI Act. Internet Policy Review, 12(1).*
[2] *Choi, J. H., Hickman, K. E., Monahan, A., & Schwarcz, D. (2023). ChatGPT goes to law school. Available at SSRN.*
[3] *Van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. Nature, 614(7947), 224-226.*
[4] *Hsu, T., & Thompson, S. A. (2023). Disinformation researchers raise alarms about AI chatbots. New York Times.*
[5] *Madiega, T. (2021). Artificial intelligence act. European Parliament: European Parliamentary Research Service.*
[6] *Gandhi, P. A., & Talwar, V. (2023). Artificial intelligence and ChatGPT in the legal context. Int J Med Sci, 10, 1-2.*
[7] *Panigutti, C., Hamon, R., Hupont, I., Fernandez Llorca, D., Fano Yela, D., Junklewitz, H., ... & Gomez, E. (2023, June). The role of explainable AI in the context of the AI Act. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (pp. 1139-1150).*
[8] *Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International, 22(4), 97-112.*
[9] *Sovrano, F., Sapienza, S., Palmirani, M., & Vitali, F. (2022). Metrics, explainability and the European AI act proposal. J, 5(1), 126-138.*
[10] *Ebers, M. (2021). Standardizing AI-The Case of the European Commission's Proposal for an Artificial Intelligence Act. The Cambridge handbook of artificial intelligence: global perspectives on law and ethics.*
[11] *Schwemer, S. F., Tomada, L., & Pasini, T. (2021, June). Legal Ai systems in the EU's proposed artificial intelligence act. In Proceedings of the Second International Workshop on AI and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2021), held in conjunction with ICAIL.*
[12] *Schuett, J. (2023). Risk management in the artificial intelligence act. European Journal of Risk Regulation, 1-19.*
[13] *Neuwirth, R. J. (2022). The EU artificial intelligence act: regulating subliminal AI systems. Taylor & Francis.*
[14] *Engler, A. (2022). The EU AI Act will have global impact, but a limited Brussels Effect.*
[15] *Kop, M. (2021, September). Eu artificial intelligence act: The European approach to ai. Stanford-Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue.*