

Analysis and Recommendations on Data Involved in Legal Artificial Intelligence

Xiaofei Wang

School of Law, China Jiliang University, Huangzhou, Zhejiang, China

Abstract: This paper discusses the data problems arising from the application of artificial intelligence to legal counseling, with the purpose of revealing how to effectively use the data and avoid data risks during the operation of artificial intelligence. In order to achieve this goal, the first part of this paper analyzes the operation mechanism of artificial intelligence and the status quo of applying it to the legal field. The second part of the paper starts from the link of its operation, describes the specific data problems, such as the informed consent of the data subject, the safety of the data training, etc., and carries out a legal analysis. After this analysis, the last part of the paper gives relevant suggestions from the aspects of legislation, ethics, data grading, and risk prediction. Relevant suggestions are given.

Keywords: AI; Legal Advice; Data Security; Risk; Information Interaction

1. Introduction

At present, China has entered a new development stage of comprehensively building a modern country and promoting high-quality development, and the people's needs are growing in terms of democracy, the rule of law, fairness, justice, security and the environment. Providing the people with adequate, high-quality and convenient public legal services to continuously meet the people's needs for a better life is the proper meaning of accelerating the construction of a modern public legal service system.

One of the most important parts of the work of legal professionals is the provision of reliable, high-quality legal advisory services to lay people. However, due to the insufficient number of legal professionals, it is often difficult to ensure that lay people have access to sufficiently high-quality counseling services. Lawyers, as one of the most direct ways for

people to seek legal help, are not in reality able to solve all legal problems practically and effectively. The main reason for this is that, in addition to the economic factors that the public needs to consider, limited human resources are also one of the reasons that restrict the public's ability to defend their rights. To alleviate the judicial pressure to cope with the above problems, convenient and low-cost legal counseling network platforms have sprung up. For example, "China Legal Service Network", "Easy Law", "Fast Law", "Find Law", "Legal Express" and so on[1]. However, these platforms are still mainly relying on licensed lawyers for answers to legal knowledge, and with less than 700,000 lawyers in China, the rate of online provision of legal answers is always low. Through the domestic mainstream legal advice website research can be seen, the website daily average of more than 2,800 legal counseling questions, the total number of questions for 12.65 million, has solved the problem but only 370,000, the answer rate is less than 3%[2]. Even if Internet technology is utilized to break down offline communication barriers, relying solely on human beings to provide legal advice is not a proper response to the public's need to seek legal help.

The artificial intelligence Q&A model that has emerged in recent years has a natural advantage of low cost and high efficiency in the highly informatized and data-driven legal field. In 2019, Baidu launched "Du Xiaofa", for example, based on massive case data, laws and regulations and judicial practice, relying on the core technology of Baidu Knowledge Graph, can not only provide users with fast and accurate search of laws and regulations, but also do intelligent case adjudication and legal assessment. It can be seen that artificial intelligence, relying on data resources and using the network platform as a medium, automates the processing of information and provides legal counseling, which is of practical significance in alleviating legal needs and

helping the public to protect their rights. However, in the AI operation mechanism, in order to ensure that the AI can be comprehensive, accurate and professional in responding to legal issues, a large amount of data collection, storage, transmission and processing and continuous training and optimization are required, while legal data often contain sensitive personal information and confidential business information, which, if leaked, tampered with, or misused, may cause serious losses and harm to the parties and the community. Therefore, how to ensure the security and legality of data, and prevent the leakage and abuse of data is a problem that needs to be focused on and solved by the application of AI in the field of legal services. Most of the current research on how to guarantee data security focuses on emphasizing the close regulation of data and industry self-regulation from a technical perspective, and there are few detailed regulations on data input and processing from a legal perspective, such as ensuring that the data source is correct and legal, which may involve the openness of data and copyright issues, in addition to the design of privacy for the details of the case when inputting the data, the provision of consultants with the In addition to this, the design of privacy in case details when entering data, the extent of information to be provided to the counselor, and the protection of instantly accessible dynamic data are all issues that need to be regulated by law.

Based on this, this paper takes the legal consulting service of artificial intelligence as the object, combines the relevant academic views and judicial precedents at home and abroad, explores the data security and standardization problems in the various stages of the operation of artificial intelligence from the legal perspective, and gives reasonable suggestions, so as to make it further popularized and applied under the legal regulation, and help solve the problem of difficult for grassroots to safeguard the rights of the people.

2. Operational Mechanisms and Applications of Artificial Intelligence Technology

2.1 Operational Mechanisms of Artificial

Intelligence Technology

Artificial Intelligence (AI) is a new digital technology that studies the use of computers to simulate certain human thought processes and intelligent behaviors, mainly in the form of automated question and answer sessions, aiming to provide natural and smooth answers to user-input questions. In recent years, with the innovation of algorithmic technology and the strengthening of computing power, artificial intelligence has been equipped with the ability of deep learning, and is penetrating into all walks of human society at an alarming rate. And legal AI mainly relies on natural language processing (NLP) technology to understand and process human language, including tasks such as text analysis, semantic understanding and generation[3]. Its operation mechanism can be understood as the developer or manager of artificial intelligence will collect the law-related data sets for specialized processing after the input to the artificial intelligence, the computer according to algorithms and pre-programmed procedures to start in-depth learning, the information input by the manager for integration and processing, to build a systematic and comprehensive legal knowledge system as their own knowledge reserves, simulate some of the ideas of human beings, and output the corresponding instructions to solve problems. The computer will simulate some human ideas and output the corresponding instructions to solve the problems.

2.2 Application of Artificial Intelligence in the Legal Field

Several artificial intelligences applied to the legal field have already appeared and achieved a good response. For example, Joshua Browder designed the DoNotPay software, which can provide users with templates for various business letters, such as appeals for violation tickets, insurance claims, complaint letters, etc. Users only need to provide basic information as well as legally related situations, and DoNotPay will automatically draft the corresponding letters and send them to the relevant departments[4]; the world's first artificial intelligence legal assistant, ROSS, can help legal teams quickly and efficiently search a

huge amount of case law and find legal platforms with content relevant to the case at hand; Luminance, a British AI company that has made breakthroughs in the field of legal tech, has announced that its AI system has been able to negotiate a contract with another AI system completely autonomously without any human intervention; in addition, there are some AI software that such as Premonition, Litigate, TAX-I, etc., can not only provide users with legal advice for specific areas or cases, but also give relevant legal advice based on the adjudication results of previous cases and the prediction of the possible judgment results of current cases. China's application of artificial intelligence in the legal field started late, but has made good achievements, for example, "Law Dog", established in 2016, is a high-tech enterprise focusing on "Artificial Intelligence + Law", which is committed to providing legal advice to users through natural language processing and knowledge mapping and other advanced technologies. Knowledge mapping and other advanced technologies, to provide informationization and intelligent legal service solutions for the government, enterprises and users; launched in 2018, "Bao Xiaohei" can realize the diversified solution of massive contradictions and disputes through intelligent consulting and assessment, and the most important feature is that it can "find the answer in one second" and "find the answer in ten minutes". The biggest feature is "one second to find the answer" and "ten minutes to find a lawyer", which means that it provides 24-hour uninterrupted free legal counseling and evaluation, and finds the lawyer with the highest degree of match with the user's needs within 10 minutes. There are also AI software, such as intelligent mediation robots and intelligent petition robots, which can provide the public with online legal advice, mediation, petition and other services, effectively resolving social conflicts and improving social governance. Although artificial intelligence to provide legal advice has been applied in practice, but the risk of data leakage still occurs from time to time, and the legislation does not clearly indicate how to regulate the process

of its data operation, so this paper will analyze the data security of the various aspects of the operation of legal artificial intelligence, and talk about the risk avoidance of data from a legal perspective.

3. Data Issues in AI Legal Advice

Throughout human history, decision-making has been a fundamentally human thought process, and human beings have never created machines capable of mimicking the human decision-making process in ways that are sometimes incomprehensible and inexplicable to their creators, and due to the novel and challenging nature of the application of AI technologies in legal services, there is a lack of normative guidelines in the legislation as to how they can or should be used, and what kind of normative guidelines should be followed for their use. One of the core elements in the application of AI in legal advisory services is data, and the whole process of operation can be said to revolve around data, which makes it exceptionally important to analyze the issues in the field of law to which the data relate.

3.1 Input

In order for AI to accomplish the desired legal tasks, it must first be fed large, high-quality data sets. In other words, the credibility of an AI system depends on the quantity and quality of the input data. Two issues are inevitably involved here: data selection and data security.

First of all, from the selection of data, whether the selected data is legal and allowed to use, whether the data source is authoritative enough and accurate and comprehensive, whether it involves copyright issues and so on are all issues that need to be considered in advance, only in the source of no error to ensure that the next every step can be carried out in an orderly manner. At present, the legal data of the artificial intelligence system mainly comes from the general domain big model and fine-tuned legal model, GPT-4 realizes the network learning related information, and the more professional Beida Yuanfa Intelligence System's self-developed legal big model is the result of full pre-training from scratch on the legal domain data. It can be seen that, in order to fully apply artificial intelligence in the legal field, we

can not simply apply the "fetishism", and autonomous network learning does not seem to work in the legal field, one of which is that the legal knowledge contains a large number of laws, regulations and principles of the guidelines, which not only requires logical thinking but also requires the construction of the system, and the other is that legal information on the network is a mixed bag, and artificial intelligence lacks the ability to learn information. The second is that legal information on the Internet is a mixed bag, and the lack of practical experience of AI makes it difficult to recognize the truth from it, which may cause inaccuracies and potential bias in the data source. Thus, the pre-training mode of legal data by Peking University's Yuanfa Intelligence System can be borrowed and is more appropriate.

Secondly, after screening the data in terms of scope and quality, part of the data, such as laws and regulations, judicial interpretations, contract texts, normative documents, jurisprudential materials, etc., can be directly inputted as formalized legal texts to help the AI carry out knowledge modeling, while the other part of the data, such as adjudication documents and information on the facts of the case, etc., involves the issue of source of the right of processing, because legal data involving public and private subjects are difficult to undo the damage caused by the inherent boundaries and spreading of the cyber space as more sensitive information. more sensitive information, it is difficult to undo the damage caused by the inherent borderlessness and spreading nature of cyberspace. However, without access to sensitive case details, it is impossible to optimize the algorithm to the best accuracy. Finally, in terms of data security, the protection and strengthening of technical means is certainly important, but the legal level of protection is also indispensable. At present, the effective laws and regulations, involving data leakage of the main body responsible, mainly for the unit or individual. Although the artificial intelligence system has the function of autonomous generation, but does not belong to the category of "people", it is also difficult to set up a suitable way to bear

responsibility[5]. Therefore, how to realize the responsibility of violating data security norms to units and individuals is a problem that needs to be considered in the legislation. In addition, the desensitization of input data is also inseparable from the full cooperation of the technical and legal fields, how to effectively protect the relevant information of the data subject without affecting the optimal use of the data, balancing the relationship between the protection of information and the use of data, and reducing the "senseless harm" to those who disclose the information to protect the public's "right to be forgotten". The "right to be forgotten" is a topic worth pondering.

3.2 Training

The reason why the legal judgment made by legal workers can gain public trust to a certain extent is based on the fact that they have mastered the relevant legal knowledge through a large amount of study in the early stage and have a certain degree of legal professionalism. The artificial intelligence that can simulate human thoughts can not be equated with human beings with thinking, in answering questions, there may be inconsistent with human emotions, lack of judicial temperature, predicting the results of the phenomenon of inaccurate, and the emergence of these problems will inevitably lead to the lack of public trust. Thus, it is especially important to use data to train AI to be able to simulate the thinking of legal people to provide advice. The careful setting of the validation environment is not possible without the support of a large amount of data, but it also exposes the data to the general risks associated with the use of complex technologies, so it is important that all data for validation is set in a compliant legal framework.

The next point to note is that in common law countries such as the United States, Canada and India, judicial decisions are based on similar and representative cases from the past. Therefore, the identification of the most similar cases is the primary concern of the common law systems in adjudication. The definition of similarity can be varied and considering only terminology-level and semantic-level

similarity is not sufficient to accomplish the task; legal professionals will emphasize whether two cases have similar elements. In civil law systems, on the other hand, the outcome of a judgment is determined on the basis of facts and statutory provisions. A person will be sanctioned by the law only after violating the prohibited behaviors stipulated in the law, and the whole process of legal reasoning is essentially a combination of pre-defined rules and AI technology. In addition, complex case scenarios and complicated legal provisions may require more sophisticated reasoning to analyze. In summary, using data to model the knowledge of AI, designing the application of appropriate inference rules, and iteratively verifying them is an important part of AI's ability to provide legal services, as many predicted outcomes need to be determined based on accurate and unambiguous knowledge of the law and inference rules.

Finally, in the process of training and optimizing artificial intelligence, ethical and moral standards should be considered. Because ethics is the foundation of law, law is based on ethics and reflects people's pursuit of fairness, justice and morality. The law cannot be separated from the framework of ethics, or it will lose its legitimacy and credibility. Translating abstract ethical elements into data symbols and building them into a system is a complex task that may slow down the development process and limit some of the system's functionality. These "side effects" of ethical development may directly contradict the goals of opportunistic and quick or short-term private interests. As a result, some developers may voluntarily pursue ethical goals and assume social responsibility; for others, however, legal pressure through mandatory ethical standards will be required. I argue that if AI is to be used as a tool for providing legal counseling services to the public, then its ethical audit should be an important prerequisite for certification, and while it is conceivable that robots might be subject to some kind of performance standard, no such standard exists at the moment, and the only viable source of legally binding standards is legislation.

3.3 Output

Judgment is essentially the process of arriving at a conclusion and reasoning about it adequately, and legal AI is the process whereby a human being provides a machine with information about a case that has been designed and defined, and the machine provides the human being with the required predictive information. Applying AI to legal services requires that in addition to its ability to accurately answer legal questions, it is equally important to provide process reasoning to convince the audience. Therefore, artificial intelligence in the output data, not only on the legal questions raised by the user "results" response, but also must be quoted in the legal provisions and reasoning process integrated and provided to the user, especially for complex legal issues more need to be carefully analyzed and reasoning, so this requires the user in the consultation of the legal issues. When consulting legal issues, it is necessary for the user to provide as much information as possible related to the case to the AI, so that it can extract key information from it, so as to find out the relevant data in the massive data and fully analyze and reason. From the point of view of data security, the entire counseling process will inevitably generate new information data.

It also needs to be taken into account that the limited ability of the public to understand the operation of the system leads to a limited ability to seek help, make choices, and make predictions, and that questions may be phrased very differently between professionals and non-professionals, with some questions likely to emphasize the interpretation of legal concepts, and others to involve an analysis of a specific case, especially when describing the terminology used in a particular field; for example, one can very easily name words such as homicide, arson, robbery, theft, divorce, inheritance, etc., but words such as good faith, self-help, personality, causality, legal person, labor relations, labor relations, etc. are somewhat semantic and difficult to express.

4. Recommendations for Data Issues in AI Legal Advice

4.1 Utilizing the role of Jurists

The application of artificial intelligence systems to legal advice is a long-term and beneficial work, and the formation of a high-quality, specialized team of legal experts to lead legal artificial intelligence to maturity is a key element. In terms of the selection of legal AI data, it is not enough to rely on technicians to select the data, and this process must involve professional jurists to analyze the scope and authority of the data, review the legal and legitimate sources of the data, and at the same time avoid data infringement. The data inputted into the legal AI system should cover, but not limited to, authoritative legal texts such as laws and regulations, judicial interpretations, normative documents, adjudication documents, contract texts, jurisprudence materials, legal consultation data, etc., to ensure comprehensive coverage of legal terms, concepts and theories. On this basis, a team of jurists will then categorize, sort and analyze the inputted data, and build a fine-tuned, sophisticated, and reliable system to analyze and analyze the inputted data, taking the model of the Yuanfa Intelligence System of Peking University for pre-training of the data. On this basis, a team of legal experts classifies, sorts and analyzes the input data, drawing on the mode of data pre-training by Peking University's Yuanfa Intelligence System to build a refined and systematic legal knowledge map to help the AI to learn knowledge and build systems[6].

4.2 Establishment of a Sound Data Security and Privacy Protection System

Data security and privacy protection are the foundation and prerequisite for the application of AI in the field of legal services, and it is necessary to establish a sound data security and privacy protection system to safeguard the security and legality of data[7]. Specifically, it is necessary to formulate and improve relevant laws, regulations and standards and norms, clarify the rules and requirements for the collection, storage, transmission and processing of data, standardize the scope and conditions for the use and sharing of data, and protect the data subject's right to

know, right to choose, right to access, right to correct, right to delete, and so on. At the same time, data security technology and management measures need to be strengthened, and encryption, desensitization and anonymization need to be adopted to protect the integrity, availability and confidentiality of data, and to prevent the leakage, tampering or misuse of data[8]. In addition, there is a need to establish effective oversight and accountability mechanisms for data security and privacy protection, strengthen data auditing and monitoring, identify and address data security and privacy issues in a timely manner, and provide penalties and remedies for data violations and misconduct.

4.3 Strengthening Ethical review and Regulation

Ethics and legal responsibility guide and constrain the application of AI in the field of legal services, and principles of ethics and legal responsibility for AI need to be developed and adhered to in order to safeguard the rationality and fairness of AI. Thus, there is a need for ethicists to participate in the creation and training of ethical standards of conduct for AI and for the establishment of an independent ethical review committee composed of professionals who are responsible for reviewing and overseeing the application of AI in legal counseling services to ensure that both the process and the outcome are in accordance with ethical standards[9,10].

At the same time, the appropriate involvement of academic and non-profit organizations should be supported in order to ensure comprehensive oversight. In Germany, a non-profit organization called "Monitoring Algorithms" has emerged, spearheaded by technologists and media veterans, with the aim of evaluating and monitoring algorithmic decision-making processes that affect public life. Specific regulatory tools include auditing the rigor of access agreements, agreeing on ethical guidelines for digital management, appointing specialists to monitor information, tracking the reuse of personal information online, allowing users to withhold personal data, setting timelines for data access, and not reselling data to third parties without consent[11]. This approach is worthwhile for our country.

5. Conclusion

Taking the operation mechanism of AI as the basic analysis framework, this paper analyses the data security and regulation problems in the input, training and output stages from a legal perspective. We believe that in the input stage, there are problems such as irregular data selection and difficulty in defining the subject of data responsibility; in the training stage, although the AI can simulate human thoughts, it does not have unique human feelings, so how to make the AI both rational and judicial temperature is a challenge in the AI legal consulting service; in the output stage, how to ensure that the answer result is logically sophisticated and consistent with the jurisprudence, and at the same time, it can also correspond to the level of understanding of the general public. correspond to the level of understanding of the general public. In response to these problems, we put forward suggestions such as the establishment of an expert team, the establishment and improvement of a data security system, and the strengthening of supervision, with the aim of promoting the effective integration of AI technology and legal consulting services, and accelerating the solution to the status quo of the mismatch between the social judicial demand and resources.

References

- [1] Zhang Jiahui. A Study on Service Quality Improvement of Online Legal Consulting Company. East China Normal University, 2022.
- [2] Du Shuai. Research and Implementation of Intelligent Q&A for Legal Counselling. Central China Normal University, 2021.
- [3] Zhong H, Xiao C, Tu C, et al. How does NLP benefit legal system: A summary of legal artificial intelligence. arXiv preprint arXiv:2004.12158, 2020.
- [4] Zhou Yi-wen. Research on question and answer system oriented to legal field. Hunan University, 2018.
- [5] LI Chao, LI Xuying. Challenges, Regulation and Its Limits of Online Litigation Data Security, Shanghai Law Research Collection 2022, Volume 22 - Anthology of Intelligent Rule of Law Academic Community. Shanghai Changning District People's Court, 2023: 8.
- [6] Huang Weiyi. Deep legal content Q&A model based on knowledge graph. University of Chinese Academy of Sciences, 2020.
- [7] Zhou Xiang. The auxiliary value and realization path of judicial artificial intelligence for adjudication reasoning. Journal of Law, 2024, 45 (01): 111-127.
- [8] Zhang Weili, "An Analysis of the Current Status of Information Security Level Protection," Information Security and Technology, No. 9, 2014.
- [9] Zhang Le, "The Challenge of Emerging Technology Risks and Their Adaptive Governance", Journal of Shanghai Administrative College, No. 1, 2021.
- [10] Long Keyu. Research on Legal Regulation of Generative Artificial Intelligence Application Misbehavior - A Perspective on ChatGPT and Social Bots. Oriental Law, 2023, (04): 44-55.
- [11] Collingridge D., The Social Control of Technology, Milton Keynes, UK: Open University Press, 1980.