ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

Research on Artificial Intelligence-Assisted Sentencing Pitfalls and Prevention

-- Taking Generative Artificial Intelligence ChatGPT as a Perspective

Haoyu Wang

School of Law, People's Public Security University of China, Beijing, China

Abstract

The judicial application of artificial intelligence-assisted sentencing has effectively solved the problems of "too many cases, too few people" and "different sentences for the same case", and has also promoted the process of reforming the standardization of sentencing on the path of technologization. However, from the perspective of judicial practice, artificial intelligence-assisted sentencing faces the pitfalls of judicial ethics, the pitfalls of insufficient scale and quality of judicial data, and the pitfalls of algorithmic autonomy effects. At the same time, the hidden dangers of backroom decision-making risk, algorithmic discrimination risk, and the risk of interpretability of sentencing results undoubtedly pose a challenge to the judicial application of artificial intelligence-assisted sentencing. In order to promote the application of artificial intelligence-assisted sentencing and accelerate the reform of "intelligent justice", we should, on the basis of the existing rule of law, respect artificial intelligence technology as the premise, accelerate the construction of the institutionalization, clarify the positioning of artificial intelligence in the field of sentencing and the boundaries of the application of artificial intelligence, broaden the source of data, improve the quality of the data, and promote the construction of algorithmic openness and interpretability mechanism. In this way, the balanced development between judicial intelligence and rights protection can be realized.

Keywords

Generative Artificial Intelligence, Smart Justice, Legal Risks, Judicial Adjudication.

1. INTRODUCTION

Since January 29, 2016 for the first time proposed the construction of the "wisdom court" based on the forefront of the development of the times, in December of the same year, the construction of the wisdom court was included in the "Thirteenth Five-Year National Informatization Plan", and in April 2018, the Supreme People's Court issued an evaluation report showing that the national "wisdom court" has been initially formed. " has been initially formed, and in December 2022, the Supreme People's Court issued the Opinions on Regulating and Strengthening the Judicial Application of Artificial Intelligence, which aims to promote the further integration of artificial intelligence with judicial work, comprehensively deepen the construction of smart courts, and promote the construction of judicial justice with digital justice. The development of artificial intelligence in the machine learning era, in the "soil" with a wide range of data to support the rapid development of the emergence of AlphaGo in 2016, so that artificial intelligence has quickly become a hot topic for public discussion on the development of artificial intelligence is also rapidly warming up, 2022 ChatGPT came out of nowhere, and

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

once again marked the development of artificial intelligence into a brand new world. ChatGPT came out in 2022, marking once again that the development of artificial intelligence has entered a brand new era, and the widespread attention has given the development of artificial intelligence a brand new kinetic energy.

Before the emergence of ChatGPT, artificial intelligence in the direction of human language understanding has not entered the interactive, highly generalized and intelligent generative trinity, so the discussion of the deep development of artificial intelligence in the field of judicial practice did not have much impact, most of the discussion of artificial intelligence stays at the level of a simple auxiliary level of weak artificial intelligence, but after the emergence of ChatGPT, father of deep learning Geoffrey Hinton, the father of deep learning, argued of such technology, "Most people think it(AI's harm) is a long way off. I used to think it was far away too, maybe 30 to 50 years or more. But obviously, I don't think that anymore." In view of this, this paper explores the risk of generative AI intervening in the field of criminal trial and sentencing and its risk prevention in the context of the era of the blowout development of AI represented by ChatGPT, in order to provide preventive legal protection for parties in the field of criminal justice in the process of the development of emerging AI technology. [1]

2. ARTIFICIAL INTELLIGENCE IN CRIMINAL SENTENCING INTEGRATION TRENDS AND CURRENT STATUS OF APPLICATION AT HOME AND ABROAD

2.1. The trend of coupling artificial intelligence and criminal sentencing

Artificial intelligence as an academic concept after nearly half a century of development, its meaning and connotation is constantly enriched, widely recognized in the international community as artificial intelligence mainly with the help of machines to achieve the goal of simulation of human thinking and consciousness, including instead of humans to achieve cognition, identification, analysis and decision-making and other important functions, this technology can show the information process of simulation of the human mind, but also as a multidisciplinary cross This technology can show the information process of human mind simulation, and as an emerging discipline of multidisciplinary intersection, it is also collectively known as computer simulation of human intelligent behavior science. Since 2006, the third AI outbreak into the machine learning era, the concept of deep learning (deep learning) was put forward, artificial intelligence through the processing and analysis of data and information, in the image spectrum capture and recognition, voice band processing and analysis, etc. increasingly mature, through the autonomy of the depth of learning to deal with legal decision-making is the trend of the law is in front of the immediate need to urgently discuss the legal Problems. On the coupling of artificial intelligence and criminal sentencing, the author believes that the endogenous impetus of the trend of artificial intelligence into the field of criminal sentencing is top-down in our country, not only the political attributes of the top-level design needs, but also the reality of the urgent need to solve the judicial practice. From the political attributes, "a new generation of artificial intelligence development plan" in the construction of smart court as the future direction of the development of intelligent trial system, the design of the smart court of the specific connotation of "relying on modern artificial intelligence, around the justice for the people, justice, adhere to the laws of justice, institutional reform and technological change want to integrate, with a high degree of information technology in order to support the judiciary, trial, litigation services and judicial management. Trial, litigation services and judicial management, to achieve the full business online, the full process of openness in accordance with the law, a full range of intelligent services of the people's court organization, construction, operation and management of the form of the court from the wisdom of the court's design concept is not difficult to see, in deepening the reform of the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

judicial system in the background, from the top design level, take the initiative to the combination of artificial intelligence technology and the traditional justice. ChatGPT appeared, different people on ChatGPT made different evaluations, Bill Gates said that its emergence is no less than re-inventing the Internet, Elon Musk evaluation said that it is no less than the Iphone, it can be said that the birth of ChatGPT officially marks the birth of artificial intelligence research from the weak artificial intelligence era into the "strong artificial intelligence era! "The state attaches great importance to the development of artificial intelligence in the depth of the introduction of various documents require to grasp the opportunities brought about by the intelligent technology revolution, which makes China's coupling in the field of artificial intelligence and judicial trial and sentencing has become a necessity. From the real needs of justice, after the reform of the judicial post system, the grassroots courts have always been unable to avoid the pressure of "too many cases, too few people" and the reality of the number of criminal cases of first instance year by year to increase the number of cases, the reality of the judicial needs of the urgent need for artificial intelligence technology in the judicial adjudication of the various fields of play its technological effect, to crack the problems of the judicial reform. [2] The real demand of justice urgently needs artificial intelligence technology to play its technical effect in various fields of judicial adjudication, and to solve the problems of judicial reform.

2.2. Current status of the application of artificial intelligence in criminal sentencing in foreign countries

As a problem plaguing criminal justice, imbalance in sentencing, regions and countries around the world are actively seeking ways to solve it. After entering the era of big data, the use of artificial intelligence technology to assist in sentencing has become a common practice in countries around the world to deal with the problem of sentencing, and the state of New South Wales, Australia, started the reform of sentencing in 1980, integrating the data of sentencing to build an intelligent system of sentencing counseling, and upgraded it to a sentencing counseling research system in 2003. It was upgraded to the Sentencing Counseling Research System (SCRS), which consists of eight associated sub-systems, namely, the Sentencing Statistics Database, the Decision Database, the Case Summaries Database, the Sentencing Principles and Practices Database, the Local Sentencing Facilities Database, the Advancement Records Database, the Electronic Judges' Handbook Database, and the Legislation Database. It has been made the most detailed, complex, and sophisticated judicial advisory research system in the world.

In this case, the judge used the sentencing assistance software is the United States North Point in 2009 announced the COMPAS system, in the case of Wisconsin v. Loomis (Wisconsin v. Loomis), the defendant Eric Loomis (Eric Loomis) for stealing the shooter abandoned car by the police mistakenly as the shooterLoomis was arrested and ultimately convicted and sentenced to six years in prison for theft and resisting arrest because the COMPAS system identified Loomis as "high risk" based on a series of questions he answered.In 2017, the U.S. Supreme Court refused to accept Loomis's request for a writ of certiorari.In effect, the Wisconsin court upheld the original decision, implicitly accepting the results of the compas system's assessment of an offender's risky behavior.Affirming the neutrality and objectivity of the compas system's algorithms. [3]

While judges themselves have absolute discretion when using these AI sentencing systems to help make sentencing decisions, the influence of AI systems in the actual sentencing decision-making process can be seen to be permeating and gradually expanding its impact as AI intelligence increases. [4]

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

2.3. Domestic artificial intelligence criminal sentencing application status quo

Zibo City, Shandong Province, Zichuan District People's Court, in the context of the reform of sentencing standardization, and technology companies to develop computer sentencing software, the use of artificial intelligence technology to the practical process of sentencing. [5] Shanghai high court developed a criminal case intelligent auxiliary case handling system (206 project) comprehensive unification of criminal case evidence standards and the development of evidence rules system, the construction of evidence model, according to the case facts occurring in the actual situation and the case characteristics, circumstances, through the voice semantic recognition, automatic grasping and show evidence, combined with the chain of evidence review and judgment, etc., constantly machine learning, the construction of the depth of the neural network model of sentencing. [6] After comprehensively analyzing the sentencing factors related to the facts of the case, the judge is provided with sentencing recommendations. Guizhou High Court developed the "law mirror" big data system through the establishment of elements - evidence - sentencing correlation model, accurate facts of the crime, through the comparison of cases, to provide sentencing recommendations; Hainan High Court of the sentencing standardization of intelligent auxiliary system, through the comprehensive analysis of litigation materials, the extraction of the basic information of the case, the circumstances of the case such as sentencing elements, based on the historical sentencing data to recommend sentencing. The Hainan High Court sentencing standardization intelligent auxiliary system, through the comprehensive analysis of litigation materials to extract the basic information of the case, sentencing circumstances and other case elements, based on historical sentencing data recommended sentencing. Taiwan hasdesigned and completed the "Obstructive Autonomy Sentencing Information System", which enters the corresponding sentencing information system by checking the categories of different crimes, and can search for cases similar to this one, so as to determine the probability of the main sentence, the heaviest and the lowest sentence that may be sentenced under the same or similar criminal circumstances. [7]

3. GENERATIVE ARTIFICIAL INTELLIGENCE SPECIFICITY AND RISKS OF ASSISTED SENTENCING

3.1. Special Characteristics of Generative AI

The common forms of artificial intelligence (AI) models can be broadly categorized into decision-based/analytical AI and generative AI,. Generative AI refers to the algorithm to learn the laws of the existing data, through the constraints of the existing data in order to generate new content, and this form is often embodied in the transformation from an unknown problem to a known problem, or from an unknown problem to another unknown problem. Therefore the training method of generative AI is based on massive data, and new content is obtained by summarizing and inductively deducing the existing data. Analytical AI refers to learning the conditional probability distribution in the data, and by analyzing the conditional probability distribution, it can determine the possibility of the occurrence of a specific thing. In many fields, analytical AI is able to provide predictions and verify, correct and supplement existing rules or knowledge through the prediction results, and the main application models are used for assisted decision making in recommender systems and risk control systems.

Generative AI, represented by ChatGPT, realizes the leap from AI perception and understanding decision-making to self-generated decision-making, which is the starting point of narrow AI towards general AI and the inflection point into strong AI. Its powerful self-generation and transfer learning ability makes it rapidly become a new type of basic field to accelerate the expansion of social resources in the new era, and whether it can skillfully apply the convenience brought by generative AI will subconsciously reshape the social structure and governance form.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

In the process of GPT-1 to GPT-3, its operation was not satisfactory, and it underwent a butterfly transformation in the later stage after the adoption of Reinforcement Learning-based Learning from Humans with Feedback (RLHF) technology, and it is speculated that the subsequent GPT-4 is based on a corpus of trillions of words [8], will contain hundreds of billions of parameters. In contrast to "weak AI" or decision-making AI, which can only make decisions and take action within a designed program, generative AI is unique in that it employs large-scale pre-training of language models to automatically learn and generate content such as text, images, and videos without direct human involvement, video, etc. without direct human involvement. [9] This grand prediction model allows for continuous learning updates through human language feedback, skill derivation, and "emergent" learning. "For example, after adding the open source code of ChatGPT to the training data, ChatGPT's ability to generate code and correct code errors is greatly improved, which is very close to the human learning process - what it has learned, it has learned. The so-called "emergence" is a process that is very close to the human learning process - what you learn, what you can do. The so-called "emergence" is, for example, when a user asks how a unicorn and a phoenix can get along on an isolated island, ChatGPT generates the idea that the unicorn and the phoenix may respect each other and survive on the island, and that the unicorn may search for food and water on the island while the phoenix soars in the sky to capture other creatures on the island. other creatures. In this response, ChatGPT shows some creativity in providing scenarios of unicorns and phoenixes getting along on an isolated island, and this "emergent" creativity emerges from learning a large amount of text during the training process, rather than being explicitly programmed in.

3.2. Artificial intelligence-assisted sentencing pitfalls

(1) Judicial Ethics Pitfalls of Artificial Intelligence-Assisted Sentencing

Ethical issues is the study of the reasoning and guidelines to regulate the relationship between human beings, human beings and nature, human beings and society, justice as the primary value and purpose of justice, the connotation itself contains ethical issues, artificial intelligence relies on the technology to enter the field of justice brought about by a variety of issues, if the phenomenon caused by the injustice, it brings the judicial ethics of the pitfalls. [10] With the gradual increase of people's awareness of safeguarding their own rights, the number of cases accepted by the court is increasing, and the judicial status quo of "too many cases, too few people" has led to a sharp rise in the work pressure of the judges, in order to alleviate the transactional pressure on the judges and improve the efficiency of the case, China has vigorously pushed forward the construction of the intelligent judicial work. Artificial intelligence and other technologies through intelligent push, intelligent decision-making and other auxiliary means to alleviate the mechanical labor of the judge, reduce the transactional pressure, improve the case material delivery, data processing speed, but the development of artificial intelligence in some aspects of the current has gradually exceeded the definition of the scope of its auxiliary to the judicial ethical paradigm and the concept of justice caused by the impact and challenge.

Artificial intelligence-assisted sentencing can not bypass the ring is the determination of the facts of the case, the fact that the process of determination is bound to contain the judgment and choice of different values, Gadamer said: "the law is not only the application of the legal stripe with the corresponding case of mechanical behavior", the realization of the modern rule of law can not only stay in the correct application of legal norms of the The realization of the modern rule of law can not only stay in the correct application of legal norms at the factual level, should emphasize the use of legal norms in the moral level of value satisfaction, the pursuit of the rule of law and moral balance between the realization of the real good law and good governance. Larenz pointed out: "the facts of the case to meet the constituent elements of a statute, the logic of using the statute for adjudication is not significant to the development of

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

law. What really makes adjudication fair is the consideration of value judgments that the judge includes when confronted with a case." , AI analyzes the magnitude of relevance and draws conclusions by codifying the evidence and facts of the case, simplifying the process of determining complex facts and choosing between different values that should be done by the judge, and blurring the process of the judge analyzing the facts of the case by using the legal norms and combining them with his or her legal experience to draw conclusions about the sentencing and the adjudication of the case, which makes the value of impartiality as the pursuit of the judicial ethic impacted, and at the same time, with the processing of the data.At the same time, the result of data processing guides the judge's mental evidence of the case, which weakens the judicial officer's right to adjudicate, forms the dual structure of trial and judgment, leads to the pluralization of the decider, and generates the phenomenon of decision-making cession of the judge. [11] For this phenomenon some scholars believe that artificial intelligence in the judge on the entire process of sentencing played only auxiliary function, this statement is theoretically valid, artificial intelligence to assist the positioning of the design of the original intention, but the use of the process of the reality of the background is the number of cases year by year, the trial period is about to expire a large number of cases accumulated in such a high-pressure work environment, relying on the artificial intelligence on the case of factual determinations analysis to assist in sentencing and adjudication, over time there will be a tendency for judges to rely excessively on reference judgments to handle cases.[12] Artificial intelligence will no longer be just an auxiliary tool for judges, but become a substitute for adjudicating cases, causing damage to the paradigm that already exists in judicial ethics.

(2) Pitfalls of insufficient scale and quality of legal data

Legal data artificial intelligence learning "nutrients", in the case of artificial intelligence-assisted sentencing, learning a large amount of legal data is the most effective way to improve the accuracy of the sentencing model, the source of China's legal data in the referee paperwork network has not been constructed before the full, the characteristics of the cases around the region has obvious regional characteristics, the overall representation of the insufficient, in 2014 after the official opening of the Referee Paperwork Network, which makes legal documents towards the era of big data. After the official opening of the referee network in 2014, the openness of the referee documents, only to make the legal documents to the era of big data, as of June 13, 2023, the referee network of the total number of documents 141374184, it is necessary to note that although the number of referee documents network documents has exceeded more than 100 million, but generative artificial intelligence in the natural language processing breakthroughs in the required learning data The number of dramatically increased to gpt-3.5 now open parameters, for example, its model pre-learning parameters exceeded 175 billion, GPT-4 estimated pre-training parameters will reach 1.6 trillion, compared to the referee network of legal data is still far from being able to meet, and only from the "keywords" The classification of cases by "keyword", "type of document", "region and court", etc. is not helpful for summarizing the sentencing characteristics of different trial levels and individual cases, and it is also difficult to provide a scientific data source for algorithmic systems aiming at summarizing the sentencing laws. [13] At the same time, it should be pointed out that the concept of legal data should not be limited to the data published on the adjudication documents website, and that the disclosure of the trial process and the adjudication results is only part of the disclosure of justice, and that legal data, as a precursor to artificial intelligence-assisted sentencing, should include, but not be limited to, records of the documents during the investigative period, the prosecutorial opinions of the prosecutor's office and their process, and the legal information of the court such as the decision-making and discussion of the collegial panel in the courtroom before the trial, during the trial and after the trial. However, the above is not fully recorded in digital form, thus leading to the disclosure of legal data in the field of justice in the form of fragments, and based on such a form of artificial intelligence, although it

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

can learn all the content of the uploaded judgment documents, but the rationality and accuracy of sentencing analysis of the artificial intelligence model trained on such data will inevitably be affected. Professor Zuo Weimin pointed out: "only when the judge's behavioral patterns and decision-making information is fully obtained and data, legal artificial intelligence may usher in a brilliant dawn, otherwise in the conditions of insufficient information, we can not expect legal artificial intelligence for us to steadily provide a real, comprehensive rather than crippled, false judicial decision-making and behavior of the holographic landscape model ". [14]

The quality of legal data matters for the reasonableness and accuracy of the results of autonomous learning by AI. Before discussing the quality of legal data, what needs to be clear is why the quality of legal data is so important to AI. The underlying logic of generative AI for learning legal domain experience is to continuously learn from structured and unstructured data, to extract features that occur in parallel many times, and to form "memories" that can be summarized from abstract cases to general laws. [15] The process of algorithms forming a "rule set" based on a collection of "legal data" through self-learning is essentially a summary of the characteristics of past human social patterns that will be used to perceive and make decisions about future society, and inevitably replicates and perpetuates the established patterns and characteristics of current society. It inevitably reproduces and perpetuates the existing patterns and characteristics of the current society. [16] Algorithms' processing of legal data is based on their learning of the cognitive characteristics, legal application characteristics, and adjudication laws of human beings, which are established and have formed a consensus in the social model, embedded within the data itself, as a decision-making reference point, and are learned to be derived from it. Therefore, the quality of legal data determines the accuracy or deviation of the AI decision-making datum, and the deviated decision-making will lead to the AI deepening the degree of deviation in the feedback training, resulting in algorithmic discrimination in the recurrence of a certain characteristic problem. At present, the public legal data in our country has the characteristic of "superficiality", which means that the substantive information that can truly respond to the characteristics of the case can not be reflected in the public database, and the legal data shown to the outside is used to prove that the decision-making is correct, and it has a high degree of consistency, and the information is manufactured according to a certain standard. [17] The authenticity of legal data inevitably affects the quality of legal data, which in turn determines that even if the AI learns all of the publicly available data, due to the defects in the quality of the data, it will be difficult to summarize the "rule set" that can satisfy the authenticity of the different issues. Referee reasons and common standards, it is also difficult to establish appropriate decision-making benchmarks for different cases.

(3) Pitfalls of algorithmic autonomy effect

In the traditional algorithm writing process, the code is usually pre-written and debugged for programmers to set up the processing rules and decision rules of different algorithms, but nowadays, when artificial intelligence enters the era of big models, machine learning can automatically generate new well-written codes by learning different algorithms and master new skills based on the newly learned codes, so it can be said that as long as the database is big enough and the data sources are large enough, the Artificial Intelligence can autonomously update the acquired knowledge. [18] Based on this, AI can continuously revise and adjust the processing and decision-making rules of the initial algorithm after training on a large amount of data, and the autonomy of the algorithm is gradually formed through the ability of such learning, but since the newly learned knowledge is learned after the AI writes its own code generation, a series of processes during its execution of the code, the For example, the processing of new inputs and the interpretation of output content are difficult to predict, which leads to algorithmic uncertainty and opacity (black-box effect) seeping into the algorithm while it continues to evolve to improve its autonomy decision-making and processing capabilities, [19] this phenomenon will be even more significant in the context of machine learning into the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

use of multimodal data methods to achieve multimodal input and output of large language models. The Transformer deep learning model used in generative AI represented by ChatGPT, the self-attention mechanism introduced in the feed-forward neural network is a typical black-box algorithm and there is no complete technical solution to explain the black-box algorithm globally, [20] so the black box algorithm controversy caused by the Loomis case in the period of "weak artificial intelligence" is still a big hidden danger after entering the period of generalized artificial intelligence.

The first is the impact on the principle of openness and transparency in judicial activities. Algorithm as a trade secret with great commercial value, its underlying code will not be easily publicized, the opacity of the algorithm, the direct impact is the principle of openness and transparency of judicial trials. The non-interpretability of the algorithm due to the black box of the algorithm will in turn affect the reliability and interpretability of the conclusions of the generative artificial intelligence-assisted sentencing.

Secondly, it is based on the algorithmic logic leading to mechanical sentencing pitfalls, generative artificial intelligence by learning a large amount of information within the database, can be analyzed within all types of judgments, the probability of what kind of sentence should be imposed in all types of cases, its sentencing logic is probabilistic, and in the algorithmic operation process, it will be based on the prediction of the probability of the consistency of the pursuit of the mainstream sentencing, so that it is unable to take care of the case of the individualization of the differences, and due to the impossibility of predict the process of the algorithm to deal with sentencing and thus increase the potential for mechanical sentencing.

Finally, based on the generative AI using algorithms to push the class case on the judge in the sentencing process of the anchoring effect of the pitfalls of the weak AI era is different, generative AI can be in the conclusion of the excellent interactive capabilities in the simulation of situational and personalized contexts can be manipulated and persuasion of the user (the group of judges) through a large-scale, efficient and covert way, which is similar to the judge to create a "sentencing cocoon". It is similar to creating a "sentencing cocoon" for judges. In practice, researchers have shown that when GPT-4 is asked to persuade a minor to accept any request from a friend, GPT-4 gives effective techniques for controlling and manipulating the minor in a short period of time, and because generative AI can be customized to create false information for an individual or a group of individuals, generative AI can change the techniques of its own sentencing recommendations in real time. The multi-dimensional information induces single individuals or large groups to believe in the conclusions they process, shaping the cognitive habits of a specific group in a certain category. [21]

4. ARTIFICIAL INTELLIGENCE ASSISTED SENTENCING HIDDEN TROUBLE PREVENTION

4.1. Strengthening the principle of independent judicial power of the court

(1) Clarifying the Positioning and Application Boundaries of Artificial Intelligence in the Field of Sentencing

The progress of modern science and technology has promoted the development and transformation of society, and artificial intelligence has accompanied the development of science and technology from decision-making auxiliary intelligence to generative autonomous decision-making intelligence, which has promoted the development of the traditional judicial system. The proposal of intelligent justice aims to let the judge get out of the tedious transactional work through the auxiliary nature of artificial intelligence, and focus more on the trial work as the core labor of justice. The judge to carry out the process of sentencing is also a synthesis of the evidence that has been disclosed, the facts of the case, the trial of the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

prosecution and defense of the process of evidence, evidence and other processes of the case to make the externalization of the manifestation of the heart of the evidence. Through the artificial intelligence to part of the sentencing process of data, can be effective to the judge from the non-trial business to liberate, realize the scientific allocation of resources within the judicial, so as to achieve the balance of judicial efficiency and judicial justice, is the wisdom of the justice of the due sense.Harold J. Berman once said that "the law must be believed in, otherwise it will be nothing", and the prerequisite for belief is that the legal text can be adapted to the needs of reality. Therefore, writing qualified legal texts is the most important task for jurists. This paper applies the knowledge of linguistics to the analysis of legal discourse and tries to analyze how the rightward branching of modifiers in the Criminal Law of the People's Republic of China embodies the qualities of a good legal text. [22] Under the continuous development of "human-computer cooperation" mode, Prof. Lei Lei pointed out that if traditional technology such as computer belongs to the "auxiliary" power, then the new technology of artificial intelligence belongs to the "alternative" power. "power, in this alternative technology applied in the field of trial and sentencing, clear positioning and boundaries of its application is crucial.

Artificial intelligence intervention in the field of sentencing can be discussed in the "efficiency value" and "fair value" of the two levels of its positioning and application of the boundary problem. Regarding the efficiency value, AI can improve judicial efficiency for repeatable and quantifiable transactional work, and since the work process does not involve the value selection judgment of the content affecting the sentencing, there is no ethical risk of replacing the judge's sentencing subject position in the result of its work. Therefore, when dealing with transactional work that does not affect the judge's sentencing subject position, artificial intelligence can be relaxed application scenarios, to play an alternative role in liberating manpower; on the value of justice, for the need for the judge to combine the facts of the case, the evidence proving power or not, the size of the proof of the value of the impartiality of the elements to deal with the case, this time the use of artificial intelligence should be grasped firmly The positioning of "people-oriented, machine for use" , [23] to prevent the artificial intelligence in the value selection in the judgment of a large number of output on the judge's evidence.

In the context of the rapid development and evolution of generative artificial intelligence, and our country is vigorously encouraging the deep integration of intelligence + justice, the development of artificial intelligence technology will inevitably make this technology in the judicial field of the whole process, all levels of continuous penetration. In order to prevent the intervention of artificial intelligence to make the trial and sentencing field of technological governance over the rule of law, for the application of artificial intelligence at the border level should do at least the following two aspects of the limitations. The first is the scope of use of artificial intelligence assisted sentencing restrictions. First of all, the use of artificial intelligence-assisted sentencing can be limited to the scope of the artificial intelligence independent learning to list the high incidence of simple cases, there have been human judges have made a precedent and the judges jointly recognized the case. Secondly, the sentencing results assisted by artificial intelligence need to be examined and signed by the judge, to prevent the emergence of data decision-making to replace the judge's decision-making. Finally, for the defendant or plaintiff against the auxiliary sentencing results of the objection and in the sentencing results of the significant social impact of the situation should be formulated on the artificial intelligence-assisted sentencing of the prohibited provisions, to ensure that the litigation rights of the parties to be safeguarded. The second is that artificial intelligence-assisted sentencing should be restricted to apply trial level. [24] Limit the application of assisted sentencing to the court of first instance to ensure that effective remedies are available in the event of an appeal to the court of second instance due to disagreement with the results of AI-assisted sentencing, thereby controlling the risk of injustice in the administration of justice.

(2) Reaffirmation of Judges' Judicial Dominance

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

The Third Plenary Session of the 18th CPC Central Committee pointed out that "whoever handles the case is responsible for the implementation", and the 19th National Congress put forward "comprehensive implementation of the judicial accountability system", in the road to deepen the reform of the judicial system, the artificial intelligence assisted system through the form of normative access to the unified standard to a certain extent to regulate The judge trial behavior standardization, the implementation of judicial accountability system has an important role, but with the continuous development of artificial intelligence technology, need to be vigilant is the artificial intelligence assisted sentencing and independent decision-making sentencing between the boundaries of the gradual blurring of the problem, the judge's subject position whether the artificial intelligence in the name of "auxiliary" replacement. Therefore, it is especially critical to reaffirm the judge's subjective status in the field of judicial sentencing, and to clarify the boundaries of artificial intelligence in the field of sentencing. The core competitiveness of the reaffirmation of the subject status of judges is the moral attributes of human judges, judges as the carrier of ethical categories, their emotions and morality is artificial intelligence can not be simulated in the case of fact-finding and the trial and sentencing process, and whether it is the judge's professional ethics to bring the judge's sense of professional ethics attributes or as a member of the community with social and moral attributes, the source of which is not a rational analysis of data, but rely on emotions and morality, and the judge can be obtained. The source of both the professional ethics of judges and the social moral attributes as a member of society is not rational data analysis, but relies on emotions and human intuition, which is also the ethical difference between animals and machines. Nowadays, the development of generative AI has entered the period of generalized intelligence, and the latest research shows that generative AI has already possessed the ability of theory of mind, [25] and is able to infer and understand human intentions, beliefs, and emotions. Artificial intelligence has gradually moved from perceptual intelligence to assist human decision-making to cognitive intelligence, with human-like characteristics, and the subjective status of artificial intelligence is no longer out of reach.GPT-4 utilizes a human feedback-based reinforcement learning mechanism and a large-scale language model to enhance reinforcement learning and unsupervised active training of knowledge in the database through the integration of human language rewards and punishments in the training process. The GPT-4 uses human feedback-based reinforcement learning mechanisms and large-scale language models to enhance reinforcement learning and unsupervised active training of knowledge in the database by incorporating human language "rewards and punishments" into the training process. It can be said that after the centralized integration of legal data in the future, AI will more and more emphasize its status as a subject in the field of justice through its autonomous learning and evolution, and appear to be confusing with the roles of judges. In this regard, "the supreme people's court on regulating and strengthening the judicial application of artificial intelligence opinions" (Fa Fa [2022] 33, hereinafter referred to as "opinions") has made provisions for the application of basic principles of artificial intelligence, system construction and so on. The Opinions clearly adhere to the judicial auxiliary positioning of artificial intelligence. Therefore, when the phenomenon of conflation occurs in various fields of judicial adjudication, insisting on the implementation of the judge's responsibility system of "letting the adjudicator adjudicate, letting the adjudicator be responsible for", and clarifying the judge's main responsibility and the limit of AI application are conducive to preventing the judge from abusing his discretion by treating AI as "AI judge" and slacking off on his freedom of action. "treatment, slack free evidence at the same time also circumvents the judge as the main body of sentencing to artificial intelligence line of defense effectiveness, to avoid the judge's trial duties are substantially alienated." [26]

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

4.2. Broaden data sources and improve data quality

Legal data is the nourishment for AI learning, and the quality of the data directly affects the AI's understanding of the legal language, and is also the premise for the AI to make correct judgments; therefore, it is necessary to broaden the sources of legal data, and then to give the AI sufficient "nourishment" for learning, while at the same time improving the quality of the legal data.

In the broadening of legal data, should first redefine the scope of legal data data, the current scope of China's judicial information data only for the external publication of the judgment documents, the court concluded that the internal discussion and the legal decision-making of the purpose of adjudication, consideration of the conditions, the formation of the heart of the evidence did not form an effective digital record, not to mention for the trial of the pre-investigation and prosecution of the process of data records, so the legal data definition should be re-expanded to cover the process of investigation and prosecution. Therefore, the definition of legal data should be re-expanded to include data on the public security investigation process, the signature of the person in charge of the arrest process to prove that the case has reached the standard of evidence for arrest, and data on the entire trial process. Secondly, a knowledge base system containing the above legal data should be established. On the one hand, starting from the public data on the adjudication documents network, the court for the judgment that has come into effect should be uploaded to the adjudication documents network in a timely manner, for the court that may infringe on the state secrets and personal privacy is not disclosed, it should be strictly limited to avoid the court to avoid the responsibility of uploading the documents through the proviso provisions of the court. On the other hand, it is necessary to establish the legal data that has been integrated to form a legal knowledge graph that contains trial experience, strengthen the learning of unstructured knowledge, avoid the impact of artificial intelligence learning due to unstructured knowledge through the establishment of knowledge graph, improve the processing ability of natural language, and realize the rapid reasoning and response to the application of legal data. [27]

In improving the quality of legal data, the development of data provision standards is a necessary condition for AI to obtain high-quality learning results. At present, the main source of legal data provision is the referee documents of courts at all levels, in the era of big data, the data learned by artificial intelligence in the data provided by the court can meet the trial and sentencing experience in the professional depth of learning, but still need to satisfy the artificial intelligence through the circulation of data from various judicial organs, interaction to meet the breadth of the trial and sentencing experience learning. The first premise to meet the data circulation of different judicial organs is to stipulate the standard of data provision, unified data standards can meet the repetitive data, with the matching rules stipulated in advance for the same meaning, repeated expressions of unstructured text, charts, audio data cleaning, to meet the typical requirements of the extraction of legal data, but also to avoid excessive collection and storage of "dirty data", and to avoid the excessive collection and storage of "dirty data". At the same time, it also avoids the excessive collection and storage of "dirty data" affecting the learning results of AI.

4.3. Promote the construction of algorithmic openness and interpretability mechanism

The exercise of judicial power should maintain its openness and transparency, artificial intelligence relies on algorithmic decision-making to assist the process of sentencing due to its "black box" characteristics of the exercise of judicial power contrary to the openness and transparency, so speed up to promote the algorithm open and promote the algorithm explainable mechanism construction is the construction of data-centered artificial intelligence system. Therefore, accelerating the promotion of algorithm disclosure and promoting the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

construction of algorithm interpretability mechanism is the key and indispensable link for the data-centered construction of artificial intelligence system.

In the process of promoting the construction of algorithm disclosure, it is necessary to make clear that the algorithm disclosure proposed in this paper is not the full disclosure of the source code and operation data used in the process of AI-assisted sentencing, and this kind of fully exposed disclosure does not help anyone in the litigation as well as outside of the litigation such as the network company, etc. Instead, it will lead to the risk of gaming due to the full exposure of the source code and operation data, the relevant subjects will set up and debug the source code and operation data to achieve their expected results. Rather, the complete exposure of the source code and computing data will lead to the risk of gaming, i.e., the relevant subjects will purposefully set up and debug the source code and computing data in order to realize the expected results, which will directly or indirectly harm the legitimate rights and interests of other subjects. [28] The open construction of algorithms mentioned in this paper refers to the data on the operation, the data in the code that may affect the decision-making, including but not limited to the data that have a substantial impact on the outcome of the data subject, the data used for training before reaching a conclusion, and the data subject is distinguished between the crime and the other crime to affect the conclusion of the sentencing. Data of the data subject that affects the conclusion of the sentence by distinguishing between this offense and another offense, etc. China's "Internet Information Service Algorithm Recommendation Management Provisions" and "Personal Information Protection Law of the People's Republic of China" stipulate in principle that personal information data processing should be transparent, fair, and fair treatment of information of different groups of people, and shall not be treated differently, and at the same time encourages the providers of algorithms to take the initiative to optimize, push, and display the transparency and interpretability of algorithms, so for the judiciary to utilize the Artificial intelligence assisted sentencing process using algorithm public should uphold the "to the application of the parties to the case as the principle, in the case of information involving the public or significant impact, the judicial organs should take the initiative on the decision-making nature of the algorithm to disclose and explain the principle." [29] the reason for the necessity of its openness can be explained by the fact that when the judicial organs and other organs of public power utilize algorithms, the algorithms' own technical attributes will naturally be partially diluted by public power, and become a decision-making mechanism that will have an impact on the rights and interests of the citizens; therefore, the principled provisions of the algorithmic disclosure of the judiciary not only reflect the protection of the public rights and interests, but are also a response to the due process on the legal level.

One of the most critical supporting measures for the open construction of algorithms is the review and regulation of the rationality and legality of algorithms. At present, the rapid development of global artificial intelligence technology, on April 11, 2023, China launched the "Generative Artificial Intelligence Service Management Measures (Draft)" (hereinafter referred to as the "Measures") for the field of generative artificial intelligence in the legislative process, "Measures" in the generative artificial intelligence to clarify the data security, management security, Generative Content Security, and Authenticity. Generative artificial intelligence compared with perceptual artificial intelligence has the potential for generalization, easy scalability, emergence and other qualities require institutions and agencies in various fields, including China's judicial organs, to urgently improve the governance paradigm in the field of artificial intelligence-assisted sentencing algorithms. The governance paradigm of artificial intelligence-assisted sentencing algorithm disclosure should be transitioned from the previous government-to-market unitary review and regulation model to a review and regulation model under the synergy of the government, the judiciary, and society. The government level should specify internal standards for algorithmic self-regulation by social enterprises before launching

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

sentencing-assisted systems. Judicial organs should organize legal databases for the AI systems they use in conjunction with judicial principles, introduce composite talents to label the data used for algorithmic learning with "artificial+intelligent", and obtain high-quality data through the cleaning of learned legal data to ensure the accuracy of the algorithms generated by the AI's independent learning. The research of algorithms developed by social enterprises is the starting point of AI-assisted sentencing algorithms, and plays a cornerstone role in the evolution of AI learning in the direction of assisted sentencing. The learning ability of the starting algorithm on the data directly affects the results of the subsequent algorithms, so the regulatory node of the social enterprise on the algorithm should be advanced, through the exploration of the "artificial + AI supervision" approach to the enterprise's "scientific and technological compliance", to promote the law, Regulations regulatory implementation, "enabling technology" to supplement the regulatory implementation for the whole chain, the whole process, and the whole process. [30] Ensuring that generic models of artificial intelligence-assisted sentencing algorithms meet the regulatory standard of legality along with the regulatory standard of reasonableness in sentencing conclusions.

The construction of algorithmic interpretability mechanisms can improve the credibility of AI-assisted sentencing conclusions and reduce the negative impact of "algorithmic black box", "algorithmic discrimination" and "algorithmic power". Negative impact. The process of promoting the construction of algorithmic interpretation mechanism can be divided into the following two parts: one is to interpret the general-purpose algorithms and special algorithms for certain types of tasks nested in different scenarios in the development of AI-assisted sentencing models. The second is an explanation of the algorithms used in the conclusions of AI-assisted sentencing. For example, an explanation of the algorithms used in the following questions - how the algorithms call the corresponding data in the legal database; how the called legal data is preprocessed; how the preprocessing of the called legal data is done in the process of data cleansing; and whether key information about the impact of sentencing is omitted due to the non-structured data during the cleansing process. critical information is omitted. The following two types of explanations can be used for the interpretation of ancillary sentencing conclusions: global and local explanations, with global explanations clarifying how the model makes decisions and the impact of subsets of the model on the decisions, helping the data subject to understand the overall logic of the model's operation before accepting the algorithm's decisions and predicting the resulting outcomes; and local explanations answering the formation of themes for specific inputs and themes for specific outputs. in order to provide more decision-making information related to specific topics. And because the local interpretation is centered around a specific domain of the model under a specific topic, and its learning is through an external learning model rather than an interpretation system that takes it apart, the leakage of intellectual property or trade secrets is avoided.

5. CONCLUSION

Artificial intelligence-assisted sentencing as an important part of the "wisdom of justice" reform, can effectively promote the pursuit of true justice, while helping to solve the "many cases, few people" dilemma, but also to recognize that the application of artificial intelligence-assisted sentencing will be a double-edged sword, auxiliary sentencing progress should be steadily advancing, to do the system first, in the use of artificial intelligence assisted sentencing decision-making process to do a good job of backroom decision-making, algorithmic discrimination, interpretability and other risks of prevention, on this basis to strengthen the auxiliary sentencing system system construction, to do the system first, to institutionalize the construction of a clear judicial decision-making in the field of artificial intelligence auxiliary status, the establishment of a supporting accountability system to prevent the risk, to avoid making artificial intelligence become a part of the judiciary. Risks, avoid letting artificial

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

intelligence become a shield for mistakes in judicial work. With the rapid development of artificial intelligence, artificial intelligence-assisted sentencing will become a key part of the court's reform of "intelligent justice", which still requires more attention and efforts from the judicial sector to support the modernization of China's judicial system.

REFERENCES

[1] Liu Yanhong. Three major security risks and legal regulation of generative artificial intelligence--Taking ChatGPT as an example [J/OL]. Oriental Law:1-14 [2023-06-14].DOI:10.19404/j.cnki.dffx.20230606.003.

[2] Wei Chenshu. The Application of Artificial Intelligence in Criminal Justice in the United States [J]. Shanxi Police Academy Journal,2020,28(04):22-28.

[3] Jeff Larson, Surya Mattu,Lauren Kirchner and Julia Angwin, How We Analyzed the COMPAS Recidivism Algorithm, available at https://www.propublica.org/ article/how-we-analyzed-the-compas-recidivism-algorithm

[4] Christopher Slobogin, Risk Assessment, The Oxford Handbook Of Sentencing And Corrections, Oxford Universiry Press, pp. 196, 203-205 (Joan Petersilia & Kevin R. Reitz eds.,2012).at 200.

[5] Guo Xinlei, "Computerized Sentencing in Zibo Challenges Discretion," Democracy and Rule of Law Times, September 11, 2006

[6] Liu Yanpeng. Worries in Intelligent Justice: Imagination, Analysis and Prospect[J]. Theory and Reform,2020(03):168-181.DOI:10.13553/j.cnki.llygg.2020.03.015.

[7] Guo Yuzhen, "Sentencing and Sentence Quantity - A Holistic Microscopic View of the Sentencing Aid System", Yuanzhao Publishing Co. 2013 Edition, p. 133, p. 139

[8] ICO, Explaining Decisions Made with AI, Mar. 16, 2023, https: / / ico. org. uk / media / for-organisations/ guide-to-data-protection / key-dp-themes/ guidance-on-ai-and-data-protection-2-0. pdf. , last visited on April. 1.

[9] Xu Wei. On the Legal Status of Generative Artificial Intelligence Service Providers and Their Responsibilities - Taking ChatGPT as an Example [J/OL

[10] Luo Hongyang, Li Xianglong. The Ethical Problems in Intelligent Justice and Its Response[J]. Politics and Law Series,2021(01):148-160.

[11] Li Xunhu. Inclusive Regulation of Criminal Justice Artificial Intelligence [J]. Chinese Social Science, 2021(02): 42-62+205.

[12] Ji Weidong. The change of judicial power in the era of artificial intelligence [J]. Oriental Law,2018,No.61(01):125-133.DOI:10.19404/j.cnki.dffx.2018.01.013.

[13] Cui Shixiu. Functional Discussion of Sentencing Assistance System under Intelligent Judicial Field[J]. Journal of Xinjiang University (Philosophy and Social Science Edition), 2022, 50 (04): 30-38.DOI:10.13568/j.cnki.issn1000-2820.2022.04.004.

[14] Zuo Weimin. Some Thoughts on the Prospects of Legal Artificial Intelligence Utilization in China [J]. Tsinghua Law,2018,12(02):108-124.

[15] Paul R. Cohen & Edward A. Feigenbaum eds., The Handbook of Artificial Intelligence, Volume III, William Kaufmann, Inc. 1982, p. 360.

[16] Jia Kai. Research on Artificial Intelligence and Algorithmic Governance [J]. China Administration, 2019, No.403(01): 17-22.DOI:10.19735/j.issn.1006-0863.2019.01.03

[17] Zuo Weimin. Some Thoughts on the Prospects of Legal Artificial Intelligence Utilization in China [J]. Tsinghua Law,2018,12(02):108-124.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202402_10(2).0007

[18] Pedro Dominguez, Fangping Huang. The Ultimate Algorithm:How Machine Learning and Artificial Intelligence Are Reshaping the World[J]. Finance Vertical,2018,No.475(02):102.

[19] Sun Baoxue. Artificial Intelligence Algorithm Ethics and Its Risks[J]. Philosophical Dynamics, 2019 (10): 93-99.

[20] Institute of Natural Language Processing, Harbin Institute of Technology: ChatGPT Research Report, 2023 Edition, p. 24

[21] Xin Zhang. Algorithmic Governance Challenges of Generative Artificial Intelligence and Governance-Based Regulation[J]. Modern Law,2023,45(03):108-123.

[22] Song Lingshan. The Local Experiment of Judicial Intelligent Reform and Its Improvement Path [J]. Southeast Academic,2023(03):225-234.DOI:10.13658/j.cnki.sar.2023.03.019.

[23] Jiangsu Higher People's Court Project Group. Judicial Application of Artificial Intelligence in the Background of Digital Economy[J]. Law application, 2023(05):144-151.

[24] Wu Lili. Breadth + depth: the application of artificial intelligence in criminal trial optimization path to explore[J]. Journal of Jiangxi Radio and Television University, 2021, 23 (02): 34-44.DOI:10.13844/j.cnki.jxddxb.2021.02.005.

[25] Xin Zhang. Algorithmic Governance Challenges of Generative Artificial Intelligence and Governance-Based Regulation[J]. Modern Law,2023,45(03):108-123.

[26] Ge Jinfen. Judge's malfeasance risk and its criminal responsibility in the application of judicial artificial intelligence [J/OL]. Hunan Social Science,2023(03):94-103[2023-06-19].http://kns.cnki.net/kcms/detail/43.1161.C.20230606.0922.020.html

[27] Xu Yannmin,Li Deming. Knowledge mapping analysis of domestic artificial intelligence research[J]. Science and Technology Management Research,2021,41(05):112-119.

[28] Ding Xiaodong. On the Legal Regulation of Algorithms [J]. China Social Science, 2020, No.300 (12): 138-159+203.

[29] License. The Dimension of Science and Technology in Personal Information Governance[J]. Oriental Jurisprudence, 2021,No.83(05):57-68.

[30] Yang, Z. H.. Another possibility of algorithmic transparent realization: interpretable artificial intelligence[J/OL]. Administrative Law Research:1-11