PIONEER
ACADEMIC PUBLISHING LIMITED

Studies in Law and Justice
ISSN 2958-0382
www.pioneerpublisher.com/slj
Volume 4 Number 3 June 2025

When Algorithms Testify: Addressing the Explainability Gap of AI Evidence in Criminal Cases

Yuxin Chen[1]

[1] Law School, Beijing Normal University, Beijing, China
Correspondence: Yuxin Chen, Law School, Beijing Normal University, Beijing, China.

doi:10.56397/SLJ.2025.06.01

Abstract

The expansion of generative artificial intelligence evidence in the field of criminal justice has exposed the structural risks caused by the unexplainability of algorithms. Although existing studies have revealed multiple obstacles, they have not yet touched upon the fundamental crux of the unexplainability of the algorithm. The three predicaments derived from this, namely the disruption of argumentative logic, the loss of focus in the cross-examination process, and the depletion of judicial trust, essentially stem from the subtle tension between the certainty of machine conclusions and their mystery. The solution lies in establishing a transparent evidence generation mechanism, introducing an expert-assisted review system, and setting up traceability rules for training datasets. Through certain system, a dynamic balance is achieved between technological empowerment and procedural justice to prevent the algorithm conclusions from being improperly endowed with transcendent probative force.

Keywords: artificial intelligence, algorithmic black box, criminal evidence

1. Introduction

Since the industrial revolution in the 18th century, machines have gradually replaced human beings in standardized production and driven changes in all areas of society. 21st century breakthroughs in artificial intelligence have given machines the ability to think in complex ways, such as the DENDRAL chemical analysis system, the MYCIN medical diagnosis system, the AlphaGo Go program, and the ChatGPT dialogue system. The breakthrough development of AI in the 21st century has enabled machines with complex thinking ability, such as DENDRAL chemical analysis system, MYCIN medical diagnostic system, AlphaGo program, and ChatGPT dialog system, etc., which can reach or surpass the level of human beings in professional fields. The resulting "machine evidence" is defined by Andrea Roth as machine-generated data and information. The evolution of machine evidence has gone through three generations: the first generation of semi-mechanized evidence requires human-computer collaboration to complete (such as early mechanical records); the second generation of programmed evidence to achieve fully automated generation (such as electronic data generated by standard processes); and the third generation of Generative Artificial

Intelligence (GAI) evidence[1] is generated by AI with in-depth learning capabilities to generate brand new content on their own, such as medical diagnostic reports, self-driving data, and smart interactive content, etc. The uniqueness of GAI evidence is that it generates innovative content based on self-supervised learning from multimodal big data, rather than simply executing a predefined program.[2] The rapid development of generative AI (AIGC) is reshaping the paradigm of criminal proof, and its technological features provide new tools for judicial efficiency as well as complex challenges of legal application and ethical review.

The introduction of all new types of technology in the criminal sphere is expected to be widely controversial, and so is generative AI evidence. There are views that AI can significantly improve the efficiency and accuracy of criminal proof, and advocate releasing the potential of the technology through legal adaptation.[3] Other scholars have pointed out that the U.S. federal courts have adopted a strict scrutinizing stance on AI evidence, requiring that algorithmic principles and training datasets must be disclosed. For example, in the criminal judgment involving face recognition, the court has repeatedly excluded relevant evidence due to the algorithm's "racial bias" problem, and this kind of technological skepticism is of reference significance to China's judicial practice.[4] As around 2020, the emergence of intelligent sentencing assistance systems under the wave of "intelligent justice" triggered widespread controversy. Now, the emergence of generative artificial intelligence evidence has brought a new round of "technological impact", is bound to trigger a fierce collision of different views. Admittedly, in the face of this unknown but closely related to the interests of the new things, to maintain a reverent and cautious attitude is not wrong. However, in the context of the reality of the increasingly tense judicial resources, academics and practitioners are equally eager to generate artificial intelligence evidence as a powerful tool of proof into judicial practice. Therefore, it has become imperative to clarify the key dilemmas of generative AI evidence in the field of criminal proof, deeply analyze its causes, explore feasible solution paths, and give full play to the positive role of generative AI evidence in criminal proof.

2. The Esoteric Characteristics of Algorithms

On the one hand, the criminal proof of generative artificial intelligence evidence appeared its rising demand, on the other hand, the existing generative artificial intelligence evidence exists into the criminal proof system there are many obstacles, this structural contradiction between supply and demand needs to be solved. To resolve this contradiction and break through the barriers to the application of generative artificial intelligence evidence, the key lies in the essence, clear such evidence is difficult to effectively integrate into the criminal proof system is the root cause of the existing rules of evidence review is difficult to respond to the non-interpretable characteristics of generative artificial intelligence evidence.

First of all, generative AI evidence exists a proof potential that cannot be underestimated. Human society is moving into the age of intelligence, and there are more and more scenarios in which AI can be used. As people become more willing to interact with technology and AI-powered devices, the opportunities for machines to monitor human behavior have greatly increased. The resulting machine evidence may strongly contribute to fact-finding[5]. Relying on technological breakthroughs such as face recognition technology, algorithmic recommendation technology, and intelligent trajectory analysis technology, generative artificial intelligence has entered the evidence law field of vision. At the same time, generative artificial intelligence has further demonstrated its unique potential for discovering the truth of the case on the basis of reflecting its possibility of becoming criminal evidence. Specifically, due to the generative artificial intelligence evidence

[1] The review of GAI evidence discussed in the article includes only the review of judgmental material generated by face recognition, autopilot data, etc., and excludes the review of electronic data material such as videos, recordings, etc., that have been falsified through generative AI techniques.

[2] See Xiong Xiaobiao. (2025). The Dilemma of Generative Artificial Intelligence Evidence Determination and the Normative Approach. *Legal Science (Journal of Northwestern University of Political Science and Law)*, *1*(1), pp. 72-93.

[3] See Gao Manjie & Qin Pengbo. (2024). Criminal Legal Risks and Countermeasures of Generative Artificial Intelligence. *China Judgement*, (13), p. 78.

[4] See Ben Li. (2018). Artificial Intelligence in U.S. Judicial Practice: Issues and Challenges. *China Law Review, 2*(2), pp. 54-56.

[5] Sabina Grace. (2022). Artificial Intelligence in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials. Translated by Fan Wen, *Studies in Procedural Law, 26*, p. 145.

based on massive data, and relatively has the characteristics of good accuracy and high flexibility, so it can be expected to play a unique role in the criminal proof of proof. At the same time, the discovery of the truth of the case is the basic value orientation and one of the ultimate goals of criminal procedure activities[1], and the selection of evidence should also follow this principle. At the same time, according to Article 50 of China's Criminal Procedure Law, materials that can be used to prove the facts of the case are evidence. Therefore, no matter from the perspective of potential proof value of generative AI evidence, or from the perspective of the spirit of the evidence law to encourage the adoption of evidence, the material based on generative AI should be included in the category of criminal evidence.

Secondly, although generative artificial intelligence evidence faces many obstacles in the process of entering the field of criminal procedure, its most fundamental contradiction lies in the fact that the current evidence review system is unable to comfortably cope with the inherent tension between the certainty of machine-generated conclusions and the black box of algorithms. Currently, the academic community believes that the obstacles to the application of generative AI evidence can be categorized as follows: first, the concept is unclear. For the criminal justice application of big data-related evidence, the academic community has initially formed three sets of discourse systems of "big data evidence", "artificial intelligence evidence" and "algorithmic evidence", which to a certain extent has hindered the development of relevant theories. To a certain extent, this has hindered the development of related theories[2]. Secondly, it cannot be accommodated by the legal types of evidence. Third, the risk of human rights infringement, i.e., the source data collection process of generative AI evidence may infringe on the public's right to privacy and other rights and interests. Fourth, the reliability question, i.e., whether the conclusions drawn from generative AI evidence can achieve the state of infallibility presented on its surface.

However, while the above questions better summarize the barriers to criminal proof of generative AI evidence, they collectively ignore the core reason behind the problem, namely the non-interpretability of AI, or the "algorithmic black box", which tends to blur the focus of the issue. First, different terminologies do not present insurmountable barriers, and effective dialog between discourses is often possible. Generative artificial intelligence evidence on criminal proof of the obstacle is essentially human distrust of the conclusions of the machine, so whether it is called big data evidence, artificial intelligence evidence, or algorithmic evidence, scholars are concerned about the content of the focus is the same, and will not create a fundamental obstacle to the relevant discussion. Second, the inability to be accommodated by statutory types of evidence is a constant and common pain point in the history of evidence law. On the one hand, the problem is not relevant to the entry of generative AI evidence into criminal procedure, such as accident investigation reports and other materials actively used in practice to prove the facts of the case are likewise not part of the statutory types of evidence; on the other hand, the problem is based on a specific historical background and legislative reasons, and is not a theoretical discussion of the problem can be resolved. Third, the risk of human rights violations is a problem that exists in all evidence collection, and its core lies in the regulation of the means of evidence collection (such as the additional restrictions of the Criminal Procedure Law on the "technical investigation means"), rather than the main problem of generative AI evidence. Finally, the discussion of the reliability of machine conclusions seems to hit the nail on the head, but the essence of the problem diluted the focus of the problem, for two reasons: First, the reliability of generative artificial intelligence evidence is not necessarily inferior to traditional evidence. For example, in the field of face recognition, the accuracy of artificial intelligence has exceeded that of humans. In the segment of identifying criminal suspects, AI may do better. Second, our system of evidence and proof does not require that evidence in criminal procedure is always 100% reliable, which is also unrealistic. Criminal justice has never pinned its hopes on a certain type of evidential material with transcendent probative power falling from the sky and discovering the truth and proving the

[1] Xiaona Wei. (2024). In Defense of Objective Truth. *The Jurist*, (2), p. 144.

[2] Zhang Di. (2023). Algorithmic Evidence in Criminal Proceedings: Concepts, Mechanisms and Their Utilization. *Journal of Henan University (Social Science Edition)*, (3), p. 36.

facts of a case once and for all. In fact, the criminal proof system formally fully recognizes and accepts the limited nature of individual evidence, thus deriving a set of rules of proof to make use of existing evidence in a more scientific way. This is also true for generative AI evidence. Further, all evidence has the potential to be falsified, and this is where the scientific nature of evidence comes in. In practice, for example, the testimony of witnesses, especially those with an interest in the case, and the confessions and defenses of the accused are not always reliable. Even the once blindly superstitious "appraisal opinion" is sometimes wrong. In view of this, the lack of reliability is not a fundamental obstacle to the entry of artificial intelligence into criminal proof.

Looking beyond the appearance of "unreliability" of generative AI evidence and tracing the underlying causes of this impression, it can be found that the fear of generative AI evidence stems more from its inherent mystery, i.e., its non-explainability. Because of this unknown, the criminal proof system can not give its own value equivalent to the effectiveness of the proof, and therefore unlimited worry about whether such evidence to give too much trust. However, from the perspective of the overall development trend of artificial intelligence, the massive corpus, multimodal features and autonomous production of artificial intelligence will continue to deepen, and the technical interpretability will be further weakened. Non-interpretability is essentially a natural attribute of AI products, and interpretability can only be used as a governance orientation. The White Paper on Artificial Intelligence Safety Standardization released in 2023 states that "algorithmic models are becoming increasingly complex, and the goal of interpretability is difficult to achieve", and it is becoming extremely difficult for human beings to understand the large-model AI, and it is currently being explored in the direction of explaining large models with the help of AI. It can be seen that the non-interpretable nature of generative AI is outstanding, but as a "techno-social" paradigm, the social side of the society cannot put itself "on the shelf".[1]

3. Deconstructing the Black Box: Algorithmic Bias in Criminal Justice Systems

Algorithmic black box refers to the non-disclosure and non-transparency of AI algorithms,[2] the non-explainable characteristic may either originate from the algorithmic secrecy behavior for commercial purposes, i.e., the subject concerned does not want the law of the algorithm's operation to be disclosed; or it may also originate from the nature of the algorithm itself, i.e., the algorithmic part is impossible to be interpreted. Specifically, in the field of machine learning, there is a trade-off between model interpretability and model performance (accuracy). Several models (including linear regression and decision trees) have predictive principles that are well understood intuitively, but require a sacrifice in model performance because they produce results with high bias or variance (underfitting: linear models), or are prone to overfitting (tree-based models). More complex models such as integrated models and the recent rapid development of deep learning often produce better predictive performance, but are considered black-box models because it is extremely difficult to explain how these models actually make decisions. The lack of clarity in the decision-making process makes AI face three major interlocking and stepwise obstacles, namely the lack of argumentation, the problem of qualification and the crisis of trust.

Non-interpretability itself does not point to damage, but the risks arising from non-interpretability are directly damaging. Combined with the occurrence of risk, the risk mainly appears in the security of the system itself, the value of the user, and the basic rules of social operation.[3]

3.1 The Absence of Argumentation: Institutional Alienation in the Criminal Proof Process

Criminal proof is a bridge between the evidentiary material and the truth of the case and is processual. In fact, the process of legal argumentation reflects the value of procedural justice.[4] However, the inherent algorithmic black-box characteristics of artificial intelligence

[1] Longjun Jin. (2025). The Uninterpretability of Generative AI and Its Rule of Law Response. *Rule of Law Research*, (2), p. 43.

[2] Feng Xu. (2019). Legal Regulation of the Algorithmic Black Box of Artificial Intelligence — Expanding on the Example of Intelligent Investment Guarantees. *Oriental Law, 6*(6), pp. 78-86.

[3] Longjun Jin. (2025). The Uninterpretability of Generative AI and Its Rule of Law Response. *Rule of Law Research*, (2), p. 47.

[4] See Wei Bin. (2024). Analysis of Legal Arguments for the Explanatory Difficulties of Judicial Artificial Intelligence. *Legal System and Social Development*, (4), pp. 76-92.

technology lead to the generation of artificial intelligence evidence that often manifests itself as "assertive" conclusions, which conflicts with the processual attributes of criminal proof and fails to satisfy the requirements of justice for procedural transparency, which may jeopardize procedural justice.

Specifically, the essence of criminal proof is the process of reconstructing factual knowledge through evidence retrospection. All evidentiary materials used for conviction and sentencing should satisfy the basic attribute of objectivity, and should be presented in the trial in a form that is original, direct and directly reflects the facts of the case, refusing to use processed, value-biased conclusive materials as evidence in the case. This is the best evidence rule and opinion evidence rule jointly constructed evidence jurisprudence foundation. Based on this, the adjudicator through their own professional knowledge and rational judgment to build a bridge from the evidence material to the facts of the case, and ultimately through the argument and evidence of reasoning in the process of determining the evidence and reasoning to fully explain, so that it is figurative and public, which is the fact that the reasoning and the application of reasoning of the law and the premise and foundation of the reasoning of the[1], but also criminal proof of the meaning of the due. However, generative artificial intelligence evidence is the conclusion of the machine deduction, and the general evidence material presents the basic, objective facts of the case is different, its nature and "appraisal opinion" similar, but can not and appraisal opinion of the same trace the conclusion of the deduction of the trajectory and reasoning process. The existence of algorithmic black box leads to the machine can output conclusions, but can not clarify its reasoning process. This flaw makes it difficult for the adjudicator to theoretically rule out other possibilities and substantively meet the standard of proof beyond a reasonable doubt. This inherent logical rupture, so that the generation of artificial intelligence evidence as if "foreign objects", alienated into the public prosecutor's office to pursue the conviction of instrumentalized means. At the same time, due to the generative artificial intelligence evidence of non-interpretability, so that the evidence review is alienated into a simple verification of the results. As shown in the U.S. case of Wisconsin v. Loomis (State v. Loomis), when the algorithmic logic of the COMPAS recidivism risk assessment system could not be disclosed, the judge was forced to shift the focus of the review from the reasoning process to the conclusion probability.[2]

Further, this conflict will lead to a cognitive break in the rationality of judicial decision-making. Procedural justice requires that the adjudicator must show the formation process of evidence through the "adjudication documents fully reasoned". However, the non-interpretability of algorithms leads to a cognitive break between "technical rationality" and "judicial rationality". "When technical decision-making cannot be translated into a human-understandable logic chain, the judge's discretion will be reduced to an endorsement tool for the algorithm's output." This unknowability of the decision-making process essentially violates the "duty to justify" required by procedural justice.

3.2 The Challenge of Cross-Examination: The Loss of the Voice of the Defense

Currently, the common law system and civil law system countries have formed a consensus: the right to confrontation is a fundamental right of the citizens, the right to confrontation is the basic obligation of the state to the citizens,[3] the legitimacy of the basis behind it, including, but not limited to, the right of defense, the authenticity of the government to prevent the abuse of power, and to promote the trust of the state power and a variety of other theories[4]. Article 61 of China's Criminal Procedure Law stipulates that "witness testimony must be examined in court by the public prosecutor, the victim and the defendant, the defense, and both sides and verified before it can be used as the basis for a decision", which likewise clarifies the

[1] Wang Zeshan. (2024). Study on the Reasoning of Evidence Authentication in Criminal Judicial Documents. *Journal of China University of Political Science and Law*, (1), pp. 238-252.

[2] Cited in Jiang Su. (2020). Automated Decision-making, Criminal Justice and the Rule of Law on Arithmetic — Reflections Triggered by the Loomis Case. *Oriental Law Journal*, *3*(3), pp. 76-88.

[3] Fan Chongyi & Wang Guozhong. (2006). A Brief Exploration of the Right of Criminal Defendants to Confront Evidence. *Journal of Henan Province Cadre College of Politics and Law Management*, (5), pp. 49-57.

[4] See Chen Yongsheng. (2005). On the Defense's Right to Examine Evidence in Court. *Law and Business Studies*, (5), pp. 89-96.

defense's right to examination. At the same time, the defendant's right to speak in criminal procedure in China is often crystallized in his right to give evidence in court, and the degree of its realization is closely related to whether it is possible to realize the dual values of discovering the truth of the case and safeguarding human rights. Further, with the de-instrumentalization of the value of criminal procedure, the right to confrontation has evolved in contemporary times into a symbol of procedural justice.

The criminal defendant's right to confrontation is the defendant's right to refute and question the prosecution's evidence in court[1], which can be divided into the right to confrontation and the right to cross-examination. Whether it is the right of confrontation, or the right of cross-examination implies a basic premise: that is, a comprehensive understanding of the prosecution's evidence, that is, the defense has a clear understanding of the prosecution's allegations, the evidence used to prove its allegations and its chain of logical proof. Therefore, most countries have set up a relevant system regarding the discovery of evidence. In the context of generative artificial intelligence evidence, when the algorithmic decision-making process becomes a "technological black box", the defense will encounter structural barriers to the right of defense, the defense often does not have access to all the information about the evidence, and the logical chain from the source of data to the conclusion of the evidence to the facts to be proved is broken. According to Article 13 of the European Parliament's Artificial Intelligence Act, algorithmic interpretability constitutes a prerequisite for the exercise of a party's right to object. Algorithm non-interpretability directly leads to the defense can not be generated for the logic of evidence to put forward effective questioning, in essence, hollowed out the defense to the party's "evidence, questioning, debating" rights bundle, which puts it into a "no evidence can be qualitative" predicament, which directly affects its in the court hearing process. The realization of the right to speak, contrary to the requirements of the principle of "equality of arms", leading to the degradation of the litigation structure from a "confrontation between two creations" to a "monopoly of technical authority".

[1] Wang Xiaohua. (2012). Research on the Right of Criminal Defendants to Confront Evidence in China. Southwest University of Political Science and Law.

3.3 Crisis of Confidence: Lack of Credible Outcomes

Generative AI evidence at the level of argumentation of the logical ring break and the resulting restrictions on the right of the defense to confrontation, directly leading to a crisis of credibility corresponding to the outcome of the referee. Judicial credibility refers to the general knowledge and degree of trust held by the public in the impartiality and authority of the judicial system, the essence of which is the social credibility accumulated by the judicial organs through the fulfillment of their duties by adhering to the legal norms and respecting the objective facts. This concept not only reflects the people's expectations for fairness and justice in justice, but also reveals the core objectives and operating rules of the judicial system, and is a key indicator for assessing the degree of development of the rule of law civilization in a country or region.[2]

The people's trust in the adjudication of a case mainly comes from two dimensions: substantively, whether the adjudication result is correct, i.e., whether the link between the facts of the case and the adjudication result is logical; and procedurally, whether the process of arriving at the adjudication result is flawless, i.e., whether every subject with an interest in the case equally and voluntarily expresses his or her own opinion. When machine conclusions drawn by AI are used as evidence for conviction and sentencing, these two drawbacks are simultaneously revealed by the existence of the algorithmic black box. On the one hand, the lack of argumentation leads to the substantive defects of the adjudication results; on the other hand, the impairment of the defense's right to cross-examination directly weakens the procedural legitimacy of the adjudication results. There is no reason for the public not to fear and question a conclusion that is both procedurally and substantively flawed, while at the same time disposing of the fundamental rights of the accused.

The credibility of judicial decisions is built on the dual basis of substantive legitimacy and procedural legitimacy. At the substantive level, the conclusion of the decision must form a close logical loop with the facts of the case as proven by the evidence; at the procedural level, it is

[2] See Long Zongzhi. (2015). Realistic Factors Affecting Judicial Justice and Judicial Credibility and Their Countermeasures. *Contemporary Jurisprudence*, (3), pp. 3-15.

required that the decision-making process safeguard the right of all parties to participate in the litigation, and that "visible justice" be constructed through equal dialog. This dual legitimacy constitutes the core value of the modern judicial system. However, when the conclusion of the machine generated by artificial intelligence is directly used as the basis for conviction and sentencing, the algorithmic black box will simultaneously dismantle the legitimacy of these two dimensions of the foundation. First, in the entity level, the non-traceable algorithmic reasoning makes the factual determination reduced to "technical arbitration", the correlation between the adjudication results and the facts of the case lost verifiable basis, resulting in the substance of the justice of the entity is reduced to a probabilistic judgment; Secondly, in the procedural level, the non-explanatory algorithmic decision-making essentially deprives the defense of the right to effective questioning, and the foundation of procedural justice has been hollowed out. The foundation of justice has been hollowed out. The loss of this dual legitimacy will lead to a serious crisis in the rule of law: a decision that is neither substantively correct nor procedurally participatory is able to dispose of the fundamental rights of citizens. This potential risk of "algorithmic tyranny" will not only trigger public skepticism about the adjudication of individual cases, but will also fundamentally shake the trust in the judicial system. "Justice must not only be realized, but also realized in a visible way."[1] Otherwise, the foundation of the rule of law edifice will be in danger of collapsing.

4. Deconstructing and Overcoming Algorithmic Black Boxes in Evidentiary Procedures

As mentioned earlier, the fundamental obstacle to generative AI evidence in criminal proof activities does not lie in its lack of reliability, but in the subjects of its non-interpretability of the "fear", the seemingly "unquestionable" face of science and its internal operating logic. There is a strong tension between the seemingly "unquestionable" face of science and the mysterious color of its inner logic of operation. In essence, the value of technical agnosticism and judicial refutable conflict. The process of criminal proof requires that the evidence must be interpretable, questionable, can be overturned open characteristics, and algorithmic black box created by the "technological leviathan" is eroding the procedural justice depends on the existence of the system foundation. In view of this, this structural contradiction determines the key to the problem is to re-examine the ability of generative artificial intelligence evidence boundaries, the establishment of a targeted review system, so that the subjects get to know, question, overthrow the conclusions of the machine's right and ability, rather than to ensure that the generative artificial intelligence evidence of the "one hundred percent" accuracy, this is the problem of the technical field, rather than the black box of algorithms created "technological leviathan" is eroding the system on which procedural justice is based. This is a technical issue, not a judicial one. To break the myth of certainty of generative artificial intelligence evidence in the field of criminal proof, we can establish a corresponding challenge system from the content and method of review. Generative artificial intelligence evidence that has been effectively challenged and incorporated into the basis for a final decision meets the requirements of procedural justice and can also respond to the public's expectations for justice.

4.1 Systematic Review: Transparency in the Mechanism of Evidence Generation

It is true that, theoretically, there is an inherent "cognitive blind spot" in the process of generating artificial intelligence evidence, and this technical limitation should not be a reason for abandoning regulation. On the contrary, we should uphold the principle of "limited transparency", within the boundaries of technical possibilities to actively promote the transparency of the relevant content, the part that can be reviewed to establish a perfect system. First of all, the most front-end and fundamental issue is the raw data on which the machine's conclusions are based. It needs to face at least twofold challenges of authenticity and comprehensiveness. Both the quantity and quality of data are directly related to the accuracy of the final conclusion. Secondly, on the middle end, the main focus is on the review of algorithms. Although the research on algorithmic review is not fully developed at present, there is a more mature consensus on the way to review scientific evidence of the same

[1] Alfred Thompson. (2011). *Denning: Due Process of Law*. Translated by Li Keqiang and others, Law Press.

nature: in 1923, the U.S. Court of Appeals for the District of Columbia federal appeals court in the case of Frye v. United Stated laid down a standard for measuring the reliability of expert testimony, which held that the court accepts a recognized scientific theory. In Frye v. United Stated, the U.S. Court of Appeals for the District of Columbia set forth the standard for measuring the reliability of expert testimony, holding that a court will accept expert testimony derived from an accepted scientific theory or scientific discovery, provided that what is deductively inferred therefrom is sufficiently well grounded and generally recognized in the field of which it is a part[1], i.e., the "Frye standard". Finally, at the end, there should be a qualification process for those who operate AI systems. Reference can be made to the rules for qualifying personnel in the appraisal review system.

4.2 Equally Armed: Expert Auxiliary Assistance

Aiming at the artificial intelligence into the criminal procedure, the resulting impact of the prosecution and defense power contrast, scholars put forward the concept of "evidence bias"[2] to argue that the right to confrontation has been eroded to a certain extent due to the introduction of generative artificial intelligence evidence. Undoubtedly, to give the weaker side of the stronger force clamping is the most direct means of solving the problem of "bias in".

In the application scenario of generative AI evidence, the imbalance in the ability of the defense party is centered on the structural weakness of the technical nature of the evidence power. Due to the high degree of asymmetry in algorithmic information, although the defense is generally involved in the AI service ecosystem, it faces the following dilemmas: first, it lacks accessibility to the underlying algorithmic architecture and model training logic of generative evidence; and second, it has a blind spot to the technical paths by which personal data are extracted, labeled, and embedded in decision-making systems. Even if the parties are able to recognize the importance of collecting generative AI evidence, they are often caught in a double passivity due to the black box effect of algorithms and the lack of professional dialogue capabilities: it is difficult to analyze the logic and causal chain of generating AI evidence, and they are also unable to effectively challenge the semantic completeness of its output, which ultimately results in the alienation of technological empowerment into the ability to fight against the litigation bias. Based on this, the permission to seek professional help and as an effective basis for questioning the generation of artificial intelligence evidence is the context of the prosecution and defense of the two sides of the basic requirements of equal arms. In view of the commonality between generative AI evidence and appraisal opinions in many aspects[3], reference can be made to the setting of expert assistants in the current appraisal opinion system in China.

Specifically, through a reasonable definition of the qualifications of the expert supporter, improve the rights and obligations of the expert supporter to participate in the litigation, and clarify the litigation status of the expert supporter and the attributes of his opinions, etc., to ensure that the role of the expert supporter can be given full play to, and to ensure that the rights and interests of the person being prosecuted can be effectively safeguarded. At the same time, the content of the expert auxiliary's examination of the generative artificial intelligence evidence can refer to the interpretation made by the U.S. Federal Supreme Court in 1995 in the case of Daubert v. Merrell Dow Pharmaceuticals. Inc. on the issue of scientific standard of scientific evidence. The decision held that the reliability of expert testimony should be judged from four aspects: (1) whether the scientific theory and scientific methodology relied upon to form the expert testimony can be repeatedly tested; (2) whether the scientific theory and scientific methodology used to form the expert testimony has been peer-reviewed or has been published; (3) whether the known or potential rate of error concerning the theory is acceptable; (4) whether the theory and research methodology guiding the theory in question are relevant to the case; and (5) whether the scientific standard for scientific evidence is acceptable. Methodology and research methods are accepted by the

[1] See Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

[2] See Zhang Qi and Fan Yunhui. (2024). The Risks of Artificial Intelligence Evidence in Judicial Activities and Legal Responses. *Journal of Henan Finance and Economics College (Philosophy and Social Science Edition)*, (1), pp. 35-40.

[3] See Zheng Fei, Ma Guoyang. (2022). The Triple Dilemma and the Way Out of the Application of Big Data Evidence. *Journal of Chongqing University (Social Science Edition)*, (3), pp. 207-218.

relevant scientific community and the extent of that acceptance[1].

4.3 Practical Verification: Historical Data Disclosure

The transparency of the evidence generation mechanism and the support of expert assistants to the defense try to guarantee the possibility of challenging the evidence of generative AI at the institutional level, and to convey to the public the notion that the adjudication results are logically correct. However, it cannot be ignored that due to the non-interpretability of artificial intelligence, there will always be a part of the path of machine conclusion proof that is in the gray area. Therefore, in line with the spirit of the evidence law, "strengthen the evidence rule", the proof of the existence of flawed evidence should be strengthened. Given that the underlying logic of generative artificial intelligence evidence comes from the science of machine learning, the historical accuracy of the algorithms on which it is based can be publicized for practical verification, thereby enhancing the credibility of the results.

Historical data disclosure has a number of advantages. First, the approach helps the general public make clear judgments. "Mathematical certainty is absolute." Accuracy as a number can turn ambiguity into clarity and help the public make judgments. Second, the approach is easy to understand. Size judgments are simpler. The disclosure of historical data is more publicized than the disclosure of algorithms, and has a more significant effect on improving the credibility of adjudication results.

5. Conclusion

In his book *Technopoly*, communication scholar Neil Bozeman asserts that every new technology is both a burden and a gift, not an either/or outcome, but a product of both advantages and disadvantages. Generative AI evidence may be a technological breakthrough or a Pandora's box that has already been opened in criminal proof activities. To some extent, the fear of the latter stems from the opacity of the algorithm. Humans are naturally afraid of the unknown, for humans, the algorithm operates like a "black box" — we are responsible for providing data, models and architecture, the algorithm is responsible for giving the answer, while the middle of the operation process is only carried out in the dark. This kind of cooperation seems to bring us great convenience, but the problem is that if the operation of the algorithm is not monitorable and unexplainable, it will lead to the fact that human beings can't really understand the algorithm, and they can't control the algorithm effectively, and thus can't foresee and solve the problems that the algorithm may bring[2].

In light of this, the obstacles to generative AI evidence in criminal proof should be clarified by focusing on its non-interpretability. Further, criminal justice does not need to be committed to the technological breakthroughs therein, but should focus on the subtle tension between the certainty of the machine's conclusions and its mysteriousness, so that generative AI evidence will always have the possibility of being challenged, and as far as possible to ensure that it will not be given a probative value that far exceeds its own proper value. This is the necessary path for generative AI evidence to gain trust and ultimately smooth criminal procedure.

References

Alfred Thompson. (2011). *Denning: Due Process of Law*. Translated by Li Keqiang and others, Law Press.

Ben Li. (2018). Artificial Intelligence in U.S. Judicial Practice: Issues and Challenges. *China Law Review, 2*(2), pp. 54-56.

Chen Yongsheng. (2005). On the Defense's Right to Examine Evidence in Court. *Law and Business Studies*, (5), pp. 89-96.

Fan Chongyi & Wang Guozhong. (2006). A Brief Exploration of the Right of Criminal Defendants to Confront Evidence. *Journal of Henan Province Cadre College of Politics and Law Management*, (5), pp. 49-57.

Feng Xu. (2019). Legal Regulation of the Algorithmic Black Box of Artificial Intelligence — Expanding on the Example of Intelligent Investment Guarantees. *Oriental Law, 6*(6), pp. 78-86.

Gao Manjie & Qin Pengbo. (2024). Criminal Legal Risks and Countermeasures of Generative Artificial Intelligence. *China Judgement*, (13), p. 78.

[1] See Daubert v. Merrell Dow Pharms., 509 U.S. 579, 594 (1993).

[2] Wang Huanchao. (June 13, 2019). How to Make Algorithms Explain Why They "Algorithmically Discriminate"?. Tencent Research Institute.

Jiang Su. (2020). Automated Decision-making, Criminal Justice and the Rule of Law on Arithmetic — Reflections Triggered by the Loomis Case. *Oriental Law Journal*, *3*(3), pp. 76-88.

Long Zongzhi. (2015). Realistic Factors Affecting Judicial Justice and Judicial Credibility and Their Countermeasures. *Contemporary Jurisprudence*, (3), pp. 3-15.

Longjun Jin. (2025). The Uninterpretability of Generative AI and Its Rule of Law Response. *Rule of Law Research*, (2), p. 43, 47.

Sabina Grace. (2022). Artificial Intelligence in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials. Translated by Fan Wen, *Studies in Procedural Law, 26*, p. 145.

Wang Huanchao. (June 13, 2019). How to Make Algorithms Explain Why They "Algorithmically Discriminate"?. Tencent Research Institute.

Wang Xiaohua. (2012). Research on the Right of Criminal Defendants to Confront Evidence in China. Southwest University of Political Science and Law.

Wang Zeshan. (2024). Study on the Reasoning of Evidence Authentication in Criminal Judicial Documents. *Journal of China University of Political Science and Law*, (1), pp. 238-252.

Wei Bin. (2024). Analysis of Legal Arguments for the Explanatory Difficulties of Judicial Artificial Intelligence. *Legal System and Social Development*, (4), pp. 76-92.

Xiaona Wei. (2024). In Defense of Objective Truth. *The Jurist*, (2), p. 144.

Xiong Xiaobiao. (2025). The Dilemma of Generative Artificial Intelligence Evidence Determination and the Normative Approach. *Legal Science (Journal of Northwestern University of Political Science and Law)*, *1*(1), pp. 72-93.

Zhang Di. (2023). Algorithmic Evidence in Criminal Proceedings: Concepts, Mechanisms and Their Utilization. *Journal of Henan University (Social Science Edition)*, (3), p. 36.

Zhang Qi and Fan Yunhui. (2024). The Risks of Artificial Intelligence Evidence in Judicial Activities and Legal Responses. *Journal of Henan Finance and Economics College (Philosophy and Social Science Edition)*, (1), pp. 35-40.

Zheng Fei, Ma Guoyang. (2022). The Triple Dilemma and the Way Out of the Application of Big Data Evidence. *Journal of Chongqing University (Social Science Edition)*, (3), pp. 207-218.