Proceeding of ICGPSH 2025 Symposium: Exploring Spelling Evolution and Educational Approaches to Dyslexia
DOI: 10.54254/2753-7048/2025.HT25452

Reliability Risks and Regulation of Algorithmic Evidence in Criminal Certification

Weitong Zhu

Department of Law, North China Electric Power University, Beijing, China
zwt040817@163.com

Abstract. Algorithmic evidence is the product of the deep integration of algorithm and criminal justice. In criminal proof, the reliability risk of algorithmic evidence mainly focuses on three levels: at the level of legitimacy, the definition of evidence qualification is vague, the data collection is illegal, and the algorithmic technology application program is not standardized; at the level of authenticity, the data source is biased, the algorithm is discriminatory, and the black box dilemma; and at the level of relevance, the causality is insufficient, and the predictive conclusions are easy to be mechanically applied. In this regard, the reliability of should be guaranteed through the legalization of evidence types and the traceability of data sources; the establishment of algorithmic technology visualization system and third-party verification mechanism to achieve technical reliability; and the clear logical connection between the conclusion and the facts to be proved and the setting up of manual review nodes to strengthen the logical reliability. It provides institutional and technical support for the innovation of judicial proof system in the digital era.

Keywords: algorithmic, evidence, evidence review, technology governance

1. Introduction

"Algorithm is the law and power in the network era." [1] With the development of digital technology, algorithmic evidence has gradually become one of the important forms of evidence in the field of criminal proof. In cases of cybercrime, financial fraud and other cases, algorithmic evidence tracks criminal suspects through data fusion analysis, behavioral pattern mining and other technical means, or becomes a key breakthrough in case detection. Especially in the detection of new and complex crimes, this type of evidence has significant advantages. In the case of Feng Mou and other seven people's occupation, for example, prosecutors introduced a third-party agency, the use of blockchain analysis, big data comparison and other algorithmic techniques to build a model of the correlation between the virtual currency address and the reality of the identity of the results of the technical analysis into the algorithmic evidence and the formation of a complete chain of evidence, a clear presentation of the path of money-laundering to help the detection of the case.

However, at the same time, the use of algorithmic evidence also has potential risks. Such as the procuratorial authorities "introduce third-party agencies" to establish the algorithmic model and its analysis results, in the current evidence system should be categorized as what type of evidence, its probative power and reliability can meet the standards of criminal proof and so on. Existing research

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2025.HT25452

mainly focuses on the conceptual definition and legal attributes of algorithmic evidence, the technical risks of algorithmic evidence, and the discussion of macro-regulation path.On the basis of the above research, this paper focuses on the core link of criminal proof - evidence "three" review, and comprehensively evaluates its overall reliability, and puts forward a set of systematic solutions covering the improvement of rules, technical governance and rights protection, with a view to providing theoretical support for the standardized application of algorithmic evidence and providing theoretical support for the development and application of algorithmic evidence. A set of systematic solutions covering the improvement of rules, technical governance and rights protection is proposed, with a view to providing theoretical support and technical route for the standardized application of algorithmic evidence.

2. Connotation of algorithmic evidence

2.1. Concept of algorithmic evidence

The definition of "algorithmic evidence" must begin with clarifying the concept of an "algorithm." Algorithms originally belonged to the field of mathematics and computational sciences, i.e., "any well-defined computational step can be called an algorithm, which accepts a value or set of values as input and outputs a value or set of values" [2]. With the development of digital technology, the concept of "algorithm" has been gradually extended to the field of social sciences. Scholars define algorithm as "the decision-making of human-machine interaction, i.e., a set of mechanisms by which humans make decisions through code setting, data computation and automated machine judgment." [3]

In the judicial field, algorithms, by virtue of the efficient processing of complex data, the role of judicial decision-making is increasingly prominent, some scholars believe that "algorithmic evidence is the integration of massive case information data computation, the formation of structured, information-based case scenarios automatically interpreted and automated reasoning results of the material." [4] Other scholars believe that "algorithmic evidence mainly refers to evidence based on algorithms, including both algorithms directly related to big data in the course of the case, as well as evidence after the algorithmic analysis of big data." [5] The first point of view reveals the technical generation process of algorithmic evidence, and emphasizes the results of "structuring and informatization", which is in line with the technical characteristics of evidence in the digital era. However, its definition does not clearly distinguish between the "algorithm itself" and "algorithm analysis results" of the evidence attributes, may ignore the independent value of the algorithm in the generation of evidence (such as the algorithm logic itself as the object of proof); the second viewpoint broadens the scope of algorithmic evidence, both including The second view broadens the scope of algorithmic evidence to include both the algorithmic technology itself, such as the algorithmic program and logic used in the case, and the results of the algorithmic analysis, such as the data report, which more comprehensively covers the diversified scenarios in judicial practice. However, the expression "generated based on algorithms" in the concept is relatively vague, and does not clarify the status of algorithms as the "core tool" in the generation of evidence, which makes it difficult to reflect the technical characteristics of algorithmic evidence that distinguish it from traditional electronic data.

Combined with the above definition and the theory of criminal evidence - including evidence to meet the authenticity, relevance, legitimacy attributes, "beyond a reasonable doubt" standard of proof, the author believes that, algorithmic evidence is to algorithmic technology as the core tool, through the massive case of structured analysis and integration of data. Algorithmic evidence is to

DOI: 10.54254/2753-7048/2025.HT25452

use algorithmic technology as the core tool, through the structured analysis and integration of massive case data, the formation of part of the explainable, verifiable automated reasoning conclusions or technologically generated materials. The definition emphasizes the positioning of algorithmic technology as the "core tool", distinguishes it from the traditional evidence mode relying on manual analysis, and highlights the irreplaceable role of algorithm in the formation of this type of evidence; at the same time, on the basis of the theory of "algorithmic ontology as evidence" in the second point of view, the definition of "algorithmic ontology as evidence" in the second point of view has been adopted. At the same time, on the basis of absorbing the theory of "algorithmic ontology as evidence" in the second point of view, the expression "technologically generated material" avoids the problem of conceptual broadening; finally, the definition adds the attribute of "partially explainable and verifiable", recognizing that the process of generating algorithmic evidence is subject to the constraints of technological capabilities, while retaining its automation. Finally, the definition adds the attribute of "partially explainable and verifiable", recognizing that the process of generating algorithmic evidence is subject to the constraints of technological capability, retaining the advantage of automation while avoiding excessive reliance on the absolute correctness of the algorithm, which is more in line with the requirement of "verification of doubtful evidence" in judicial practice. Eventually, algorithmic evidence is defined as a kind of "limited transparent but controllable" technical evidence, which is more suitable for the judicial practice in the digital era on the scientific, efficiency and fairness of the comprehensive needs.

2.2. Characteristics of algorithmic evidence

2.2.1. Technical dependence

The generation of algorithmic evidence is highly dependent on preset or self-learning algorithmic rules, such as cluster analysis, association mapping, decision tree, etc., and its technical kernel is embodied in mathematical computational logic; at the same time, the generation of evidence conclusions is based on the training and optimization of the case data, and the quantity and quality of the data directly affects the reliability of the conclusions; and the algorithm's prediction results will vary with the technology's iteration and update.

2.2.2. Limited interpretability and verifiability

The interpretability of an algorithm's evidence refers to the ability of its decision logic to be presented in a human-understandable way that clarifies the path of reasoning. For example, in traffic collision cases, accident responsibility determination algorithms need to explain how to calculate the proportion of responsibility through parameters such as vehicle speed, braking distance, collision angle, and so on, instead of only outputting conclusions such as "full responsibility" and "main responsibility". However, due to the complexity of the algorithm, it is not possible to completely restore all the details. For example, face recognition algorithms can explain the calculation logic of "similarity matching score", but it is difficult to trace the activation mechanism of neurons in each layer of the neural network. Verifiability means that the conclusions of the algorithmic evidence can be reproduced or tested by technical means to confirm its consistency and accuracy. This feature can be demonstrated through methods such as blockchain deposits to ensure data integrity and allowing third-party organizations to reproduce the algorithmic process.

DOI: 10.54254/2753-7048/2025.HT25452

2.2.3. Predictability

Although algorithmic evidence has different forms, its overall operation mechanism focuses more on predictive and inferential judgment. Specifically, this feature is mainly embodied in the analysis of relevant data through algorithmic models to identify abnormal behavioral patterns, and then predict potential criminal risks or facts, generating criminal risk reports or behavioral inference conclusions. From a technical perspective, this feature relies on the statistical modeling and machine learning capabilities of algorithms, the essence of which is to transform correlations between data into certain factual inferences.

3. Current status and potential risks of the application of algorithmic evidence in criminal proof

3.1. Application status of algorithmic evidence in criminal proof

The author found that algorithmic evidence has been more widely used in judicial practice by searching for relevant judging documents and combining it with interviews with judges, and obtained 312 effective samples.

After analysis, it is concluded that: The application of algorithmic evidence is mainly concentrated in three types of crimes, namely, obstructing social management (39.1%), destroying the socialist market economic order (38.8%), and infringing on property (14.1%).

In addition, technology carriers are also evolving (see Table 1). From 2019-2021, when basic data analysis and intelligent warning platforms accounted for the absolute proportion, to the last three years, emerging technologies such as blockchain depository are developing at a high speed, highlighting the growing demand for credibility deposit in the judicial field; intelligent warning platforms are also undergoing continuous iteration and have become the dominant technology; and AI models may be limited in use due to their insufficient explanatory nature.

Table 1. Distribution of technology carriers

Type of technology	Percentage (2019-2021)	Percentage (2022-2025)	Typical system
Basic data analysis	68.3%	32.1%	Excel macro tool, SQL query
Intelligent Early Warning Platform	21.5%	45.7%	Public Security "Wind Tunnel System", Ant Ant Shield
Blockchain Depository	1.2%	15.9%	Hangzhou "Judicial Chain", Beijing Tianping Chain
Artificial Intelligence Models	9.0%	6.3%	Gambling money identification model

Through interviews with judges, it was found that, firstly, judges are generally cautious when utilizing algorithmic evidence. At present, algorithmic evidence mainly plays the function of auxiliary proof, and can hardly be used as the basis for determining the facts of a case on its own. This phenomenon corroborates with the aforementioned refereeing documents' avoidance of the use of the concept of "algorithmic evidence". In cases where the objective chain of evidence is relatively complete, algorithmic evidence is easily adopted as supplementary evidence if it can form a corroborative relationship with traditional evidence; while in areas involving subjective judgment or ethical disputes, it is more reliant on rules of thumb and free evidence. Taking the report on the

DOI: 10.54254/2753-7048/2025.HT25452

results of the analysis and research of funds in economic crimes as an example, the judge believes that if it is used as evidence, it should be ensured that the samples are true and complete, the analysis method is scientific and reasonable, and the conclusion is clear and error-free, in particular, it should be corroborated with other evidence to eliminate reasonable doubt. In addition, there are differences in the acceptance of algorithmic evidence by different types of courts. The Internet Court and other specialized trial institutions have a higher acceptance of new types of evidence, and tend to explore the rules through individual cases, such as the "Ten Typical Cases of Data Algorithms" issued by the Beijing Internet Court. When dealing with complex technical evidence, grassroots courts may rely more on traditional forms of evidence due to a lack of technical understanding or limited resources for review.

3.2. Reliability of algorithmic evidence in criminal proof risks

From the empirical study of, a series of risks have also arisen in the process of integrating algorithmic evidence with traditional evidence rules. The following from the discusses the lack of reliability of algorithmic evidence perspectives of legitimacy, authenticity and relevance.

3.2.1. Legitimacy risks

First. legitimacy of evidence form Doubtful.Algorithmic evidence first faces the dilemma of vague definition of evidence qualification. Currently, the legal positioning of algorithmic evidence in academia and practice has not yet formed a unified cognition, and whether it belongs to electronic data, expert opinion or independent evidence type is still controversial, which leads to a lack of clear guidelines for evidence review in judicial practice.

Second, data collection procedures are not sufficiently legalized.The process of data collection may violate privacy and other personal rights. In the investigation of the crime of violating citizens' personal rights, the algorithm's deep mining of communication records and social data may break the principle of "necessary and minimal", resulting in in citizens' freedom of communication and private spherein Personal Information Protection Law China's excessive interference. The algorithmic formed by evidence not only such illegal collection may be excluded due to procedural violations, but even may trigger disputes over the legality of investigative behavior or the state compensation risks.

Third, the lack of legitimacy in the process application of algorithms. On the one hand, the use of algorithmic tools lacks transparent rules. For example, whether the algorithm report submitted as evidence in the trial needs to disclose in advance to the defense the algorithm principle, data source, as well as the specific disclosure procedure and method is not yet clear. Although China's "Provisions on the Administration of Algorithm-generated Recommendations for Internet Information Services" [6]requires the disclosure of algorithm principles, enterprises often use technical confidentiality as the reason, only provide general "recommendation logic description" or directly refuse to explain, resulting in the exercise of the right to cross-examine is limited the defense. On the other hand, the subject of application of algorithmic evidence presents an unequal state concentrated in the prosecution. The prosecution holds massive data resources and advanced technical tools by virtue of public power, while the defense is in in data acquisition and analysis a disadvantaged position structural inequality, due to making the main body of evidence mainly concentrated algorithmic in the prosecution, which has caused a certain impact on the principle of equality of status between the prosecution and the defense and the system of disclosure of evidence.

DOI: 10.54254/2753-7048/2025.HT25452

3.2.2. Authenticity risk

The authenticity of algorithmic evidence relies on the objective credibility of data and algorithms, but at present there are still a variety of risks at the technical level.

First, data authenticity is difficult to guarantee. First, the data collection process may be caused by technical failures, human intervention and other factors, resulting in a single source of data and unrepresentative samples and other problems. This makes the algorithmic evidence generated based on these data unable to accurately reflect the full picture of the case, and produce biased judgments on criminal behavior in different regions or platforms.

Second, algorithmic evidence also has high requirements for updating data. The social environment, criminal means and behavioral patterns are constantly changing, and if the data on which the algorithmic evidence relies is not updated in a timely manner, and is still based on past data for training and analysis, it may produce data bias, and be unable to accurately identify and judge certain new types of criminal behavior.

Third, it is difficult to trace the authenticity of algorithm operation. Algorithmic discrimination refers to discrimination caused by differential treatment of risky individuals. The basic types of algorithmic discrimination include biased agency, feature selection, and big data (algorithmic) "killing" [7].For example, the U.S. COMPAS system, which relies too heavily on arrest records rather than actual crime rates, results in black defendants being incorrectly labeled as high-risk at twice the rate of whites [8].Such belonging to algorithmic discrimination feature selection, although it is not yet generally visible in our country, with the expansion of algorithmic application scenarios, its potential risk cannot be ignored.

Algorithmic black box and interpretability dilemma In terms of algorithmic black box, "algorithmic black box" refers to the lack of transparency and due process of algorithms, This characteristic is in fundamental conflict with the principle of judicial openness. The operating logic of algorithmic models often involves complex mathematical operations and technical encryption, and there are high technical barriers that make it difficult for judicial officers to validate the algorithms solely on the basis of their own experience and knowledge. It may not be possible to meet the standard of proof of "beyond reasonable doubt". Despite the emergence of new explainable artificial intelligence (XAI) technology [9], it is difficult to fully decipher the logic of complex models, and there is a risk of secondary manipulation of the results of the explanation. The judiciary is unable to trace the authenticity of the algorithm's operation through technical means, and this may lead to technological blindness, which to a certain extent undermines the principle of presumption of innocence and the system of free heart evidence.

3.2.3. Relevance risk

The algorithmic evidence and the facts to be proved The logical correlation may due to between be in a dilemma of proof technical characteristicsits. According to the requirements of evidence review, the technical materials or output conclusions generated by algorithms must be substantially related to the facts to be proved and exclude the interference of other irrelevant factors. In practice, algorithms often incorrectly correlate accidentally related or weakly related data to causality, and this "correlation instead of causality" logic may lead to investigative authorities mistakenly including innocent groups in the scope of investigation or incorrectly locking in the subject of the crime.

DOI: 10.54254/2753-7048/2025.HT25452

4. Algorithmic evidence in the criminal proof of the path of perfection

In the face of the reliability risk of algorithmic evidence in criminal proof, there is an urgent need to clarify in that algorithmic evidence needs to satisfy the three requirements of "legitimate data sources, verifiable algorithmic logic, and sufficient correlation of conclusions"judicial practice, and to build a normative system that unifies risk prevention and control and the good use of technology.

4.1. Strengthening legitimacy review: standardizing the reliability of the application process of algorithmic evidence

4.1.1. Clarify the specific types of evidence attributed to algorithmic evidence

The legalization of the type of evidence is the logical starting point and institutional premise for safeguarding the legality of algorithmic evidence. Currently, there are three main views on the categorization of algorithmic evidence, the first is the independent evidence type, based on its technical complexity, advocating algorithmic evidence as a new class of evidence [10]; the second is the subclass of electronic data, because of its carrier relies on the electronic form, it can be categorized into the category of electronic data, but need to formulate a special review rules [11]; and finally, the last is a technical report: analogous to the appraisal opinion, the requirement of algorithmic generation process needs to be certified by a third party technical agency. process needs to be certified by a third-party technical organization.

Different doctrines have their own merits, the author believes that, due to the algorithmic evidence in the collection, review, application and other aspects of the existing types of evidence with the existence of large differences, it is difficult to be directly categorized as a class of evidence, our law will be stipulated as an emerging independent type of evidence is conducive to a better regulation of it.

4.1.2. Strengthen the control of data source legality

Establish rules for reviewing the legitimacy of data sources, and ensure that the collection of algorithmic evidence complies with the legitimacy basis of Article 13 of the Personal Information Protection Law by verifying the authorization documents for data collection, such as users' informed consent and administrative license approvals, etc., and prohibit the acquisition of data by means of unlawful intrusion and excessive collection.

4.1.3. Improve the algorithm filing system

With regard to the requirement of "algorithm filing" in Article 24 of the Administrative Provisions on Algorithm Recommendation for Internet Information Services, it is necessary to refine the content of the filing in the field of criminal proceedings: add the "Commitment for Judicial Application", make clear the obligation of the technology provider to cooperate with the examination of evidence in the litigation, and refrain from refusing to provide evidence on the grounds of commercial confidentiality. Refuse to provide basic technical information that affects the qualification of evidence on the grounds of commercial secrets; for algorithm evidence that fails to fulfill the filing obligation or the filing information is false, directly exclude its qualification as evidence.

DOI: 10.54254/2753-7048/2025.HT25452

4.2. Authenticity review: ensuring the of algorithmic evidence technical reliability

4.2.1. Establish a negative list system

Establish a "negative list of algorithmic evidence" and explicitly include crime prediction and risk assessment algorithms that have not passed reliability certification, as well as algorithms that are discriminatory in nature, in the scope of prohibited use, and strictly prohibit their use as a basis for conviction and sentencing. In this way, eliminate illegal evidence from the application of the source,and avoid the recurrence of "COMPAS-style" systematic bias in China's judicial practice.

4.2.2. Improve technical visualization and logical verification techniques

Develop technical assistance systems, such as algorithmic evidence visualization tools, to transform complex algorithmic logic into flow charts, decision trees or dynamic simulation demonstrations, so as to facilitate understanding by non-technical personnel. At the same time, establish a database of algorithms for similar cases, which can compare the application logic of algorithms in similar cases, and identify data deviations or potential discriminatory patterns through statistical analysis. Such technical means not only reduce the understanding threshold of non-technical personnel, but also provide a verifiable technical path for the review of the reliability of algorithmic evidence through transparent logical display and historical data comparison, effectively reducing the authenticity determination dilemma due to the black box characteristics of the algorithm.

4.3. Relevance review: strengthening the reliability of the logical connection between algorithmic evidence and facts to be proved

4.3.1. Clarify the sufficient elements of relevance

Clarify that algorithmic evidence must meet the requirement of "sufficient relevance", that is, algorithmic evidence must form a direct and strict logical chain with the facts of the crime, the suspect's behavior and the facts to be proved. Specifically, the conclusion of the algorithmic analysis and its technical generation of materials can not only stay in the level of risk assessment or probability prediction, but should be traced back through the data, model validation and other ways to ensure that the derivation process and the case of the core controversy has a substantial correlation, and must be corroborated with the traditional physical evidence, witness testimony, defendant statements and other types of legal evidence, together constitute a complete system of evidence to exclude reasonable doubt.

4.3.2. Setting up manual review nodes and prohibiting complete reliance on automated decision-making

Evidence relevance assessment algorithms are high-risk algorithms, and the decision-making results of such algorithms directly affect the rights and fate of suspects and defendants. Therefore, in the key aspects of algorithmic decision-making, manual review should be introduced to examine whether the algorithm has over-expanded the scope of relevance or ignored the specificity of individual cases, and to ensure that the conclusions are in line with judicial logic and rules of thumb. Reviewers should be composed of judges and prosecutors with legal expertise and rich judicial experience, as well as professionals familiar with algorithmic technology, to form a cross-disciplinary review team. During the review process, the reviewers should not only review the

DOI: 10.54254/2753-7048/2025.HT25452

output results of the algorithm, but also trace the input data, model training process and decision-making logic of the algorithm to ensure that the algorithm complies with the legal provisions and judicial justice requirements in data processing, model selection and rule application. This not only makes use of the algorithm's efficient advantage in data processing and analysis, but also avoids judicial injustice caused by algorithmic defects, realizing the organic integration of technology and law in the field of criminal procedure.

5. Conclusion

The application of algorithmic evidence is like a double-edged sword, both for criminal proof to bring technical innovation opportunities, but also the risk of impact on procedural justice. The technical dependence of algorithmic evidence determines that it can't run independently from the judicial rules, algorithmic black box, data bias and rights protection imbalance and other problems, in essence, is the conflict between technical tools and judicial value performance. In this regard, it is necessary to establish a governance mechanism for algorithmic evidence with the synergistic evolution of legal rules and technical standards. By establishing algorithmic evidence as an independent evidence category, refining the review rules, and developing visual validation tools,risks can be effectively controlled while preserving technical effectiveness; by setting up mandatory procedures for manual review and introducing a technical defense assistance system,the defendant's right to confrontation can be safeguarded, and erosion of the judge's free evaluation of evidence by "algorithmic dictatorship" can be prevented.

References

[1] Jamie Saskander, The Power of Algorithms: How Humans Survive Together, translated by Dabai Li, Beijing Daily Press 2022, p. 297.
[2] Thomas Coleman, Charles Reasonson, Ronald Levist, Clifford Stein, etc.: Introduction to Algorithms, translated by Yin Jianping, etc., Mechanical Industry Press, 2013 edition, p. 5.
[3] Algorithmic Accountability Applying the Concept to Different Country Contexts p. 5 http: //webfoundation. org / docs / 2017 / 07 /Algorithms_ Report_ WF. pdf May 3 2018
[4] Yang Jiwen, "Algorithmic Evidence: a Prospective Look at Algorithms as Evidence and the Rules for Their Application, " in Local Legislation Research, Vol. 3, No. 3, 2022, p. 37.
[5] Zhang Di: "Algorithmic Evidence in Criminal Proceedings: Concept, Mechanism and Its Application", in Journal of Henan University (Social Science Edition) 2023, No. 3, p. 36.
[6] Article 17 of Provisions on the Administration of Algorithm-generated Recommendations for Internet Information Services
[7] Zheng Zhihang, Xu Zhaoxi. Legal Regulation and Judicial Review of Algorithmic Discrimination in the Era of Big Data—Taking U.S. Legal Practice as an Example [J]. Comparative Law Research, 2019, (04): 111-122.
[8] Pan Jingui. The Risks of Criminal Evidence Utilization and Its Regulation in the Digital Era—Taking Algorithmic Evidence as an Analytical Perspective [J]. Rule of Law Research, 2024, (06): 26-42.
[9] Wang Ying. Ruminations on the Framework of Algorithmic Infringement Liability [J]. Chinese Jurisprudence, 2022, (03): 165-184.
[10] Xu Hui, Li Xiaodong. Study on the Evidence Attributes of Big Data Evidence [J]. Journal of the People's Public Security University of China (Social Science Edition), 2020, 36(01): 47-57.
[11] He Jiahong, Deng Changzhi, Zhang Guiyong, et al. Challenges brought by big data investigation to evidence law [J]. People's Prosecution, 2018, (01): 54-57.