Education and Social Work
ISSN Print: 3079-515X
ISSN Online: 3079-5168
https://doi.org/10.63313/ESW.9074

The Evidentiary Value of AI-Generated Data: A Framework for Reliability and Admissibility

Hua Zhang
School of Law, Anhui University of Finance and Economics, Bengbu Anhui, China
Email: 1669074340@qq.com

How to cite this paper: Zhang, H. (2025). The Evidentiary Value of AI-Generated Data: A Framework for Reliability and Admissibility. Education and Social Work, 2(2), 34-41. ISSN Print: 3079-515X; ISSN Online: 3079-5168.
https://doi.org/10.63313/ESW.9074
Published: 2025-07-04

Abstract

The increasing use of artificial intelligence (AI) in criminal investigations has introduced complex evidentiary challenges. AI-generated data—such as facial recognition results, predictive risk assessments, and automated surveillance outputs—often lack transparency, auditability, and human accountability. This paper examines whether such data should be admissible as evidence in criminal trials, and under what conditions. It argues that traditional evidentiary rules are insufficient to address the unique risks posed by opaque and autonomous sys-tems. Drawing on comparative insights from multiple jurisdictions, the article proposes a normative framework based on three principles: verifiability, pro-cedural accountability, and proportionality. The goal is to ensure that efficiency gains from AI do not override fundamental rights such as the presumption of innocence, the right to a fair trial, and the ability to challenge adverse evidence. The paper concludes by offering concrete recommendations for legal reform, including disclosure requirements, independent technical review, and stricter judicial reasoning when admitting AI-generated evidence.

Keywords

AI-Generated Evidence; Admissibility; Criminal Procedure; Algorithmic Bias

1. Introduction

Artificial intelligence (AI) technologies are rapidly transforming criminal investigations, enabling authorities to process massive amounts of data through facial recognition systems, predictive algorithms, and automated surveillance tools. These systems generate outputs that are increasingly relied upon as evidence in criminal proceedings. However, the use of AI-generated data raises serious concerns regarding its admissibility, reliability, and impact on fair trial rights. Unlike traditional forms of evidence, AI outputs are often opaque, difficult to verify, and not subject to human cross-examination. This presents unique challenges to established evidentiary standards and procedural safeguards. While legal systems strive to maintain a balance between efficiency and rights protection, there remains a lack of coherent

frameworks governing the use of AI evidence. This article argues that existing rules are insufficient to address the risks posed by algorithmic decision-making, and proposes a normative framework to guide courts in assessing the admissibility and probative value of AI-generated data.

2. Conceptual Foundations of AI-Generated Evidence

The increasing adoption of artificial intelligence (AI) in criminal investigations has produced a new category of evidence that challenges conventional legal frameworks. This chapter outlines the core characteristics of AI-generated data and explains how it fits—often uneasily—within traditional evidentiary concepts such as relevance, admissibility, and reliability. It also explores the epistemological and procedural tensions posed by the unique nature of algorithmic evidence.

2.1. Definition and Nature of AI-Generated Evidence

Artificial intelligence (AI)-generated evidence encompasses outputs produced by automated or semi-automated computational systems that analyze data using algorithms, machine learning, or deep learning techniques. Typical examples include facial recognition results, automated license plate recognition, predictive risk assessments, and algorithmic surveillance alerts. Unlike traditional evidence derived from human senses or physical objects, AI-generated data is created through complex statistical models and pattern recognition processes. This renders the evidence inherently non-transparent or "black box" in nature, as the reasoning or decision pathways of the AI system are often not accessible or interpretable even to its developers.

This characteristic presents a foundational challenge for the law. Legal evidence traditionally relies on its source being identifiable, trustworthy, and subject to challenge. However, AI evidence emerges as a product of sophisticated software with proprietary algorithms, often protected as trade secrets. As a result, courts, parties, and experts lack full access to the AI system's inner workings. Moreover, AI outputs are probabilistic rather than deterministic, meaning they reflect likelihoods rather than certainties, which complicates their interpretation in a legal context that demands definitive proof beyond a reasonable doubt in criminal cases.

2.2. Traditional Evidentiary Principles and Challenges with AI Evidence

In most jurisdictions, evidence must satisfy three core principles to be admitted in court: relevance, admissibility, and reliability. Relevance requires that evidence logically relate to a fact in dispute. AI-generated data typically meets this standard because it is often introduced to support or refute material facts. However, the problems arise with admissibility and reliability. Legal rules exclude evidence that is unlawfully obtained, prejudicial, or unreliable. AI-generated evidence challenges these criteria because of its lack of explainability and verifiability.

Unlike human witnesses or physical evidence, AI outputs do not testify or possess an

independent provenance that can be examined. The so-called "black box" problem means the reasoning process behind AI conclusions is obscure, limiting the ability of courts or opposing parties to evaluate its accuracy or potential bias. Furthermore, proprietary systems often limit the disclosure of algorithmic details, hindering meaningful scrutiny. This opacity risks allowing evidence that cannot be effectively challenged or tested, undermining the adversarial process. Thus, existing evidentiary standards — which emphasize transparency, opportunity for cross-examination, and trustworthiness — may be insufficient or ill-fitted to AI-generated data.

2.3. Traditional Evidentiary Principles and Challenges with AI Evidence

The introduction of AI-generated evidence in criminal trials raises profound procedural and epistemological questions. Procedurally, defendants have a right to a fair trial, which includes the ability to confront and challenge adverse evidence. When the evidence is produced by an inscrutable AI system, this right is jeopardized. The inability to meaningfully question how an algorithm reached its conclusion challenges fundamental principles of due process and procedural fairness.

Epistemologically, the justice system is built on the idea that truth is discovered through human testimony, cross-examination, and rational evaluation. AI systems rely on statistical inference and machine learning, often without providing comprehensible explanations for individual decisions. This "black box" nature creates a gap between legal proof and technological certainty. Additionally, algorithmic bias—stemming from flawed training data or design—can systematically disadvantage certain groups, threatening the principle of equality before the law.

Finally, issues related to the chain of custody and data integrity emerge because AI outputs are digital and often remotely accessed or transmitted. Courts must determine how to authenticate such evidence reliably. Without clear standards, the risk of manipulation or errors increases. These challenges demand legal frameworks that can balance the efficiency and utility of AI with the need to protect defendants' rights and maintain public confidence in judicial processes.

3. Comparative Approaches to AI-Generated Evidence

3.1. Common Law Jurisdictions: Emphasis on Admissibility Standards and Fair Trial Rights

In common law jurisdictions such as the United Kingdom and jurisdictions influenced by it—including Hong Kong—the use of AI-generated evidence is subject to stringent admissibility tests. Courts emphasize the fair trial principle, and any evidence that cannot be fairly challenged by the accused may be excluded, even if relevant. One critical principle is that of equality of arms, ensuring both parties have a reasonable opportunity to examine and contest the evidence.

For instance, in HKSAR v. Chan Chun Kit (2020), the court admitted CCTV footage

that had undergone AI-based enhancement but required expert testimony to explain the process, ensuring the defense could challenge its accuracy. Similarly, when facial recognition results are presented in UK courts, they must be accompanied by human expert validation and comply with the Criminal Procedure Rules requiring disclosure of methods used and potential limitations.

The hearsay rule also plays a gatekeeping role. Since AI systems do not "perceive" events and cannot be cross-examined, their outputs often resemble hearsay, requiring exceptions or supplemental safeguards. The common law approach thus reflects caution: courts are willing to admit AI evidence, but only when its reliability is demonstrable and procedural fairness is preserved.

3.2. Civil Law Jurisdictions: Flexibility in Evidence Admissibility, Concern for Substantive Justice

In contrast, civil law systems—such as those in Taiwan and many European countries—tend to adopt a more flexible stance on admissibility. Instead of rigid rules, they follow the principle of free evaluation of evidence by judges (Beweiswürdigung nach freier Überzeugung), where the trial judge has broad discretion to assess the probative value of all submitted materials, including AI-generated data.

For example, in Taiwan, the Code of Criminal Procedure does not contain detailed exclusionary rules comparable to the common law. As a result, AI-based outputs such as traffic surveillance analyses or voice recognition results may be accepted into the record, provided the judge considers them trustworthy and consistent with other evidence. However, this approach has raised concerns about insufficient procedural safeguards, especially when defendants lack the technical resources to challenge such evidence.

The lack of adversarial scrutiny in many civil law trials makes it more difficult to test the inner logic or bias of an AI system. Furthermore, judicial reliance on court-appointed experts may lead to overreliance on technocratic authority, without the critical tension typical of adversarial proceedings. Some legal scholars have proposed supplementing this system with mandatory explainability standards or introducing algorithmic accountability laws, but such reforms remain at an early stage.

4. A Normative Framework for Admitting AI-Generated Evidence

4.1. Verifiability and Algorithmic Transparency

A foundational requirement for admitting AI-generated evidence is verifiability. In order to maintain the integrity of the fact-finding process, courts must be able to examine how the evidence was created, what data it relied upon, and what logical or statistical methods were applied. This demands a minimum level of algorithmic transparency—not necessarily full source code disclosure, but at least access to the operational parameters, performance metrics, and development context of the sys-

tem. Without such transparency, judges and litigants are forced to accept the outputs of AI as given, undermining the adversarial process and introducing a layer of epistemic opacity into criminal adjudication.

Verifiability also requires that AI-generated outputs be accompanied by sufficient documentation to enable independent review. This includes records of training data sources, validation procedures, and known error rates. Inaccuracies in facial recognition or predictive policing systems have been well-documented, and courts must be made aware of these limitations in order to appropriately weigh the evidence. An unverified AI-generated result should not be treated as dispositive, nor should it carry a presumption of correctness merely due to its technological origin. Admitting such evidence without scrutiny would effectively delegate parts of judicial decision-making to systems beyond the court's control or understanding.

4.2. Procedural Fairness and the Right to Challenge

The second normative requirement is procedural fairness, which underpins the right to a fair trial in both adversarial and inquisitorial systems. AI-generated evidence challenges this principle by producing results that may not be subject to meaningful confrontation. Unlike human witnesses, an algorithm cannot be cross-examined. While expert witnesses can provide context, they may not be able to fully explain individual decisions produced by complex models, particularly those built on neural networks or deep learning techniques. This weakens the defendant's ability to challenge adverse evidence and risks creating a power imbalance in the courtroom.

To uphold procedural fairness, courts should insist that AI evidence be intelligible and contestable. This means ensuring that the defense has access not only to the output itself but also to the underlying logic of the system that produced it. Protective orders and in-camera review may be used to balance proprietary concerns with the defendant's rights, but complete opacity is unacceptable. Furthermore, judges must be prepared to reject AI-generated evidence where its inclusion would prejudice the proceedings or where the defense lacks a reasonable opportunity to respond. The use of AI must not erode the principle of equality of arms or transform the courtroom into a forum where technological authority overrides legal reasoning.

4.3. Proportionality and Contextual Admissibility

Finally, the principle of proportionality must guide decisions on the admissibility of AI-generated evidence. Not all uses of AI raise the same level of concern. A low-risk application—such as automated transcription of an interview—should not be subject to the same level of scrutiny as a high-stakes output, such as a probabilistic match in a criminal identification. Courts must evaluate AI-generated evidence in light of its potential impact on the defendant's rights, the seriousness of the offense, and the role the evidence plays in the prosecution's case.

Proportionality also calls for careful contextual assessment. Evidence that is only marginally probative but highly prejudicial should be excluded, especially where jurors may overestimate its accuracy due to the "technological halo effect." Conversely, if the AI output is corroborated by independent evidence and its limitations are fully disclosed, it may be admissible as part of a broader evidentiary mosaic. The goal is not to reject AI evidence outright, but to ensure that its use is fair, justified, and proportionate to the legal consequences it carries.

Incorporating proportionality into the admissibility analysis empowers judges to exercise discretion with regard to both the reliability of the technology and the procedural safeguards available in each case. This nuanced approach is essential for avoiding both over-reliance on AI and wholesale exclusion, offering a pathway toward responsible integration of emerging technologies into the criminal justice process.

5.Conclusion and Policy Recommendations

5.1. Rethinking Evidentiary Standards in the Age of Artificial Intelligence

The rapid integration of artificial intelligence into law enforcement and judicial processes calls for a fundamental re-evaluation of traditional evidentiary principles. While AI-generated evidence offers efficiency and expanded investigative capacity, it simultaneously exposes critical weaknesses in existing legal frameworks—particularly regarding admissibility, transparency, and fairness. This paper has demonstrated that AI evidence is often epistemically opaque, procedurally imbalanced, and disproportionately influential in criminal trials. Therefore, evidence law must evolve to account for these new epistemological and structural challenges.

Courts can no longer assume that evidence generated by machines carries intrinsic objectivity. In fact, AI systems encode human biases, reflect the limitations of their training data, and often operate in ways that even their designers cannot fully explain. Consequently, the legal system must shift from a presumption of neutrality to a requirement of demonstrable reliability and contextual scrutiny. The essence of justice lies not merely in the correct outcome, but in the fairness and transparency of the process. AI must be subjected to the same demands.

5.2. Institutional and Legislative Reforms

Meaningful regulation of AI-generated evidence requires not only judicial discretion but also comprehensive institutional and legislative reform. At the legislative level, jurisdictions should consider enacting AI-specific evidentiary statutes that establish clear criteria for admissibility, including thresholds for transparency, expert disclosure, and algorithmic testing. Such laws could mandate documentation of data sources, model performance under case-specific conditions, and error rate disclosure. Legislatures must also define who bears the burden of proving reliability—

ideally the party seeking to introduce the evidence.

On the institutional level, courts should develop procedural rules that facilitate robust challenge to AI evidence. This includes granting defendants timely access to algorithmic details, encouraging the use of independent technical experts, and allowing in-camera review of proprietary systems where necessary. Training judges and lawyers in AI literacy will also be crucial. Judicial actors must be able to ask the right questions, weigh the risks, and avoid both over-deference to technological systems and uninformed skepticism.

In addition, jurisdictions should consider creating public certification systems or regulatory sandboxes to vet AI tools before they are introduced into criminal trials. These systems can establish baseline technical standards, promote accountability from private developers, and build public trust. Without such forward-looking structures, the risk of inconsistent judicial outcomes and undetected injustice will only grow.

5.3. Future Outlook: Towards Responsible and Rights-Based AI Integration

As artificial intelligence continues to reshape legal practice, its integration into evidentiary processes must be guided by a principled commitment to human rights, transparency, and democratic accountability. The use of AI in criminal adjudication cannot be separated from broader questions about surveillance, digital inequality, and power asymmetries between the state and individuals. Therefore, any framework for AI-generated evidence must be embedded in a larger normative commitment to responsible innovation and rights-based governance.

Future developments should focus not only on legal admissibility but also on normative legitimacy. Even the most technically accurate AI tools may be rejected if they undermine core values of the legal system. For example, predictive policing tools may generate evidence that is formally reliable but substantively unjust, particularly where they entrench historical patterns of discrimination. Similarly, automated facial recognition systems may pass technical tests but still raise proportionality concerns due to their invasive nature.

The legal community must remain vigilant as new technologies emerge. This includes fostering interdisciplinary collaboration among jurists, engineers, ethicists, and civil society actors. Rather than reactively responding to crises, the justice system should proactively shape the terms under which AI enters the courtroom. A legal order that harnesses the power of AI while safeguarding procedural justice will not arise spontaneously—it must be deliberately designed.

References

[1] Weimin Zuo.Some Reflections on the Prospects of Legal Artificial Intelligence in China[J]. Tsinghua Law Journal 12, no. 2 (2018): 108-124.

[2] Changshan Ma.The Reshaping Effects and Limits of Judicial Artificial Intelligence[J].

Chi-nese Journal of Law 42, no. 4 (2020): 23-40.

[3] Fanqing Cheng.Problems and Countermeasures in the Development of Judicial Artificial Intelligence in China[J]. Oriental Law no.3(2018):119-130.

[4] Xunhu Li.Inclusive Regulation of Artificial Intelligence in Criminal Justice[J]. Social Sci-enc-es in China no. 2 (2021): 42–62, 205.

[5] Jingping Huang.A Negative List Approach to Criminal Justice Artificial Intelligence[J]. Ex-ploration and Free Views no. 10 (2017): 85–94.