ADVANCE SOCIAL SCIENCE ARCHIVE JOURNAL
Available Online: https://assajournal.com
Vol. 05 No. 01. Jan-March 2026.Page#.7706-7717
Print ISSN: 3006-2497 Online ISSN: 3006-2500
Platform & Workflow by: Open Journal Systems
https://doi.org/10.5281/zenodo.23019647

Artificial Intelligence and the Integrity of Criminal Evidence: Authentication, Admissibility, and Fair Trial Rights

Kashif Javed
(Corresponding Author)
LLB Hon's 5 years
Kashifchabcd@gmail.com
Ali Shehr Yar Khan
LLB Hon's 5 years
alikhanpunya513@gmail.com
Dr. Ishfaq Ahmad
Lecturer in Law
BZU Sub Campus Vehari
ishfaq.ahmadvcamp@bzu.edu.pk

Abstract
Artificial Intelligence (AI) is now being used in many aspects of evidence production, examination, processing and presentation in criminal proceedings. As AI can improve the efficiency of investigations and analysis in the forensics world, the answers to the following questions are not trivial and arise when using AI tools: Where did it come from? What is it? Is it authentic? Can it be reliable? Is it admissible in court? These considerations are of particular importance when it comes to the use of AI to create or significantly edit photographs, audio files, videos, documents, or investigation telling. With the advent of deepfakes and other synthetic media, it is growing increasingly difficult for the courts to determine which media are real and which ones are flawlessly produced. This paper discusses implications of AI in relation to the integrity of criminal evidence centrally focusing on authentication and admissibility, reliability and fair trial rights. It examines the difference between AI-generated, AI-assisted, AI-enhanced, and AI-analysed evidence, and discusses the issues of provenance, chain of custody, metadata, technical verification, expert examinations and algorithmic transparency. The paper also reflects on and discards the connection of the AI evidence to key procedural justice principles, such as the presumption of innocence, equality of arms, disclosure, and the ability of the accused to contest prosecution evidence. Specific focus is on laws and statues in Pakistan related to the admission of electronic evidence and the right to fair trial in constitution. The paper takes a stance between automatic disregard of AI evidence and automatic inclusion, asserting that it ought not to be either ignored or dismissed outright. Rather, technology-enhanced evidentiary protections should be used that can demonstrate authenticity, determine reliability, maintain a balance of fairness in the proceedings, and allow for meaningful judicial review. Balancing act can enable courts to enjoy the positive aspect of AI applications and at the same time safeguard the sanctity of criminal adjudication.

Keywords: *Artificial Intelligence; Criminal Evidence; AI-Generated Evidence; Digital Evidence; Authentication; Admissibility; Evidentiary Reliability; Deepfakes; Synthetic Media; Fair Trial Rights; Algorithmic Evidence; Electronic Evidence; Criminal Justice; Pakistan.*

1. Introduction

An integral and fundamental element of the administration of criminal justice is that courts can recognize that the evidence is correct and separate the accurate evidence from that which is inaccurate, tainted or cannot be meaningfully checked. So, for centuries, evidentiary systems have been built around more familiar types of evidence, such as oral evidence that has appeared before documentation, physical items, photos, aural or visual recordings. The use of artificial intelligence (AI), however, is changing the way evidence is created as well as how evidence is analyzed. The AI systems have the ability to now analyze extensive data sets, improve the quality of images or recordings, create synthetic audio and video walls, complete missing information, help investigators with report writing, and automatically produce new output that can then be used as factual information in court cases. The problem in the case of an evidentiary challenge is thus more than the standard one associated with determining if a digital file has been properly preserved. It involves issues of whether the material reflects what actually occurred in the event; whether its production/processing can be the subject of independent scrutiny; and whether the accused has an opportunity to comment on it to the extent that it is meaningful.

An important context to understand these challenges is provided by Grimm, Grossmann and Cormack who differentiate the Validity and Reliability of an AI application. Validity refers to the ability of the AI system to actually complete the task it was programmed for while reliability concerns the consistency of the AI with repeatable use on rather similar circumstances (Grimm et al., 2021a). The significance is especially in criminal circumstances. It is possible to get the same result again and again and yet still get the wrong answer. Repetition, or sophistication of technology and methods, therefore cannot, in itself, create evidentiary reliability.

Adding the ability of AI to create evidence adds to the problem, if AI can create evidence, as opposed to merely analyze existing material. When using generative AI, it is possible to create images, voices and videos and written stories that could seem realistic but may have no direct connection with the event itself. In addition to false information, deepfake technology introduces other risks, such as eroding criminal justice trust. Proof of these risks can be found in Sandoval et al. (2024b)'s systematic review of studies on deepfakes and criminal justice (Sandoval et al., 2024a). They found that the issue with deepfakes as evidence is not just about these specific cases, but also suggests a broader impact on digital evidence as a whole, such that if justice and the public in general are unable to agree with any degree of confidence that audiovisual evidence is real, this could negatively impact all digital evidence.

AI can also impact the generation of seemingly routine evidence in the investigation phase. This is seen in Ferguson’s analysis of how police reports were improved in 2025 by AI assistance. Generative AI can output a draft police report, body-cameras can report in some instances, and an officer can edit and adapt the AI's output and then use that as an official police report (Ferguson, 2025a). The resulting paper might seem to be a typical police report, but part of its factual description came from an algorithm. These are issues that relate to issues of authorship, accuracy, omissions, hallucinated details and whether the underlying recording is even still required to assess the reliability of the report.

This evidentiary value of AI should then be explored on the basis of the following concepts which are interconnected: Authenticity, Provenance, Reliability, Disclosure, Adversarial Testing, Probative value. The question of authentication is whether evidence is indeed what its proponent holds it to be. Provenance: about its source, following use, processing and transformation. Reliability, This is about whether or not the processes for

making, analyzing or understanding the material can be reasonably trusted. These ideas take on added importance in scenarios where an AI system is proprietary or overly complicated, as the adversary may not be able to replicate the process or analyze the model used.

The technological evidence-fair trial rights issue is also crucial. Stoykova's emphasis is that the right to a fair trial offers a conceptual structure to engraft the rules governing digital evidence and highlights equality of arms and presumption of innocence as central features of the right to a fair trial (R. Stoykova, 2023a). This is particularly relevant to computer-generated evidence a science gap between prosecution and defense can be experienced due to the technology's complexity. Even if the prosecution does rely on a result that is formally subject to cross they may not give the defense access to relevant data, methodology, model limitations or relevant processing history to provide a genuine opportunity to cross this evidence.

So it is not a question of whether or not AI should be accepted as a category of evidence. It would be technologically and legally 'over-ambitious' to do so. AI is a large category of technologies, and the evidentiary reliability of a specific one will vary based on the specific tasks, datasets, and deployment methods, and the specific intended uses of the results of the AI. It is instead the proper question of law whether any specific piece information obtained from the AI meets the criteria of authenticity, reliability, relevance, procedural fairness and lawful admission.

This paper discusses the dividends for criminal evidence integrity when considering Artificial Intelligence (AI) and explores three related questions. What principles of evidence should apply to the authentication and admissibility of AI created or AI processed documents? Second, what is the limits of reliance on synthetic media, algorithmic outputs and AI-assisted investigative material where courts are concerned? Third, what measures must be put in place to safeguard fair-trial rights enjoyed by the Accused such as the presumption of innocence, equality of arms, the right to disclosure and to contest evidence adduced by the prosecution, in the use of AI?

It is written in a doctrinal and comparative manner and takes into account the new developments of the literature concerning evidence and AI, digital forensics, deepfakes and algorithmic decision-making. It also puts the Pakistani law context into account including the evidentiary framework of Pakistani Law, and the constitutional rights for fair trial. The main thesis is that the neutrality of technologies is not synonymous with their procedural neutrality. Evidentiary rules and practices can still be appropriate for AI-generated content, but there needs to be a more defined process relating to the technology verification, disclosure, expert opinion, and reasoned faculty by the Court where AI-generated content is part of evidentiary proceedings.

The goal is therefore not to develop an evidentiary system for each of the various forms of artificial intelligence. Instead, it is about finding a principled way for courts to apply the requirements of criminal adjudication to AI-generated evidence to assess whether it is sufficiently authentic, reliable and contestable for such purposes. The verdict on the integrity of the criminal process does not depend ultimately on whether evidence is acquired from a human or machine, but whether the court can determine what the evidence is, how it is produced, if the evidence has limitations, and whether the accused was given a meaningful opportunity to challenge the evidence.

2. Conceptualizing AI-Related Criminal Evidence

In order to explain various aspects of evidence that could be generated using artificial intelligence, it might be best to use an expression such as "AI-involved evidence" rather than terming them "AI-generated evidence." AI can serve very different roles in criminal cases, it can create something new, help sort through things already created, enrich an original recording, identify a person or thing, or help an investigator build a nice fact narrative. These various functions have varying import as regards authenticity, reliability and judicial evaluation. Thus, a meaningful evidentiary framework starts with delineation of the main types of evidence pertaining to an AI, not with considering artificial intelligence to be a single type of evidentiary phenomenon.

AI–created evidence is the first. This is the matter created in huge degree via generative frame of works instead of being directly set down by an underlying occasion. Synthetic images of photos, videos made of false footage, replicated voice files, and likewise produced documents and computer-generated stories. The main problem is that the digital file created might be legally 'real' but false as an illustration of an event. Presented side by side, the difference is great, as traditional authentication usually poses the question: Is the work on offer what its claimant claims? But with synthetic material, the more challenging question could possibly be whether there is some real-world event behind which the material stands or not.

Many of the usual assumptions about authentication have to be reconsidered, as McPeak's deepfake analysis shows. Because of the advances in synthetic media, she contends, the litigation environment poses the greatest danger, as seemingly credible audio-visual evidence can be fabricated and the event portrayed never actually happen. Her suggested answer is not to disallow digital media wholesale but to reevaluate the volume and kind of circumstantial evidence that must be evaluated when assessing authenticity (McPeak, 2020a). This is especially true of trials, as criminal cases are especially topical on the point of invented evidence being admitted for them.

The second type of evidence is AI-powered. Here, the source of the underlying material is associated with a human activity or physical event, but the activity of an investigator or Forensic Examiner is aided by AI. A case in point would be automated facial recognition. The picture or video could be taken by a regular camera, then an algorithm runs to match the picture, or video, with a database and provides a possible identification. The authors of the book, Kotsoglou et al., point out that although the result obtained from an automated facial recognition is computational, it does not necessarily be objective. Of course, its probabilistic nature, assumptions and potential rates of error play a factor here as well, especially if the result was used to justify police action and will be put forward later as evidence (Kotsoglou & Oswald, 2020a).

This category is significant due to the fact that the evidence is not originally developed by the AI system. Rather, it's constructed an interpretation of evidence. The court has, therefore, to make a distinction between the original material, and the algorithmic conclusion based on that material. For instance, if a "match" was made using facial recognition and there was no actual witness to the accused's face, it should not be accepted as if the witness actually witnessed the offense. The reliability and importance of the algorithmic output as evidence depends on the performance of the system, its data compilation/collection, operating parameters, error rates, and how the output was validated, as well as the context of its use.

The third type are evidence rich with AI. Enhancement is when an image, aural recording or video is manipulated to make it easier to see, hear or understand. Enhancement can include reduction of noise, sharpening of images, improving the resolution and filling in missing parts, as well as other transformations in the computer. This processing can help the court to grasp material which already is before it but there is also a possible threshold following the processing between clarification and alteration. The information reproduced by an AI system that was not originally recorded can turn into an inference or reconstruction instead of just an enhancement of the original recording.

This distinction has become more and more significant when the apparent "enhancement" gives the impression of more detail present in the original recording when this is not really the case. Therefore, the court should be able to ascertain what original data was used, what was done to the original data, what software and/or model was used to work with the original data and the parameters used, and to what extent output differs from the original data. It is not whether the enhancement is technological per se that is legally significant, but whether there are any preserved characteristics of the evidence.

The fourth type is AI analyzed evidence, where a system leverages an extensive amount of data to detect patterns, correlations, classifications and/or probabilities that can help with an investigation. An instance from which a

more general picture can be drawn is algorithmic risk assessment. König and Krafft show that there can be a significant variance in algorithmic decision making systems when it comes to how they are built, validated, documented and perform(König & Krafft, 2021). Their analysis is to do with risk assessment before trial and not more conventional evidence of a criminal act, but the evidentiary issue raised is relevant: whether the evidence or output of a number/algorithm is any good by itself is not enough to say whether it is good for a legal decision.
AI analysis can thus generate a chain of evidence as opposed to a single piece of evidence. This can involve the original data, how it was collected, how it was preprocessed, the algorithm or model, the parameters used, the data product it outputs, but also the human interpretation of the precise nature of the produced data. Weakness at any step can impact the amount of weight which should be given the final conclusion. Therefore, the evaluation of such evidence analyzed by an AI tool applied to a court case should not only be based on the result but on the processes that led to it.
Another challenge comes into the picture when using AI for creating traditional investigative documents. Increasing use of generative systems in policing shows that AI can have an impact on the evidence even if it doesn't show up literally in the final product. An AI system can take the footage from a body camera, interviews or other information and generate a draft investigative narrative, which can then be reviewed and/or used by the officer. The document can look like a regular police report, but was run through an algorithmic production process. Thus, the boundary between the human observer and the nature of the language he entered and edited by machine becomes a factor in considering the nature of the evidence.
It's starting to be important when researchers look at digital evidence to also consider the entire evidentiary process, and not just the final digital file. In this regard, Stoykova's proposed “right to procedural accuracy” is very relevant. She believes that the processing of digital evidence should include a protective mechanism against unreliable processing, access to the digitized evidence, explanations of the evidence processing methods (R. A. Stoykova, 2024a). Her work shows that digital evidence integrity is not just an issue of the end points of the process, but an issue of process accuracy along the entireity of evidence-processing chains.
The differences among the 'AI generated', 'AI assisted', 'AI enhanced' and 'AI analysed' evidence consequently creates different legal questions in the considerations. The actual issue with AI evidence is if it's genuine or what is synthetic. The issue with AI-aided evidence is whether or not the interpretation of the original by the algorithmic procedure is reliable enough. The question with respect to AI-enhanced evidence is whether computational processing has left the content of the evidence unchanged or whether it has materially altered the content of the evidence. Where analysis by artificial intelligence is used, factors like quality of the underlying data, methodology used, validation and interpretation should be examined by the court.
This classification also has the benefit of keeping the courts from taking an overly expansive view of admissibility. Don't allow AI to be the reason for an automatic boost or demotion in evidentiary weight. Rather, particular questions should be examined based on the function that the technology performs. A court can make reasonable distinctions between a party using an Original surveillance recording noise reduced vs an AI generated recreation of an event.
This process-centered way of thinking about digital evidence is also corroborated by newer research about digital evidence. The safeguards of the access to chain of evidence, proceeding with explanation and forensic assistance are revealed as important in criminal proceedings' context of role relevance to the procedural accuracy of Stoykova's work (R. Stoykova, 2021a). They are especially important if the defense is unable to replicate the AI process. When predicting whether an accused person is likely to challenge an conclusion, meaningful challenge at least requires information adequate to allow such a person to understand what has been done to the original material and how the queed conclusion was arrived at.

As such, AI-related evidence needs to be understood not as a new 'species' of evidence, but as a set of technological interventions, at various stages of the evidentiary process. Once courts have acknowledged that evidence has been created, modified or understood using AI, the question remains how that evidence should be verified and admitted as evidence. The main equipping should be the more AI transformed the evidentiary material/has given an input to a finding, the more necessary is there a provenance and methodological transparency and independent verification.

3. Authentication and Admissibility of AI-Generated Evidence

The first evidentiary hurdle in AI generated and processed content is how to deal with authentication. Authenticity is usually a prerequisite for the introduction of evidence - the party producing evidence must first prove the evidence is what it appears to be. The investigation is complicated by the fact that a digital file can be fully legitimate as a file yet the content of such a file could have been "made up" or edited or built up synthetically. The other question resulting from this is no longer only whether a file has remained unchanged after the has been collected but also whether the file is representative of an underlying even.

Considerations of the admissibility of AI generated evidence cannot only be based on the sophistication of the technology used to create it, but also on the validity and reliability of the technology as a whole (Grimm et al., 2021b). Their analysis is significant as courts might otherwise take the results of computations as having a built-in objectivity to them. An algorithm, on the other hand, only implements an approach, it does not mean the approach is suitable, the data used is correct, and the results of the conclusion are error-free.

Provenance therefore plays a key role in the authentication process. Where reasonably possible, a party that relies on evidence generated by artificial intelligence should identify the source of the original evidence, how the evidence was obtained, which artificial intelligence system was used to create the evidence and the processing steps that the evidence went through to be created. An original video for an AI enhanced video, for instance, should be kept, and the court should have access to it to compare with the AI's video. In the case of AI-generated images, the relevant people may be interested in the process used to produce them, in the prompt or input data, in the history of creation and in the extent to which any original substance was used in the creation of the output. Although chain of custody is still applicable, it needs a wider understanding in the context of AI. Traditional chain of custody practices is mainly directed towards the physical handling and preservation of evidence. There are lots of things that need to be done to the digital evidence, copying, storage, metadata, conversion and other technical processing. The procedural accuracy of Stoykova's research suggests a need for more enhanced governance, involving safeguard mechanisms for digital evidence, such as who should process the data and provide access to information needed for proper examination (R. A. Stoykova, 2024b). his is particularly important in instances where the AI has substantially changed the content, as there is the potential for changes and mistakes during every step of processing.

Whilst metadata can be informative with regard to the history of a digital file, metadata should not be regarded as definitive proof of authenticity. Metadata can then potentially be modified, removed or re-created. The authenticity of the image should thus be determined from the totality of the evidence, such as from examinations on the original device, where appropriate, hash values, system logs, time and location of storage, and forensic examination and evidence of people familiar with the process of creation or treatment.

Expert evidence can play a vital role, especially where the court is not able to independently evaluate the level of AI process. The expert should not simply provide information that an AI system has generated a certain outcome. When relevant this should include an explanation for the purpose, operating methodology, validation, known error levels, any relevant limitation and the conditions in which the particular production came about in relation

to this system. This is in line with the general rule that expert evidence is meant to help the court to understand expert knowledge and not to do the court's task.

It's more complex at the point at which AI systems are proprietary or commercially private. The prosecution could try to use the algorithmic output without requiring the publication of the algorithm itself or the resource code which created it, and allege it was not needed and preserve its trade secrets. Things like this can cause an imbalance if the defense can't establish if the system is reliable or if some fault may have occurred in favor of the accused. Transparency and explainability have been key issues in algorithmic decision making, which have been noted by legal academics who have studied the use of automated systems in making consequential legal decisions (Pasquale, 2015). The problem is not so much that complete disclosure, that is the disclosure of source code, is always required, but that by themselves, confidentiality restrictions do not bar the practical defense of the accused by the presentation of evidence.

Also weight is not authentication. Evidence can be given a "preliminary" authentication even if it does not later warrant full evidentiary weight. A court could find that video captured is the same copy as the video that has been recovered from a specific device and yet be unclear as whether AI has changed the depiction of the event. On the other hand, a product of an artificial intelligence can be obviously recognized as made in a specific artificial intelligence system and yet a few arguments on the point of fact.

This distinction is pertinent as evidenced on Deepfakes. As synthetic media becomes more sophisticated, courts may need to be presented with greater evidence of authenticity to take synthetic media that appears to be realistic (McPeak, 2020b). It is not the problem that digital evidence is unreliably produced or culled, but rather the issue of visual realism as a reliable authentication point is not reliable anymore.

Therefore, there should be a structured inquiry of admissibility. The court ought to learn what kind of AI support is being invoked, who originated the material, the processing and chain of custody history, the technical reliability of the system, provide for proper expert review and provide a meaningful opportunity for the other party to challenge the evidence. The evidentiary weight to be attached to the material should be determined only after the above questions are answered.

Automatic acceptance or exclusion of AI-generated evidence is thus the wrong way of going about it. In the era of Artificial Intelligence, authentication means a shift from a file to process centred system by courts. It is not only attribution of authenticity to the digital "work" that courts consider; courts ask for evidence of how the work was created and the transformations that took place, and the court considers the degree to which the evidence gives it trust.

4. Reliability, Deepfakes and Synthetic Evidence

Authentication proves the source and author of evidence; not necessarily its reliability. This distinction is significant in instances involving the use of AI to create, manipulate, categorize or enhance evidence. Digital files can be proven to be the object and the data in the file can be wrong. The inquiry into reliability should therefore consider whether the technology and methodology of producing the evidence is reliable for the specific fact to which the evidence tends to prove or disprove.

As seen in the case of deepfakes the issue is most apparent. The deepfake technique can control facial expressions, voices, and video sequences to coherently create fake content. Also pointed out by Chesney and Citron's analysis of deepfakes are its potential to reduce trust in audiovisual communications and as the ability to produce falsified accounts of people and events (McPeak, 2020c). Their pieces would have been created before the latest generation of generative AI systems, but the challenge they grappled with is still quite pertinent: Once created material looks or sounds highly convincing, courts can no longer safely assume that it is factually true.

Recent empirical studies have shown that this issue is still ongoing. Nightingale and Farid discovered that AI created faces are perceived as extremely realistic and they can in certain cases be seen as more trustworthy than real faces (Nightingale & Farid, 2022). The discovery has an immediate bearing on the matter of judicial... It is not always safe to rely on the human visual intuition as an authentication method when synthetic media are concerned. The honest conviction of a judge, lawyer, police officer or witness that an image is authentic can be incorrect as to its source.

Sandoval et al.'s 2024 systematic review further identifies deepfakes as a threat to criminal justice through evidence falsification, difficulties of attribution and damage to institutional trust (Sandoval et al., 2024b). These findings are of importance because it does not stop one potentially fabricated exhibit from being brought forward. Once those caught up in criminal proceedings orient themselves to the idea that anything intimate can be a deception, anything from the jury or prosecution seems fair game, too. This has been referred to as the "liar's dividend," and perhaps synthetic media does offer a chance for the individuals to question the credibility of real media.

Another methodological issue with AI-generated evidence is that it poses a challenge. Detection systems are also technological; they can give false positive or false negative results. Such a court should therefore be wary of relying on the AI detector to provide a conclusive verdict as to whether a recording comes from a true source. Instead, the methodology of the detector, its validation and error rates, its data set and expect ability towards the particular file should be investigated.

The same can be claimed about the reliability with regard to AI-assisted forensic evidence. Probabilistic conclusions can be made by a facial-recognition system, for instance. In previous work, Kotsoglou and Oswald show that the accuracy of the performance of automated facial recognition has a material impact beyond the trial itself, as it can trigger and serve as evidence for police interventions (Kotsoglou & Oswald, 2020b). As a consequence an algorithmic identification must therefore be considered as an identification within the element and the conditions in which the algorithmic identification occurs.

There is a further concern with generative AI – that it can create text or rebuilt information which can be made to look credible, yet not supported. When a narrative is generated as part of Ferguson's research on AI-powered police reports, it can impact certain aspects of the regular investigative process and ultimately affect the determination of probable cause, detention, discovery and trial practice (Ferguson, 2025b). Continuing the scale from part 1, just because an AI statement looks coherent doesn't mean all of the factual statements in it are backed up

The reliability enquiry should therefore take into account at least four questions: the quality of the data used by the AI, how the AI was built and how it works, the known limitations of the AI, and the connection with the fact in dispute and the AI's output. Uncertainty should not be overlooked by the court when the seeming accuracy of an algorithmic result is used in lieu of proof beyond reasonable doubt when the uncertainty is substantial.

The fundamental evidentiary requirements of the evidence remain the same but are heightened in the use of AI. Technological skepticism, as a reaction to synthetic evidence, should thus be presented in the form of a structured evaluation of the factual basis on which the synthetic evidence rests and of the ability to challenge the reliability of synthetic evidence in respect of this specific evidence.

5. AI Evidence and Fair Trial Rights

AI related evidence admissibility cannot be divorced from the right to a fair trial of the accused. But it is not just whether or not the AI has generated accurate results, but whether the accused is able meaningfully to understand and challenge the evidence that the prosecution relies on. When implementing AI systems in criminal processes, Quezada-Tavárez et al. state disclosure, the reliability of evidence and the realization of a fair right to defense as

important issues (Quezada-Tavárez et al., 2021a). Their analysis shows that traditional notions of fair trial should be held true for technological processes and not only for traditional witness testimony.

presumption of innocence is a further protection. Lack of reliability testing, a lack of adaptation of procedural rules to digital evidence and therefore fairness and the presumption of innocence are identified as a threat (R. Stoykova, 2021b). Thus, any algorithmic output should not be given evidence-based status just because they are given in a numerical way or are the result of a sophisticated technology. The prosecution has the same duty as any other civil or criminal case to prove the factual basis to its case by more than just police electrified notebooks and legible records, by legally admissible and sufficiently reliable evidence.

Armed with technical information not enjoyed by the defense, equality of arms is all the more important. This is further elaborated upon in Stoykova's follow-up work, where she demonstrates how notions of fair trial should be utilized as a framework for developing digital-evidence rules that rely on equality of arms and the presumption of innocence (R. Stoykova, 2023b). Appropriate challenge might involve the release of pertinent data, the processing history, methodology, limitations and expert information. However, it is likely that, if there is no access to the necessary facilities, the defense can formally challenge the evidence without having the means to go further.

The case of algorithmic opacity also touches on the question of the reasoned decision of the judge when it comes to the accused. Both authors note that such use of AI in criminal court might spark some inquiry about transparency, legal certainty, fairness and judicial independence (Franssen, 2023). Therefore, a judge should explain himself as: why does he consider an AI-generated conclusion sufficiently reliable and how he has handled clashes of challenges to its reliability. Just relying on an unsupported algorithmic judgment could result in an evidentiary judgment being delegated to a technological system.

Marshall et al. also warn against the notion that online evidence can simply be assumed to be reliable and highlight the need for proper disclosure and assessment tactics (Marshall et al., 2021). The same principle has a direct bearing on AI evidence: "Technological origin may not be a substitute for evidence of reliability.

Therefore, there needs to be a meaningful opportunity for the accused to access, comprehend and contest evidence generated by an AI. When the limitations of a system can't be meaningfully explored, courts should then look to whether relying on the output of the system would substantially prejudice the defense.

6. Comparative and Pakistani Legal Perspectives

Experience with the use of AI in evidence shows a trend in evaluating AI evidence based on established evidence law principles, rather than considering a separate category of evidence. For an example of scholarship in Europe, it is stressed that evidence generated or processed with the aid of an AI must not lose its characteristics of being subject to fair-trial requirements, specifically disclosure (Quezada-Tavárez et al., 2021b). This is applicable to Pakistan as it shows that there is room for technological innovation and there will still be traditional aspects and mechanisms put in place.

There is already provision in the Pakistani framework that could be adapted to deal with certain types of electronic evidence. The recording of evidence or witness by means of modern devices or techniques is allowed in courts by this Order in case-by-case basis. In 2023, the provision was updated to make a specific reference to contemporary communication technologies (Hayat, 2025). In certain situations also, Article 73 provides recognition of electronic outputs, such as print-outs, or output of an automated information system for which the statutory requirements for the system's operation, or for security procedures, are met (Idrees et al., 2025). These provisions are significant as they offer an important statutory base to address technologically generated material.

But electronic evidence and computer-generated evidence do not equal. An electronic recording might just digitally preserve an event, or AI could actually produce, manipulate or render the content. As a consequence, the

mere existence of an electronic file cannot be used to answer any questions about the possibility of AI manipulation or about synthetic creation or algorithmic interpretation. The existing evidentiary provisions of Pakistan should thus be interpreted and used in combination with the general requirements of authentication, relevance, reliability and fair trial.

Recent Pakistani scholarship has begun examining the effect of AI on criminal evidence and trial practice. Tahir, Saeed, Murtaza, and Abdullah's empirical study considers the presentation and examination of AI-related evidence within Pakistan's criminal justice system and highlights the need to address technological developments within the existing judicial framework (Tahir et al., 2025). Although this emerging literature remains limited, it indicates the increasing practical relevance of AI evidence to Pakistani courts.

A recent Sindh High Court judgement further points to the impact of technologically created evidence in the court of law. The court noted that while evidence generated under modern technology techniques should be preserved under the relevant provisions under Qanun-e-Shahadat Order, the omissions in doing so may have an impact on the admissibility of such evidence (Ashraf et al., 2026). The decision does not announced a full AI-evidence doctrine, but the Court has a focus on preservation which is relevant to AI-generated content and should be met prior to any court evaluating the contents.

Thus, the development of the Pakistani position can be done step by step in a gradual manner. Instead of developing a robust statutory framework for all AI uses at once, the courts can continue to follow current evidentiary principles and demand further technical information whenever there has been a “material impact” from AI on the evidence. Requirements should include: identification of original source, preservation of original material, disclosure of significant AI processing, and an appropriate technical and/or expert examination, and a judicial assessment of reliability.

This way would also be consistent with the guarantee in the constitution of a fair trial and due process. AI has a part to play in adjudications, but should not be the final word in evidentiary authority. The court needs to have responsibility regarding admissibility of evidence and its weight.

Conclusion

The criminal evidentiary process, from production to processing to presentation, is being revolutionized by the use of artificial intelligence; this new development poses challenges not fully addressed by traditional evidentiary rules. But it is not the technological source of evidence that poses the greatest hurdle, rather it is the courts ability to identify the authenticity, source, reliability and integrity of evidence.

Evidence generated and processed by an AI system, should be carefully authenticated, technically verified, disclosed and put under adversarial inquiry. Technology power, by the means of media synthetic, algorithm identification and AI-assisted investigation, cannot be used as the substitute of the evidentiary reliability. Taken to extremes, this would imply that the accused should be given a meaningful chance to have the technology behind the evidence presented by the prosecution examined and rebutted.

Existing principles that Pakistan's law prescribes regarding electronic evidence alongside the constitutional provision of issuance of a fair trial can form the basis for dealing with these developments. What is needed is an application of familiar evidentiary protections in a technologically advanced manner, not blanket bans on the use of AI evidence. Finally, the continuing validity of decision making by AI in criminal proceedings will for example depend on whether it still is possible for courts to independently assess the evidence, recognize their limitations and for them to make a reasoned decision in accordance with the needs of justice and justice in trial.

References:

Ashraf, M. A., Aslam, S., & Ashraf, M. R. (2026). FROM STATUTE TO COURTROOM: ADMISSIBILITY, JUDICIAL REASONING, AND INSTITUTIONAL GAPS IN PAKISTAN'S TREATMENT OF ELECTRONIC EVIDENCE. *Journal for Social Science Studies*, *4*(2), 228–237.

Ferguson, A. G. (2025a). Generative suspicion and the risks of AI-assisted police reports. *Nw. UL Rev.*, *120*, 299.

Ferguson, A. G. (2025b). Generative suspicion and the risks of AI-assisted police reports. *Nw. UL Rev.*, *120*, 299.

Franssen, V. (2023). The use of AI tools in criminal courts: Justice done and seen to be done? *Artificial Intelligence (AI) and Criminal Justice. Automated Decision-Making and Its Application in Criminal Law*.

Grimm, P. W., Grossman, M. R., & Cormack, G. V. (2021a). Artificial intelligence as evidence. *Nw. J. Tech. & Intell. Prop.*, *19*, 9.

Grimm, P. W., Grossman, M. R., & Cormack, G. V. (2021b). Artificial intelligence as evidence. *Nw. J. Tech. & Intell. Prop.*, *19*, 9.

Hayat, M. F. (2025). The Constitutionality of Article 40 of the Qanun-E-Shahadat Ordinance 1984 in the purview of Article 13 (B) of the Pakistani Constitution. *Journal of Law, Justice and Human Rights*, *1*(1).

Idrees, R. Q., Kiyani, S., & Shahid, A. (2025). The Competency and Reliability of Child Witnesses under the Qanun-e-Shahadat Order, 1984: A Critical Analysis in the Context of Pakistan. *Pakistan JL Analysis & Wisdom*, *4*, 65.

König, P. D., & Krafft, T. D. (2021). Evaluating the evidence in algorithmic evidence-based decision-making: The case of US pretrial risk assessment tools. *Current Issues in Criminal Justice*, *33*(3), 359–381.

Kotsoglou, K. N., & Oswald, M. (2020a). The long arm of the algorithm? Automated Facial Recognition as evidence and trigger for police intervention. *Forensic Science International: Synergy*, *2*, 86–89.

Kotsoglou, K. N., & Oswald, M. (2020b). The long arm of the algorithm? Automated Facial Recognition as evidence and trigger for police intervention. *Forensic Science International: Synergy*, *2*, 86–89.

Marshall, P., Christie, J., Ladkin, P. B., Littlewood, B., Mason, S., Newby, M., Rogers, J., Thimbleby, H., & Thomas, M. (2021). Recommendations for the probity of computer evidence. *Digital Evidence & Elec. Signature L. Rev.*, *18*, 18.

McPeak, A. (2020a). The threat of deepfakes in litigation: Raising the authentication bar to combat falsehood. *Vand. J. Ent. & Tech. L.*, *23*, 433.

McPeak, A. (2020b). The threat of deepfakes in litigation: Raising the authentication bar to combat falsehood. *Vand. J. Ent. & Tech. L.*, *23*, 433.

McPeak, A. (2020c). The threat of deepfakes in litigation: Raising the authentication bar to combat falsehood. *Vand. J. Ent. & Tech. L.*, *23*, 433.

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, *119*(8), e2120481119.

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. In *The black box society*. Harvard university press.

Quezada-Tavárez, K., Vogiatzoglou, P., & Royer, S. (2021a). Legal challenges in bringing AI evidence to the criminal courtroom. *New Journal of European Criminal Law*, *12*(4), 531–551.

Quezada-Tavárez, K., Vogiatzoglou, P., & Royer, S. (2021b). Legal challenges in bringing AI evidence to the criminal courtroom. *New Journal of European Criminal Law*, *12*(4), 531–551.

Sandoval, M.-P., de Almeida Vau, M., Solaas, J., & Rodrigues, L. (2024a). Threat of deepfakes to the criminal justice system: A systematic review. *Crime Science*, *13*(1), 41.

Sandoval, M.-P., de Almeida Vau, M., Solaas, J., & Rodrigues, L. (2024b). Threat of deepfakes to the criminal justice system: A systematic review. *Crime Science*, *13*(1), 41.

Stoykova, R. (2021a). Digital evidence: Unaddressed threats to fairness and the presumption of innocence. *Computer Law & Security Review*, *42*, 105575.
Stoykova, R. (2021b). Digital evidence: Unaddressed threats to fairness and the presumption of innocence. *Computer Law & Security Review*, *42*, 105575.
Stoykova, R. (2023a). The right to a fair trial as a conceptual framework for digital evidence rules in criminal investigations. *Computer Law & Security Review*, *49*, 105801.
Stoykova, R. (2023b). The right to a fair trial as a conceptual framework for digital evidence rules in criminal investigations. *Computer Law & Security Review*, *49*, 105801.
Stoykova, R. A. (2024a). A new right to procedural accuracy: A governance model for digital evidence in criminal proceedings. *Computer Law & Security Review*, *55*, 106040.
Stoykova, R. A. (2024b). A new right to procedural accuracy: A governance model for digital evidence in criminal proceedings. *Computer Law & Security Review*, *55*, 106040.
Tahir, M. I., Saeed, A., Murtaza, M., & Abdullah, H. O. (2025). The Impact of Artificial Intelligence on Evidence in Criminal Trials. *Social Science Review Archives*, *3*(3), 1882–1894.