Πάντειον Πανεπιστήμιο Κοινωνικών & Πολιτικών Επιστημών
ΣΧΟΛΗ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΤΜΗΜΑ ΚΟΙΝΩΝΙΚΗΣ ΑΝΘΡΩΠΟΛΟΓΙΑΣ
Panteion University of Social and Political Sciences
SCHOOL OF SOCIAL SCIENCES DEPARTMENT OF SOCIAL ANTHROPOLOGY

Digital Evidence and Due Process: Are Criminal Trials in the Age of AI Still Fair?

Zoi Kaloudi, 1424M016

10016, ΨΗΦΙΑΚΟΣ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ: ΔΙΠΛΩΜΑΤΙΑ, ΕΠΙΚΟΙΝΩΝΙΑ, ΔΙΚΑΙΟ (DIGI-DCL)

ABSTRACT

The accelerating integration of artificial intelligence (AI) into criminal justice systems has begun to reshape not only investigative practices but also the structure of evidentiary evaluation within criminal proceedings. Tools such as probabilistic DNA analysis, forensic pattern-recognition systems, and automated data-analysis software increasingly participate in the production and interpretation of legally relevant facts. While these technologies promise efficiency and analytical capacity, their growing influence also raises fundamental concerns regarding transparency, contestability, and the conditions under which legal proof is established.

This dissertation examines whether AI-mediated forms of evidentiary evaluation can be reconciled with the procedural guarantees enshrined in Article 6 of the European Convention on Human Rights, the Charter of Fundamental Rights of the European Union, and the emerging EU regulatory framework on artificial intelligence. Rather than approaching AI as a neutral technical instrument, the study treats it as an epistemic actor whose modes of inference may sit uneasily with the requirements of adversarial procedure and reasoned judicial decision-making. Through a combined doctrinal and conceptual analysis, informed by critical algorithm studies and sociotechnical systems theory, the dissertation explores three core tensions: algorithmic opacity and the right to challenge evidence; the relationship between probabilistic inference and the criminal standard of proof; and the shifting balance between machine-generated outputs and human judicial deliberation. The analysis raises serious doubts as to whether certain forms of AI-based evidence can be reconciled with the normative logic of criminal proof. It suggests that, in the absence of robust standards of epistemic contestability, transparency, and human accountability, important limits emerge to the compatibility between AI-mediated evidentiary evaluation and the right to a fair trial.

TABLE OF CONTENTS

1: INTRODUCTION

1.1 Digital Justice and the Rise of AI in Criminal Procedure

The digitalization of criminal justice systems across Europe marks one of the deepest structural transformations since the codification of modern legal institutions. Artificial intelligence (AI) has entered this terrain not as an external innovation but as an internal mechanism increasingly woven into processes of investigation, adjudication, and evidentiary reasoning. Predictive policing algorithms, forensic data-mining tools, and probabilistic DNA interpretation systems now participate in what was once a purely human act of judgment. They do not merely assist the police in tracing suspects or managing caseloads, they help determine what counts as reliable evidence and, in some cases, shape the paths leading toward conviction or freedom.

The promise of these technologies' rests on efficiency and objectivity. By processing vast datasets and identifying hidden correlations, AI tools appear to deliver neutral, data-driven insights free from human bias (Hildebrandt, 2020). Yet at the same time their operation simultaneously destabilizes foundational legal assumptions: transparency, accountability, and contestability. When the internal logic of a machine-learning model remains opaque to judges or defense lawyers, the evidentiary process risks becoming an act of deference to technical authority rather than deliberative reasoning (Burrell, 2016). What emerges, then, is a paradox: digital justice as both an emancipatory horizon and a potential erosion of due process. Examples of this paradox already appear at the European legal landscape. Systems such as the Netherlands' SyRI welfare, fraud algorithm, the United Kingdom's probabilistic DNA evaluators, or Germany's digital forensic analytics software illustrate the growing reliance on automated assessments in evidentiary stages. These tools are often embedded in proprietary infrastructures, shielded from scrutiny by trade, secret protections, and technical complexity. As Yeung (2018) observes, algorithmic systems in the public sector frequently operate through "regulatory intermediaries" whose decision logic remains inaccessible to those subjected to it. Within the criminal courtroom, this opacity directly challenges the defendant's right to understand and contest the evidence presented against them.

From what I've examined so far not merely technological modernization /digitalization of the public sphere in general, but the epistemic authority of the law itself is at stake. If courts begin to treat algorithmic outputs as self-validating truths, they risk displacing the juridical ideal of reasoned judgment with what Danielle Citron (2019) calls "automation bias", the human tendency to over-trust computational results. As Mireille Hildebrandt (2015) warned nearly a decade ago, "the rule of law cannot survive if legal decision-making becomes a black box." The dissertation therefore takes as its starting point a pressing question: can AI-based evaluation of evidence coexist with the procedural and epistemic guarantees of a fair trial

1.2 The Problem: Can AI-Based Evaluation of Evidence Ensure a Fair Trial?

European criminal law is built upon a procedural architecture designed to preserve fairness through human deliberation. The right to a fair trial under Article 6 of the European Convention on Human Rights (ECHR) presupposes transparency, equality of arms, and the ability of the accused to challenge the evidence presented. AI-driven systems, however, often intervene at precisely those moments where contestability is most fragile: the weighing of probabilistic results, the assessment of credibility, and the interpretation of digital traces. In essence, the core question of this dissertation evaluates whether AI-based evaluative tools are normatively compatible with Article 6 due process guarantees in the legal system as we know it today.

The central problem is one of translation: how legal reasoning, which relies on normative interpretation, can meaningfully engage with algorithmic reasoning, which operates through statistical inference. The law seeks justification, the algorithm offers prediction. When a probabilistic classifier assigns a 93 percent likelihood that a digital trace matches a suspect, what epistemic status does that figure hold within the standard of "beyond reasonable doubt"? This conceptual dissonance forms the core of the inquiry of this dissertation.

Beyond epistemology lies a deeper democratic concern. The automation of evidentiary evaluation risks creating what Pasquale (2015) famously called a "black box society," where decisions with profound consequences for liberty are rendered inscrutable. If defendants cannot access the model's internal parameters, and judges cannot explain why an algorithm's finding is persuasive, procedural fairness becomes performative rather than substantive. The legal system thus faces a dual imperative: to harness technological innovation while preserving its own conditions of legitimacy.

This dissertation does not examine fully automated judicial decision-making or the replacement of judges by artificial intelligence. Its focus is narrower and more specific: the use of AI-mediated tools in the evaluation of evidence within criminal proceedings. Judicial discretion formally remains human, yet the increasing reliance on algorithmic outputs raises concerns about how evidentiary authority is constructed, interpreted, and contested. The central problem addressed is therefore not the automation of adjudication as such, but the epistemic and procedural consequences of introducing opaque, probabilistic, and proprietary systems into the evidentiary process.

1.3 Research Objectives and Scope

This dissertation aims to critically assess whether AI-driven evaluative mechanisms used in criminal evidence assessment are compatible with European principles of due process and fair trial. It focuses on the evaluation rather than collection of evidence; that is, the stage where data transforms into legally meaningful proof. The objectives are fourfold:

1. To map the primary categories of AI systems currently influencing evidentiary reasoning in Europe.
2. To examine the conceptual and doctrinal tensions these systems generate within the framework of human rights law.
3. To analyze how algorithmic opacity affects procedural safeguards such as contestability, equality of arms, and the presumption of innocence.
4. To propose normative principles for integrating AI tools into evidence of evaluation without compromising justice.

The scope remains confined to European jurisdictions governed by the ECHR and EU legal order, incorporating select comparative references to highlight conceptual contrasts rather than jurisdictional detail.

1.4 Research Questions

From these objectives flow several guiding questions:

1. How do AI tools reshape the epistemic foundations of evidentiary reasoning in criminal trials?
2. To what extent can algorithmic inference satisfy the standard of proof required by law?
3. Is existing human rights framework, particularly Article 6 ECHR and Articles 47,48 of the EU Charter, adequate to safeguard defendants' rights against opaque technological mediation?
4. What institutional and procedural reforms might reconcile the efficiency of AI with the normative integrity of due process?

1.5 Relevance and Justification

The relevance of this research lies in the convergence of two trajectories: the acceleration of digital governance and the constitutionalizing of human rights within the European legal space. As states turn to data-driven governance, courts become laboratories for testing the limits of algorithmic authority. The European Court of Human Rights has already encountered cases implicating digital surveillance and automated decision-making, though it has yet to structure a coherent doctrine for algorithmic evidence. Anticipating this jurisprudential evolution is not merely an academic exercise but a legal necessity.

Moreover, the dissertation contributes to an emerging field sometimes described as digital constitutionalism, where legal scholars such as De Gregorio (2022) argue that the rule of law must be reanchored in the algorithmic era. It also draws on critical work by legal theorists who challenge the technocratic ideal of neutrality. As Latour (2010) reminds us, technologies are not external tools but "delegations of agency" that reshape the very form of social and institutional order. Recognizing this helps move beyond a binary of acceptance or resistance toward a nuanced understanding of how AI becomes internal to the grammar of justice.

1.6 Methodological Approach

The research adopts a conceptual, doctrinal methodology, enriched by insights from critical algorithm studies and sociotechnical systems theory. Rather than conducting empirical fieldwork, it engages in theoretical synthesis across jurisprudence, philosophy of technology, and regulatory analysis. The method proceeds through three interlinked levels:

1. Doctrinal analysis, examining how current European legal instruments (ECHR, EU Charter, GDPR, and AI Act) articulate procedural guarantees relevant to algorithmic evidence.
2. Conceptual critique, interrogating underlying assumptions about knowledge, objectivity, and decision-making that accompany the legal reception of AI.
3. Normative synthesis, proposing principles for recalibrating due process in light of machine-mediated evaluation.

The conceptual focus does not preclude attention to real, world examples rather; it situates them as illustrations of broader tensions between legal rationality and computational logic. This dissertation does not aim to provide a comprehensive empirical assessment of specific AI systems, but rather to interrogate the conditions under which such systems may acquire evidentiary authority within criminal proceedings.

2. LITERATURE REVIEW AND DOCTRINAL POSITIONING

Over the last decade, artificial intelligence has moved from the margins to the core of contemporary criminal justice systems. This shift has generated a rapidly expanding body of academic literature, spanning computer science, criminology, legal theory, and regulatory studies. Yet despite its volume, this literature remains fragmented. Technical scholarship tend to emphasize efficiency, predictive accuracy, and system performance, while legal scholarship has focused primarily on data protection, privacy, and discrimination.

What is notably underdeveloped, however, is a sustained examination of how AI affects the logic of criminal proof itself. In particular, the implications of AI-mediated evidentiary evaluation for adversarial procedure, standards of proof, and the defendant's ability to contest evidence remain insufficiently theorized.

This dissertation positions itself within that gap. Rather than treating AI as a peripheral administrative tool, it approaches algorithmic systems as epistemic actors that increasingly shape how facts are produced, interpreted, and validated in criminal trials. The literature reviewed below is therefore organized not simply by technology or legal instrument, but by the specific tensions AI introduces into the structure of criminal adjudication.

2.1 AI Technologies in Contemporary Criminal Justice Systems

Early academic engagement with AI in criminal justice framed algorithmic systems largely as instruments of modernization. Scholars highlighted their potential to enhance efficiency, improve resource allocation, and support decision-making in policing and prosecution (Ferguson, 2017; UNODC, 2020). Predictive policing tools, automated risk assessments, and forensic pattern-recognition systems were often presented as neutral technologies capable of optimizing existing practices.

As these tools moved from pilot projects into operational use, scholarly attention shifted. Empirical studies, particularly in the United States, exposed how algorithmic systems used in bail, sentencing, and parole decisions could produce systematically skewed outcomes (Angwin et al., 2016; Dressel & Farid, 2018). These findings destabilized the assumption that algorithmic decision-making is inherently objective, redirecting the literature toward questions of fairness, accountability, and bias.

European scholarship developed along a somewhat different orbit. Rather than focusing primarily on predictive accuracy, European legal scholars situated AI within broader frameworks of fundamental rights, data protection, and constitutional governance (Hildebrandt, 2015; Veale & Borgesius, 2021). Law-enforcement databases, biometric identification systems, and large-scale data analytics were analyzed as infrastructures of power rather than mere technical tools.

Despite these advances, much of the literature continues to examine AI applications in isolation. Policing tools, forensic algorithms, and risk assessments are often treated as separate phenomena, without sufficient attention to their cumulative impact on criminal proceedings as a whole. This fragmentation obscures the ways in which AI systems jointly reshape evidentiary reasoning across the criminal process.

2.2 Epistemic Opacity, Explainability, and Algorithmic Accountability

One of the most persistent themes in the literature concerns epistemic opacity. Many AI systems, particularly those based on machine-learning architectures, generate outputs through processes that are not readily intelligible even to their designers (Burrell, 2016). In legal contexts, this opacity becomes more than a technical inconvenience; it directly affects the legitimacy of evidentiary reasoning.

Criminal proceedings presuppose that evidence can be examined, challenged, and tested through adversarial procedures. Pasquale's notion of the "black box society" captures the tension that arises when legally decisive outcomes are produced by systems shielded from meaningful scrutiny (Pasquale, 2015). If neither judges nor defense lawyers can access or understand the internal logic of an algorithm, the possibility of effective contestation is significantly weakened.

The literature on explainable AI has attempted to address this problem by developing models that offer post hoc explanations for algorithmic outputs. Legal scholars, however, remain skeptical. Explanations generated after the fact may provide surface-level narratives without revealing the causal structure of decision-making (Hildebrandt, 2015). From a procedural perspective, such explanations risk functioning as legitimizing devices rather than genuine tools of scrutiny. Opacity also produces accountability gaps. Responsibility for algorithmic decisions is dispersed across developers, data providers, public authorities, and end-users (Kroll et al., 2017). Traditional legal concepts of fault and liability struggle to accommodate this distributed architecture. As a result, errors affecting evidentiary outcomes may remain institutionally unaccounted for.

2.3 Algorithmic Bias and Structural Inequality

Alongside opacity, algorithmic bias has emerged as one of the most extensively documented risks associated with AI in criminal justice. Empirical research has demonstrated that algorithmic systems frequently reproduce existing social inequalities, particularly when trained on historically biased data (Barocas & Selbst, 2016; Selbst et al., 2019).

In criminal justice contexts, the consequences of such bias are especially severe. Algorithmic outputs may influence investigative priorities, evidentiary weighting, or assessments of credibility. What appears as statistical correlation can easily translate into legally consequential inference. The literature emphasizes that biased outputs often acquire heightened epistemic authority precisely because they are produced by automated systems, making them harder to challenge in court.

European legal scholarship has approached algorithmic bias primarily through principles of equality before the law and non-discrimination. Yet beyond distributive outcomes, bias also generates procedural distortions. When defendants lack access to the data or logic underlying algorithmic assessments, their ability to contest evidence is compromised. Bias thus operates not only at the level of outcomes, but at the level of procedural fairness itself.

2.4 Digital Evidence and the Theory of Criminal Proof

The integration of AI into criminal justice reconfigures the very concept of digital evidence. While earlier legal discussions treated digital evidence as a technical extension of traditional documentary proof, more recent scholarship recognizes that algorithmically generated outputs raise distinct epistemic challenges (Casey, 2011; Mason & Seng, 2017).

Criminal proof has traditionally been understood not simply as a matter of factual accuracy, but as a structured procedural practice governed by normative commitments. As Ashworth and Redmayne observe, the criminal trial is not designed to identify truth at any cost, but to do so under conditions that respect fairness, accountability, and the moral legitimacy of state punishment (Ashworth & Redmayne, 2021). Evidence, in this sense, acquires its probative value not only from its reliability, but from the manner in which it is produced, presented, and tested within adversarial proceedings. This framework already signals a point of tension with algorithmic systems whose internal reasoning may remain inaccessible or only partially intelligible.

Within this tradition, proof is not reducible to probability. While probabilistic reasoning has long played a role in evidentiary assessment, criminal adjudication ultimately requires a form of reasoned judgment that cannot be exhausted by numerical outputs alone. The standard of proof “beyond reasonable doubt” operates as a normative threshold, not a statistical one (Ashworth & Redmayne, 2021). It presupposes that judges can articulate why evidence persuades them, and that such reasoning remains open to challenge. Algorithmically generated outputs, however, are often expressed precisely in probabilistic terms, creating the risk that quantified likelihoods acquire undue authority simply because they appear mathematically precise.

The adversarial structure of criminal proceedings further complicates the integration of algorithmic evidence. As Roberts and Zuckerman emphasize, adversarial justice depends on the capacity of the defense to scrutinize not only evidentiary conclusions, but also the methods through which those conclusions are reached (Roberts & Zuckerman, 2010). Effective challenge presupposes access to the reasoning process behind the evidence. When evidentiary weight is attached to outputs generated by proprietary or opaque algorithms, this condition is undermined. The defense may be confronted with an evidentiary object whose probative force it cannot meaningfully test, transforming adversarial contestation into a largely formal exercise.

This difficulty is not merely technical, but structural. Adversarial procedure allocates epistemic responsibility across the parties and the court, ensuring that factual claims are exposed to contradiction and critical assessment (Roberts & Zuckerman, 2010). When algorithmic systems function as epistemic intermediaries, producing conclusions without transparent reasoning, responsibility becomes diffused. The prosecution may rely on the authority of the system, while the defense lacks the tools necessary to challenge it. In such circumstances, the balance that adversarial procedure seeks to maintain is quietly destabilized.

Jackson and Summers have further shown that advisability serves not only evidentiary efficiency, but human rights protection. The ability to contest evidence is inseparable from the legitimacy of criminal conviction (Jackson & Summers, 2012). Where decisive elements of evidentiary reasoning are effectively insulated from challenge, the right to a fair trial is placed under strain, even if formal procedural guarantees remain intact. Algorithmic evidence thus raises a fundamental question: can proof still be considered the outcome of adversarial testing when its underlying logic cannot be examined by those affected?

Taken together, these strands of criminal-procedural theory reveal that the challenges posed by AI-mediated evidence are not external to the law of proof, but internal to it. The difficulty is not simply that algorithmic systems may err, but that they may do so in ways that evade the procedural mechanisms through which error is ordinarily exposed. This creates a growing mismatch between the normative architecture of criminal proof and the epistemic characteristics of algorithmic inference, a

tension that becomes particularly acute once such systems are introduced into the courtroom as evidentiary authorities rather than investigative aids.

Algorithmic outputs occupy an ambiguous position within the traditional typology of evidence. They are neither testimonial nor purely material. Instead, they function as hybrid epistemic artefacts, produced through statistical inference rather than direct human observation. This hybridity complicates their integration into criminal proceedings, which presuppose that evidentiary value can ultimately be traced back to human sources capable of explanation and cross-examination. The principle of free evaluation of evidence further intensifies this tension. Judicial discretion presupposes that judges can understand and critically assess the evidence before them. When evidentiary weight is attached to outputs whose internal reasoning is inaccessible, judicial evaluation risks becoming an act of deference rather than deliberation (Hildebrandt, 2015).

The literature also highlights the persuasive force of probabilistic reasoning. Numerical outputs may appear precise and authoritative, potentially overshadowing normative judgments that remain central to criminal proof. This raises concerns about whether algorithmic evidence subtly reshapes standards of proof by introducing quasi-mathematical logics into a domain traditionally governed by reasoned judgment.

2.5 Fair Trial Guarantees and Algorithmic Evidence under Article 6 ECHR

Article 6 ECHR constitutes the cornerstone of fair trial guarantees in Europe. Its requirements of adversarial proceedings, equality of arms, and reasoned judgments were developed in a legal environment shaped by human-generated evidence. The introduction of algorithmic systems tests the adaptability of these guarantees.

Strasbourg jurisprudence consistently emphasizes that defendants must be able to understand and challenge the evidence against them. When algorithmic systems are used by prosecution while remaining opaque to the defense, structural asymmetries arise. Formal equality between the parties' risks being transformed into substantive inequality. The requirement of reasoned judgments presents an additional challenge. If judicial decisions rely heavily on algorithmic outputs, courts must still articulate reasons that demonstrate genuine engagement with the evidence. The literature warns that uncritical reliance on algorithmic results may lead to a form of algorithmic deference, undermining the rationale-giving function of judicial decisions.

2.6 The AI Act and the Regulatory Framework for Law-Enforcement AI

The EU Artificial Intelligence Act represents a significant shift in the governance of algorithmic systems. By classifying many law-enforcement AI tools as high-risk, the regulation acknowledges their potential to affect fundamental rights. Transparency, risk management, and human oversight obligations are central features of this framework.

Nevertheless, the literature highlights a structural limitation. The AI Act operates primarily at the level of market regulation and system design. It does not directly address how algorithmic outputs should be assessed, challenged, or weighted as evidence within individual criminal proceedings. Compliance with regulatory standards does not automatically ensure compliance with fair trial guarantees.

This regulatory–procedural gap has led scholars to question whether ex ante compliance mechanisms can adequately substitute for procedural safeguards at trial. The concern is that courts may treat regulatory compliance as a proxy for evidentiary reliability, effectively shifting the burden onto defendants to demonstrate unreliability.

2.7 Research Gap and Positioning of the Present Study

The literature on AI in criminal justice has generated valuable insights into efficiency, bias, transparency, and regulation. Yet comparatively little attention has been devoted to the evidentiary status of algorithmic outputs within criminal trials themselves. In particular, the interaction between AI-mediated inference and the procedural guarantees of Article 6 ECHR remains under-theorized.

This dissertation addresses that gap by examining AI as an epistemic contributor to criminal proof. It focuses on the conditions under which algorithmic evidence can be meaningfully contested, understood, and evaluated within adversarial proceedings. By foregrounding the concept of epistemic contestability, the study situates itself at the intersection of criminal procedure, human rights law, and critical algorithm studies. Rather than asking whether AI can be regulated effectively in the abstract, the dissertation asks a more specific and pressing question: under what conditions, if any, can AI-mediated evidentiary evaluation be reconciled with the normative structure of a fair criminal trial?

This dissertation approaches artificial intelligence neither as an inherently harmful intrusion into criminal justice nor as a neutral instrument whose legitimacy follows automatically from technical performance. The author's engagement with the

field is shaped by an educational background grounded in law, digital governance, and human rights, which informs a deliberately critical, yet non-rejectionist, stance toward technological integration. AI is treated as a tool that may be used within criminal proceedings, but only under conditions of strict legal constraint, procedural transparency, and meaningful contestability.

Rather than opposing the use of algorithmic systems as such, the analysis proceeds from the premise that technological adoption in criminal justice must remain subordinated to the normative structure of the fair trial. Efficiency, consistency, or predictive capacity cannot operate as independent justifications for evidentiary use. Where AI-mediated outputs cannot be understood, challenged, or effectively scrutinized by the defense and the court, their use becomes incompatible with the foundational principles of due process. The critical orientation adopted here is therefore not anti-technology, but jurisprudential: it insists that legal legitimacy must precede, and condition, technological deployment.

3. THEORETICAL AND CONCEPTUAL FRAMEWORK

3.1 The Rule of Law and Fair Trial Theory

The challenges outlined in the introduction of opacity, accountability gaps, and the reconfiguration of evidentiary authority, demand a strong theoretical foundation before any legal analysis can proceed on them. Chapter 2 will set up the three main pillars of the conceptual lens through which the legal and ethical analysis of AI-based evidence evaluation is conducted, and they are as follows.

First and foremost, the Legal-Doctrinal Lens focuses on how existing legal frameworks apply to AI-generated evidence. Key considerations here include admissibility standards, such as the rules of evidence governing relevance, reliability, and probative value, as well as due process and procedural fairness, which ensure that defendants' rights to confront evidence and challenge expert testimony are preserved. Equally important is accountability and liability; that is, determining who bears responsibility if AI errors affect case outcomes, whether programmers, forensic analysts, or institutions. Essentially, this lens asks: Can the law accommodate AI as a legitimate source of evidence, and under what limitations?

The second vital pillar is the Ethical Lens, which focuses on normative questions concerning justice, fairness, and societal trust. From my perspective, the primary ethical concerns include bias and discrimination, as AI can reproduce or exacerbate systemic biases, disproportionately affecting minority or marginalized groups; transparency and explainability, given that the "black box" nature of many AI models challenges ethical obligations to provide clear reasoning for legal decisions; and, notably in my opinion, autonomy and human oversight, which addresses whether human judges or juries can meaningfully evaluate AI recommendations. This lens therefore asks: Is it morally acceptable to rely on AI in evidence of evaluation, and under what safeguards?

Last but not least, the Sociotechnical Lens, increasingly emphasized in contemporary scholarship, examines the interaction between AI technologies and legal institutions. The main factors identified in academic research include institutional readiness, or whether courts can handle the technical complexity of AI; public perception, addressing whether AI use in trials enhances or undermines trust in justice; and regulatory alignment, which considers whether procedural norms keep pace with rapid technological developments.

This mosaic leads to Conceptual Synthesis. The overarching lens is often described as a "normative, legal, technical" lens, in which AI is examined through the prisms of law (what is permissible), ethics (what is just), and technical reliability (what is valid). Scholars frequently frame this analysis in terms of accountability, fairness, and transparency, integrating legal doctrine with philosophical ethics and technical assessment.

3.2 The Rule of Law and Fair Trial Theory

In this chapter, I will seek to explore the intersection between the principles of the rule of law and the theory of fair trial. The central inquiry concerns how these foundational legal doctrines can be preserved, protected, and effectively maintained in an era where decision-making processes increasingly involve artificial intelligence. More specifically, the question arises: how can the fundamental right to a fair trial and the corresponding right to challenge and dispute the outcome be upheld when one finds oneself effectively in contention with an invisible, algorithmic machine? Due process stands as one of the fundamental pillars of any modern criminal justice system. It serves not only as a procedural guarantee but also as a moral safeguard ensuring that justice is both done and seen to be done. The idea that every person is entitled to a fair trial, impartial adjudication, and the opportunity to challenge the evidence against them lies at the very core of democratic legality (Ashworth & Redmayne, 2021). In my opinion, due process represents the moral compass of criminal law, and it demands that no conviction or punishment occurs unless it follows a process grounded in fairness, transparency, and reasoned deliberation.

In competitive or adversarial legal systems, fairness is achieved through what might be described as a balance of contest. Both the prosecution and the defense are allowed to present their evidence, question witnesses, and test the credibility of the opposing case. This balance ensures that truth emerges not from authority, but from structured debate (Packer, 1968). Through my research, I have come to appreciate that fairness in such systems is not merely a formal requirement, it is what gives legitimacy to the final judgment. Without procedural equality, even a factually correct verdict risks being perceived as unjust.

A related and equally essential component of due process is legal certainty and foreseeability. Individuals must be able to predict, to some reasonable degree, the legal consequences of their actions and understand how evidence will be assessed in court (Craig, 2018). This principle prevents arbitrary decisions and ensures that justice operates within known and consistent parameters. In the evidentiary context, foreseeability means that the methods used to gather, process, and interpret evidence must be transparent and subject to scrutiny. When new technologies such as AI are introduced into this framework, they must therefore meet these same expectations. If an AI system produces results that cannot be explained or verified, this threatens the predictability that legal certainty depends upon.

Another key principle flowing from due process is the equality of arms, a notion developed by the European Court of Human Rights under Article 6 of the European Convention on Human Rights. It requires that both parties in a trial have a fair opportunity to present their case and challenge that of the opponent (ECtHR, 1993). This includes the right to know and contest the evidence being used. However, when AI-generated evidence or algorithmic analysis is introduced, this right can become difficult to exercise. For instance, if a facial recognition system identifies a suspect, but the defense cannot access the system's training data or algorithmic logic, then effective cross-examination becomes almost impossible. In my view, this undermines the equality of arms, because one side possesses technological knowledge that the other cannot realistically counter.

Balanced access to information is therefore essential for maintaining fairness in trials. Transparency does not solely mean revealing results; it requires disclosing the reasoning process behind those results. Through my research, I have noticed growing concern among scholars and practitioners that AI systems used in criminal investigations such as predictive policing tools or forensic algorithms are often shielded by intellectual property rights or technical secrecy (Larsson & Heintz, 2020). This lack of openness limits the ability of courts and defense lawyers to evaluate the reliability of such evidence. As a result, there is a real danger that defendants could be convicted on the basis of evidence they cannot meaningfully contest or even understand.

The final but perhaps most profound issue concerns the role of human judgment in legal decision making. Law has traditionally relied on human reasoning, deliberation, and moral evaluation because it deals not only with facts but with values. Judges and juries interpret meaning, intent, and credibility tasks that inherently involve empathy and ethical consideration. Machines, however advanced, do not share this human sensibility. They process data according to pre-defined patterns but cannot appreciate the moral weight of human actions or the contextual subtleties of justice (Leib, 2019). In my opinion, this is why the complete outsourcing of evidentiary evaluation to AI should be approached with extreme caution.

Delegating parts of evidentiary analysis to machines might increase efficiency, but it also shifts responsibility. If an AI system errors, who is to blame the developer, the operator, or the court that relied on it? The diffusion of accountability could erode public confidence in justice itself. Moreover, as algorithmic systems become more complex, their influence may subtly reshape legal reasoning, prioritizing statistical probability over moral reasoning (Binns, 2018). Through my research, I have come to believe that technology should assist but never replace human judgment. The legitimacy of criminal adjudication rests on the perception that decisions are made by people who can be questioned, persuaded, and held morally responsible.

In conclusion, due process remains the cornerstone of criminal justice precisely because it embodies fairness, equality, and accountability. The growing role of AI challenges each of these principles in subtle but significant ways. While artificial intelligence can enhance accuracy and efficiency, its integration into the evidentiary process must not compromise transparency, contestability, or the human element that defines justice itself. As the law continues to evolve, it must do so without surrendering the essential human values that give it meaning.

3.3 Critical Algorithm Studies

Algorithmic Decision-Making as a "Black Box". A central concern in discussions about the use of artificial intelligence in the criminal justice system is the so-called "black box" character of many algorithmic decision-making systems. Machine-learning models, especially deep learning systems, often operate through layers of mathematical transformations that even their developers might struggle to fully interpret (Burrell, 2016). This lack of transparency presents a serious challenge in legal contexts where decisions must be reasoned, explainable, and open to scrutiny. In the courtroom, the inability to clarify how an algorithm reached a particular output may directly conflict with due process requirements and the defendant's right to challenge evidence. In my opinion, a legal decision that cannot be properly justified is already on ethically shaky ground, and delegating such opaque processes to machines only deepens the problem.

In more traditional forms of expert evidence, courts can question the expert, assess the methodology used, and evaluate whether the conclusion logically follows from the scientific reasoning provided. With algorithmic systems, however, the reasoning steps are often inaccessible. They may be proprietary, too complex, or even impossible to break down into a human-readable explanation (Doshi-Velez & Kim, 2017). Through my research, I found that this raises real challenges when AI-generated outputs, such as risk scores, facial recognition matches, or predictive assessments, are presented as evidence. Without transparency, the defense cannot meaningfully interrogate or rebut the conclusions. This, in effect, undermines the equality of arms and the adversarial principle that each party must be able to contest the claims of the other side.

Explainability matters tremendously in the law, not only for fairness but also for legitimacy. Courts must show why they reached out a particular judgment, and parties must be able to understand how evidence was evaluated. When algorithmic systems operate in ways that cannot be articulated in human terms, they jeopardize this fundamental requirement (Wachter et al., 2017). For example, if a judge relies on an algorithmic risk assessment tool to justify denying bail, and the tool's inner logic cannot be disclosed or explained, the defendant is placed in a position where challenging the basis of the decision becomes nearly impossible.

In my opinion, this lack of explainability also threatens legal foreseeability. Parties should be able to anticipate how evidence will be weighed and what factors influence judicial outcomes. An opaque algorithm disrupts this expectation by hiding the evaluative criteria behind a curtain of technical complexity. And while there are efforts to create "interpretable" machine-learning models, these often come at the cost of predictive accuracy, or at least are claimed to, which becomes part of the problem.

Another major issue concerns bias embedded within AI systems. While algorithms are sometimes promoted as objective or neutral, numerous studies have shown that they can reproduce and even amplify existing societal biases (O'Neil, 2016). This happens primarily through biased training data, flawed system design, or unrecognized assumptions encoded into the model. For instance, a predictive policing system trained on historical arrest data may disproportionately target communities that have already been overpoliced, thereby reinforcing discriminatory outcomes.

Through my research, I noticed that this becomes especially troubling when AI tools are used to evaluate or produce evidence. Selective weighting of factors such as postcode, prior arrests, or demographic data may produce correlations that appear statistically strong but are legally prejudicial. For example, an AI system might correlate certain neighborhoods with higher crime risk, not because individuals there are more likely to offend but simply because of historical policing patterns. If such outputs are introduced into court without critical examination, they can distort evidentiary evaluation by smuggling in biased assumptions under the guise of scientific analysis.

In my opinion, one of the greatest dangers is that biased outputs can appear legitimate precisely because they are generated by an algorithm. Judges, lawyers, and jury members might assume that the "machine" is objective, when in reality it is merely reflecting the biases embedded in its data and design. This illusion of neutrality can be far more harmful than open discrimination, because it is harder to detect and challenge.

Automation bias refers to the human tendency to trust or overvalue the conclusions of automated systems, especially when they appear sophisticated or authoritative (Skitka, Mosier, & Burdick, 1999). In institutional settings such as courts or police departments, this bias can lead actors to rely excessively on algorithmic assessments without adequately questioning their accuracy or limitations.

Judges, for instance, may view algorithmic outputs as more precise or reliable than traditional human judgments. This effect is amplified by institutional pressure to appear efficient, consistent, and evidence based. When an AI system presents a neat numerical assessment, such as a 78% likelihood of reoffending, it may carry an aura of scientific credibility that discourages deeper scrutiny. Through my research, I have found that this problem is not only psychological but also structural: institutions may adopt AI tools with the implicit assumption that automation reduces error, when in fact it simply obscures different kinds of errors.

The danger, in my opinion, is that such overreliance can lead to situations where human decision-makers defer to algorithmic recommendations even when those recommendations conflict with contextual judgment or common sense. As a result, judicial reasoning may become indirectly shaped by machine logic, even if the final decision formally rests with a human judge. This creates a subtle but significant shift in how evidence is weighed, how risk is interpreted, and ultimately how justice is administered.

In summary, algorithmic decision-making raises profound legal concerns related to transparency, explainability, bias, and human overreliance. While AI tools promise efficiency and analytical power, they also introduce new risks that threaten the fairness and integrity of the judicial process. When the reasoning behind an algorithm cannot be understood, when its inputs

are biased, or when its outputs are accepted uncritically, the right to a fair trial is put at risk. Ensuring accountability, transparency, and meaningful human oversight is therefore essential if AI is to play any legitimate role in the evaluation of evidence.

A well-known illustration of these risks can be seen in the United States with the COMPAS risk assessment tool, which was used in criminal sentencing. In the case of State v. Loomis (2016), the defendant challenged the fact that his sentence had been influenced by an algorithm whose inner workings were completely opaque. In my opinion, the most troubling part was that neither the defense nor the court could examine how the tool produced the individual "risk score," since the company behind COMPAS claimed its model was a trade secret. Through my research I found that COMPAS has also been criticized for reproducing systemic biases, particularly racial ones, which makes its appearance a supposedly "objective" tool even more problematic. I believe this example clearly shows what it means to introduce a black box algorithm into a criminal proceeding: the defense cannot meaningfully challenge it, the judge cannot fully explain it, and the whole process risks losing the transparency and fairness it is supposed to guarantee.

3.4 Sociotechnical Systems Theory

Sociotechnical Systems Theory provides the third major pillar for understanding how AI becomes embedded within the administration of justice. Unlike strictly legal or ethical approaches, this framework treats technologies not as neutral instruments but as entities shaped by, and simultaneously shaping, the social and institutional structures in which they operate. Its core premise, that technological systems and social arrangements coproduce one another, is analytically powerful, yet also raises complicating questions. If we accept that AI is not an external tool but part of a broader ecosystem of people, practices, norms, and infrastructures, then any assessment of its evidentiary role must look beyond questions of accuracy or doctrinal fit.

The first implication concerns institutional readiness. Courts often lack the technical expertise, procedural mechanisms, or resource capacity needed to meaningfully scrutinize algorithmic systems. It is tempting to assume that, because AI tools offer efficiency or enhanced analytic capability, their integration into judicial processes is inherently desirable. A sociotechnical perspective challenges this intuition: are institutions structurally equipped to absorb such tools without displacing established norms of reasoning, evidentiary assessment, and judicial independence? One risk is that reliance on automated systems subtly reshapes adjudicative practice, shifting authority from human decisionmakers to computational outputs in ways that are neither formally acknowledged nor normatively justified.

A second dimension involves social legitimacy. Technologies do not function in a vacuum; they operate within environments in which public trust, cultural expectations, and collective anxieties profoundly influence how judicial authority is perceived. While one might argue that improved accuracy or consistency should suffice to legitimate the use of AI in criminal trials, a sociotechnical analysis underscores that legitimacy is relational rather than purely technical. Historical experiences with biased policing, unequal access to justice, or opaque state decision-making can affect whether the public interprets AI-mediated evidence such as enhancing impartiality or as deepening structural inequities. Thus, statistical performance does not automatically translate into democratic acceptance.

The third issue concerns regulatory adaptability. Legal systems evolve incrementally, whereas AI technologies change at a pace that exceeds traditional legislative or doctrinal cycles. Sociotechnical theory highlights how this temporal asymmetry creates conceptual friction: established legal categories such as "expert testimony," "forensic method," or "evidentiary reliability" may not map neatly onto machine-learning tools that operate through fundamentally different logics. One could counter that law has always adapted to technological shifts, from DNA analysis to digital forensics. However, the rate and depth of AI innovation suggest a more profound challenge, not simply updating doctrinal categories but rethinking how legal concepts capture systems that are dynamic, probabilistic, and continuously modified.

Taken together, Sociotechnical Systems Theory encourages a more nuanced understanding of AI as part of an interdependent network of actors, norms, and infrastructures. It does not aim to replace legal or ethical reasoning, but to reveal the conditions under which such reasoning becomes meaningful. The question, therefore, is not solely whether AI can operate reliably as an evidentiary tool; it is whether the broader justice ecosystem can integrate these technologies in a manner that preserves accountability, transparency, and the foundational principles of the rule of law.

3.5 Conceptualizing Digital Justice

The notion of digital justice has emerged as a key thematic anchor in contemporary debates about the digital transformation of legal systems. Yet despite its increasing visibility, the term remains conceptually fluid. Some scholars employ it to describe the modernization of judicial processes through technological means, while others use it to critique the ways in which digital infrastructures reshape power, rights, and institutional authority. This conceptual ambiguity is not a weakness; it reflects the fact that digital justice is best understood as a contested and evolving framework rather than a fixed doctrinal category.

At its most basic level, digital justice concerns the capacity of legal systems to uphold fairness, transparency, and accountability in environments saturated with algorithmic decision-making. However, this definition already rests on an implicit assumption: that these foundational principles can be preserved even as the modes of producing, evaluating, and interpreting evidence to undergo profound technological change. A more critical reading would ask whether digital justice requires merely adapting existing norms to new tools, or whether it demands rethinking what justice means when mediated through computational infrastructures.

The first dimension of this inquiry involves the redistribution of epistemic authority. Traditional conceptions of justice rely on the idea that fact-finding is a human exercise grounded in deliberation, reasoning, and contestation. AI-based tools challenge this norm by introducing forms of knowledge production that operate outside human interpretive control. Should digital justice accept algorithmic inference as a legitimate epistemic contributor to legal truth? Or does meaningful justice require that inferential processes remain intelligible and contestable to all parties? This tension lies at the heart of digital justice as both a practical and theoretical construct.

A second component concerns procedural equity. Digital tools are often presented as neutral mechanisms that enhance consistency, yet sociotechnical research repeatedly demonstrates that algorithmic systems can reproduce or amplify existing structural biases. A narrow view of digital justice might focus on minimizing technical bias, assuming that procedural fairness can be restored through better datasets or improved model design. A broader perspective, however, asks whether the very act of embedding algorithmic systems into legal processes alters the distribution of procedural burdens, potentially privileging institutions with access to technical expertise while disadvantaging defendants who cannot meaningfully interrogate the logic of automated assessments.

A third element relates to institutional legitimacy and democratic accountability. Justice is not solely an outcome, but a process that must be recognizable as fair by those subjected to it. Introducing opaque or proprietary AI systems into adjudication risks creating interpretive asymmetries: defendants, lawyers, and even judges may lack the knowledge necessary to understand the basis of a machine-generated conclusion. One could argue that if outcomes become more accurate, legitimacy naturally follows. A counterargument, grounded in critical theories of technology, holds that legitimacy requires not only reliable outputs but also transparency, explainability, and the preservation of human agency in decision-making.

Bringing these strands together, digital justice can be understood as a normative framework that evaluates how digital technologies reshape the conditions under which justice is produced, perceived, and contested. It is not simply "justice with digital tools," but a lens for examining how legal rights, procedural safeguards, and institutional practices interact with, and are transformed by, technological infrastructures. This conceptualization does not presuppose a pessimistic or optimistic view of AI's role in the courtroom; instead, it provides the analytical space to identify which aspects of justice remain stable, which require recalibration, and which may be fundamentally incompatible with algorithmic mediation.

3.6 Conclusion

Taken together, the different lenses explored in this chapter show that the introduction of AI into criminal adjudication is not a marginal development but one that presses on the basic conditions under which justice is produced. The legal-doctrinal analysis highlighted how evidentiary rules, originally crafted for human experts and human reasoning, strain when applied to systems whose inner workings cannot easily be made transparent. The ethical discussion suggested that fairness and bias cannot simply be "engineered away," at least not without careful reflection on the moral values the criminal process is meant to protect.

The sociotechnical perspective complicated the picture further, revealing how institutional capacity, public trust, and the pace of technological innovation influence what actually happens in practice. It is within this intersection that the idea of digital justice becomes useful, not as a fixed definition, but as a way of questioning how much technological change the procedural foundations of the trial can absorb without losing their purpose. The challenge is not simply to regulate AI well, but to identify which elements of adjudication must remain unambiguously human.

1. LEGAL AND REGULATORY FRAMEWORK

4.1 Introduction

The legal frameworks that are governing criminal trials in Europe present themselves as safeguards for procedural fairness, but whether they can withstand AI-based evidentiary reasoning remains deeply uncertain. In theory, Article 6 ECHR, the EU Charter, GDPR, and now the AI Act should collectively preserve the defendant's right to challenge the evidence against them. In practice, however, procedural guarantees may collapse when confronted with opaque algorithmic systems that courts cannot fully interpret, let alone contest.

In my view, the central question is not simply whether these legal instruments offer protection, but whether they can do so under conditions of epistemic asymmetry, where neither the defendant nor the judge can fully understand the AI system that shaped the outcome. This raises a troubling possibility: European law may already contain the seeds of its own erosion, if it assumes that legal concepts, like human oversight, proportionality, or evidentiary reliability, can simply be transposed onto algorithmic systems without conceptual modification.

A parallel line of reasoning emerges from Digital Rights Ireland (CJEU, 2014), where the Court struck down blanket data retention on the grounds that state access to digital information requires strict safeguards, proportionality assessment, and meaningful avenues for redress. Although the case did not address evidentiary evaluation directly, its underlying logic challenges the assumption that technological tools can operate within legal processes without full contestability. If the mere storage of data demands ex ante justification and strict supervision, then the transformation of that data into legally salient inferences by AI systems must meet an equal, if not higher, threshold of procedural scrutiny. Otherwise, defendants risk facing algorithmically produced evidence whose reasoning they cannot interrogate, a situation that the CJEU's proportionality doctrine implicitly warns against. In this sense, Digital Rights Ireland signals a broader constitutional anxiety: once digital infrastructures enter legal processes, they must be governed not only as technical tools but as potential reallocations of institutional power.

4.2 The European Convention on Human Rights (ECHR)

Article 6 ECHR guarantees the right to a fair trial, including equality of arms, adversarial proceedings, and the ability to scrutinize evidence. However, established jurisprudence suggests tension between these guarantees and the inherent opacity of AI systems.
In Hertel v Switzerland (ECtHR, 1998), the Court held that procedural fairness requires access to relevant materials that may affect the trial's outcome. If even scientific publications had to be disclosed to ensure fairness, how can AI-generated evidence often inaccessible, proprietary, and technically obscure be justified unless its internal processes are made available to the defense?

Even more critically, adversarial justice assumes that evidence can be challenged. But what happens when the "expert witness" is a commercial algorithm whose logic and trail of thought process cannot be fully explained? In my opinion, this is confronting us with a legal contradiction: if evidence cannot be contested, is it still legally admissible?

The ECtHR has repeatedly affirmed that defendants must be able to "understand the essence of the evidence used against them" (Dombo Beheer B.V. v. the Netherlands, 1993). Yet current AI tools do not satisfy this condition. This suggests that either AI evidence should currently be inadmissible, or (B) Article 6 needs to be reconceptualized altogether. Here exactly lies the doctrinal tension: European law promises contestability, but often AI makes contestability impossible.

4.3 The Charter of Fundamental Rights of the European Union

The Charter introduces a broader perspective: not just fairness (Articles 47-48), but also dignity (Article 1) and data protection (Article 8). Here lies an interesting paradox. Many argue that better datasets and audited algorithms will eventually "fix" the fairness problem. But human dignity, in my opinion, cannot be satisfied merely through statistical refinement. A risk score of 82% likelihood of reoffending may be statistically valid, but it still fails a deeper constitutional test: can a defendant respond meaningfully to a number they cannot interpret? This goes beyond fairness, it touches the very idea of legal personhood. This is reinforced indirectly by Digital Rights Ireland (CJEU, 2014), where the Court held that mass data processing, regardless of its efficiency, violates fundamental rights when it undermines autonomy, self-determination, and legal foreseeability. If large-scale predictive policing or probabilistic forensic inference is built on similar logic, then the Charter may already contain grounds to challenge AI-assisted evidence as unconstitutional in principle, not merely in application.

Thus, a contradiction emerges: The Charter protects dignity and autonomy, but algorithmic evidentiary tools may reduce defendants to statistical profiles. If so, does legality require explainability? Or does it require refusing certain forms of AI entirely?

4.4 The GDPR : A Tool for Criminal Procedure by Accident?

GDPR is often dismissed as irrelevant to criminal adjudication, but its architecture directly confronts algorithmic opacity. Article 22 GDPR prohibits decisions "based solely on automated processing" when they have legal effects. Some might argue this is irrelevant to criminal trials, since judges formally retain decision-making power. Yet that assumption is fragile. Consider the phenomenon of automation bias: if a judge receives a "neutral" AI-generated score, psychologically, rejecting it becomes harder. The decision looks objective, even when it isn't. Article 15 GDPR provides a right to "meaningful information about the logic involved." But no current forensic or risk assessment tool used in Europe meets that threshold. So, either GDPR is currently being ignored, or criminal proceedings rely on technical fiction.

This suggests a provocative alternative: Maybe AI evidence is only permissible if the defense receives full access to its model, parameters, and training data. Without that, GDPR arguably blocks its use. This would radically reshape how AI is deployed in trials and could slow down digitalization rather than accelerate it.

4.5 The EU AI Act: Innovation or Paradox?

The AI Act recognizes forensic, predictive, and risk-based criminal justice tools as "high-risk" systems requiring human oversight, documentation, and transparency. Yet there is a crucial gap: the Act regulates AI before it reaches the courtroom but says almost nothing about how such AI behaves as evidence within litigation.

So, while developers face obligations, judicial actors do not. In practice, this may create a legal vacuum: if vendors comply, courts may assume compliance equals admissibility. These risks reversing the burden of proof, placing defendants in the position of proving unreliability rather than requiring the state to establish reliability in the first place.

In my view, the strongest interpretation of Article 6 ECHR and Article 47 Charter would require pre-emptive admissibility standards, not merely impact assessments done by manufacturers. Until this exists, AI risks entering courts through administrative compliance rather than evidentiary scrutiny.

4.6 National Comparative Insights, Three Models of Judicial Response

Jurisdiction	Court Attitude	Risk	Opportunity
Germany	Highly cautious	Slow progress	Strong dignity-based safeguards
Netherlands	Experimental but reactive	Normalization of automation	Courts still willing to reject AI when needed
UK (pre-Brexit influence)	Mostly pragmatic	Overreliance on probabilistic tools	Case law shows potential for restraint

The integration of AI into evidentiary reasoning has not evolved uniformly across Europe. Instead, what emerges is a fragmented legal landscape that reflects deeper assumptions about expertise, technological progress, and the nature of judicial authority itself. In my view, these differences are not merely procedural; they reveal distinct understandings of what justice should tolerate when confronted with opaque computational systems.

Germany represents the most cautious position. Its constitutional jurisprudence places a strong emphasis on human dignity and informational self-determination, which appears to introduce a kind of conceptual resistance to opaque tools in criminal trials. Rather than treating AI as an inevitable upgrade, German legal reasoning often begins with the question of whether certain forms of automation are compatible with the very idea of adjudication. This approach has slowed down the integration of AI in criminal evidence, but perhaps deliberately so. One could argue that German courts are not resisting innovation but protecting the epistemic core of the rule of law, where decisions must remain attributable to accountable human agents. Yet there is a difficult question here: does excessive caution risk rendering justice systems technologically obsolete, or is it a principled refusal to sacrifice deliberation for speed?

The Netherlands offers a more experimental, almost pragmatic, alternative. Dutch institutions have shown willingness to deploy AI tools, but courts have demonstrated clear limits when constitutional values are threatened. The SyRIjudgment (District Court of The Hague, 2020) is particularly revealing. The algorithm was struck down not simply because of technical flaws but because its opacity eroded transparency, contestability, and human dignity. What matters here, in my point of view, is that the court did not accept efficiency as an excuse for opacity. This suggests a judicial philosophy that permits technological experimentation while reserving the right to withdraw it once democratic accountability is endangered. The paradox, however, is that this reactive approach allows potentially problematic systems to operate until they are contested. What if certain evidentiary tools influence trials before their flaws become visible?

The United Kingdom (pre-Brexit legal influence) has taken a path that could be described as cautiously pragmatic. AI-based forensic tools, especially probabilistic DNA methods, have been admitted in court, sometimes with judicial skepticism but rarely rejected outright. The decision in R v T (2010) indicated discomfort with statistical evidence that lacked a solid methodological foundation, yet it did not provide a clear standard for assessing such tools in the future. This, in my view, reflects the core dilemma of the UK approach: it neither embraces automation fully nor subjects it to systematic scrutiny. The risk is normalization through practice, where AI is gradually absorbed into evidentiary reasoning simply because it becomes common, not because it becomes explainable or contestable.

A Shared Problem: Evidence Without Understanding?

Across these jurisdictions, one recurring problem remains unresolved: none has yet produced a doctrinal test for determining when an AI system should be admissible as evidence. Courts continue to treat AI-based tools as if they were traditional forms of expert evidence, without acknowledging that their reasoning processes are often inaccessible even to their creators. This raises an uncomfortable possibility: evidentiary assessment in Europe may be moving toward formal compliance without epistemic transparency.

In my opinion, the deeper issue is not how to integrate AI into criminal trials, but whether current legal concepts, like expert testimony, evidentiary reliability, and human oversight, are still sufficient when the "expert" is a proprietary algorithm. If no one can understand the reasoning behind a classification or risk score, does its accuracy matter? Or does procedural fairness require transparency even when accuracy is high? This dilemma lies at the heart of Europe's legal struggle with AI: the law demands contestability, but AI often makes contestability technically impossible.

4.7 Soft Law and Ethical Guidelines, Helpful or Illusionary?

Alongside binding legal instruments, a dense body of soft-law initiatives has emerged to govern the development and use of artificial intelligence in criminal justice. Ethical guidelines, codes of conduct, and non-binding frameworks issued by international organizations and expert bodies consistently emphasize principles such as fairness, transparency, accountability, and human oversight. Prominent examples include the Ethics Guidelines for Trustworthy AI developed by the European Commission's High-Level Expert Group on Artificial Intelligence (2019), the OECD Principles on Artificial Intelligence (2019), and the Council of Europe's work under the CAHAI framework aimed at establishing common standards for AI governance (European Commission, 2019; OECD, 2019; Council of Europe, 2020).

These instruments play an important normative and coordinative role. As scholars of governance and regulation have noted, soft law often functions as a flexible regulatory layer capable of shaping institutional behavior in areas characterized by technological uncertainty and rapid innovation (Abbott & Snidal, 2000; Yeung, 2018). In the context of artificial intelligence, ethical guidelines contribute to the articulation of shared values, influence system design at early stages, and may serve as precursors to binding regulation. From this perspective, soft law can be understood as part of a broader process of digital constitutionalism, whereby normative principles are gradually translated into legal constraints (De Gregorio, 2022).

However, the reliance on soft-law instruments in criminal proceedings raises serious concerns. Unlike statutory law or binding procedural rules, ethical guidelines lack legal enforceability. Courts are not obliged to apply them, and defendants cannot rely on their breach as an autonomous ground for excluding evidence or establishing a violation of procedural fairness. As Abbott and Snidal observe, the strength of soft law lies in its flexibility, but this very flexibility also constitutes its principal weakness when legal accountability is at stake (Abbott & Snidal, 2000). In the criminal context, where the exercise of state power directly affects individual liberty, non-binding ethical commitments risk operating as symbolic assurances rather than as effective safeguards.

This risk is particularly acute in relation to evidentiary evaluation. The right to a fair trial under Article 6 ECHR presupposes legally enforceable guarantees of adversarial proceedings, equality of arms, and reasoned judicial decision-making. Ethical principles, however robust in abstraction, cannot ensure that defendants are granted meaningful access to the logic, assumptions, and limitations of AI systems whose outputs are introduced as evidence. Nor can they compel disclosure of proprietary models, training data, or error rates when such information is shielded by claims of trade secrecy or technical complexity (Pasquale, 2015; Hildebrandt, 2015; Veale & Borgesius, 2021).

4.8 Conclusion: The Core Contradiction

European legal frameworks often appear, at least when read in the abstract, to have all the ingredients needed to manage the arrival of AI in criminal trials. Yet once these frameworks are placed alongside the kinds of opaque computational systems now entering evidentiary practice, a more complicated picture emerges. Many of the guarantees that seem robust on paper,

transparency, adversarial challenge, meaningful human oversight, depend on assumptions about how evidence is created and explained that automated systems simply do not share.

The tension is not merely technical. It goes to the foundations of procedural justice: if courts cannot understand or interrogate the reasoning behind algorithmic outputs, the form of fairness may persist while its substance gradually erodes. The unresolved issue is not only how the law should respond to AI but whether certain technologies, by their very design, sit outside what legal proof can legitimately accommodate. Until this is confronted directly, European criminal procedure risks maintaining its doctrinal vocabulary while quietly absorbing tools whose epistemic logic it cannot fully supervise.

5. MAPPING AI TOOLS IN EVIDENTIARY EVALUATION

5.1 Introduction

The legal debate around AI in criminal trials often remains vague, focusing on theories of fairness, transparency, and procedural guarantees. Yet courts do not deal with abstractions. They confront technologies: specific tools, specific outputs, and specific cases. In my view, unless we identify and examine the actual systems already shaping evidentiary evaluation, our legal analysis risks becoming detached from the realities of judicial practice.

The purpose of this chapter is therefore not to just catalogue AI tools but to interrogate them. Each technology carries implicit assumptions about truth, probability, reliability, and human judgment. The law tends to treat them as "enhanced forensic instruments," but their logic often differs fundamentally from traditional forms of expert testimony. This creates an epistemic gap: courts continue applying old evidentiary standards to tools built on statistical inference and proprietary design, without fully grasping what that shift entails for due process. The core concern is simple: can these tools be used without reshaping the meaning of evidence itself? Or, put differently: does evidentiary evaluation risk becoming a process of human validation of machine outputs, rather than reasoned human judgment informed by evidence?

5.2 Typology of AI Applications in Criminal Evidence

The following categories provide a structural map, but each carries constitutional implications that require scrutiny, not passive acceptance. A. Forensic AI (Pattern Recognition & Data Matching). These systems include facial recognition, voice analysis, image pattern detection, and forensic data mining. They are often treated as objective tools that merely assist investigators. Yet they introduce several contradictions that courts have yet to address. The core risks that are hidden in this first category are that the statistical outputs are being presented as categorical truth ("match/no match") secondly that the unintentional outsourcing of interpretation: judges may not see probabilities, only just conclusions thirdly dangerous are also the hidden datasets: training data is rarely disclosed for scrutiny, and lastly the enormous risk of automation bias is mandatory to be taken into consideration : courts may defer to machine precision

In my point of view, the legal danger is subtle but profound: the more precise these tools appear, the less they are interrogated. Yet precision is not the same as evidentiary reliability and certainly not the same as legal admissibility. Courts often treat "confidence scores" as forensic certainty, which arguably clashes with the standard of proof beyond reasonable doubt, particularly when the threshold of doubt itself becomes dependent on proprietary statistical inference. B. Credibility Assessment Tools (AI-based Behavioral Analysis) These are perhaps the most controversial. They claim to detect deception, emotional states, or risk behavior through micro-expressions, voice stress patterns, or biometric data. Some law enforcement agencies already experiment with interviews and border control profiling. But epistemologically, they may represent a category of error. Can machine analysis of facial tension or pitch variation ever count as evidence? Or is this merely a digitalized version of pseudoscience dressed in algorithmic language? Legal tension: The law recognizes deception as intention: AI detects correlation; not intent Courts require interpretation of AI: generates classification. Expert testimony is subject to cross-examination: AI models often cannot be fully explained

In my opinion, these systems push the law toward a dangerous threshold: they infer internal mental states through external pattern recognition. That shifts evidence from the observable to the speculative. If adopted in court settings, they risk undermining the presumption of innocence by reversing the logic of inference: rather than proving guilt, they probabilistically predict it. C. Probabilistic Genotyping (DNA Evidence via AI) This is where AI arguably performs the strongest but also where the risk of judicial over-reliance is greatest. Probabilistic DNA tools, such as STRmix or TrueAllele, assign statistical likelihoods that a genetic sample matches a suspect. Courts in the UK, US, and increasingly Europe have admitted such tools into evidence. I would argue this technology exposes the great epistemic dilemma of AI in trials: does probability qualify as evidence or only as investigative guidance?

Problems arise when probabilistic figures are treated as decisive rather than indicative. Judges may see "96% match" and assume near certainty, yet that figure may be statistically valid while legally insufficient. The standard of proof in criminal

law (“beyond reasonable doubt”) does not translate smoothly into statistical terms, courts must interpret probability, not merely receive it. Yet most judges are not trained statisticians. This creates an epistemic vulnerability rarely acknowledged in admissibility hearings. If the interpretation of evidence depends more on an algorithm than on legal reasoning, has judicial authority subtly shifted from the courtroom to the data pipeline? D. Digital Triage & Prioritization Systems These tools operate in policing and early-stage investigation. They rank suspects, select digital evidence for analysis, or predict investigative leads. At first glance, they appear merely operational, not evidential. But once they determine what counts as relevant evidence, they indirectly frame the evidentiary landscape itself. The danger is not that they produce evidence, but that they determine what is seen as evidence at all.
This could quietly reshape legal objectivity: if data-driven relevance replaces investigative discretion, do we risk codifying bias into case structure before a trial even begins? In my view, digital triage may be the most underestimated form of evidentiary automation, precisely because it influences procedural fairness before courts even recognize its presence.

E. Predictive & Risk Assessment Systems. In pre-trial decisions, especially remand and bail, AI risk tools are increasingly used to assist judicial decisions. They claim to improve consistency, but consistency can also conceal bias. The COMPAS controversy in the United States, and critical debate over its statistical bias, reveal the potential consequences of this approach. But the hardest question remains: If prediction shapes judicial decision-making, is judgment still a forward-looking human act, or a validation of an algorithmic forecast? The answer will determine not only the legitimacy of AI in criminal trials, but the future nature of judicial reasoning itself.

5.3 Case-Based Survey of Deployment in Europe, Uses documented instances, pilot programs, or controversies

If AI tools remained on theoretical prototypes, the debate on evidentiary integrity could stay comfortably abstract. But this is no longer the case: real systems already shape investigations, structure evidence, and even inform judicial decision-making across European jurisdictions. What emerges is not a uniform approach, but a fragmented and reactive legal landscape, one where practice often advances faster than doctrine.

One of the most revealing examples is the Netherlands’ SyRI welfare-fraud algorithm, whose downfall before the District Court of The Hague (2020) marked a turning point. The Court did not merely critique technical inaccuracy; it declared the system incompatible with fundamental rights due to its opacity, disproportionate scope, and absence of contestability. In my view, this case is pivotal because it reveals something deeper: AI was not rejected because it failed technologically, but because it proved conceptually incompatible with democratic accountability. The court implicitly recognized that opacity is not just a technical flaw, it is a constitutional threat.

In the United Kingdom, the use of probabilistic DNA tools: such as TrueAllele and STRmix,has been even less restrained. Cases like R v T (2010) revealed judicial discomfort but failed to establish clear admissibility thresholds. The judiciary acknowledged its unease with purely statistical inference, but ultimately allowed it. Here we see a different form of risk: not rejection, but normalization through ambiguity. When evidentiary tools are admitted without doctrinal guidance, their integration feels less like reform and more like habituation. In practice, algorithms begin to shape legal reasoning before the law decides how to understand them.

Germany presents a striking contrast. Courts remain cautious, invoking dignity-based reasoning when assessing forensic automation. Yet paradoxically, this hesitation has not led to clear doctrinal guidance. If anything, it has generated legal uncertainty: judges fear using AI tools without regulation, but legislators have not provided an admissibility standard. This results in a conceptual vacuum where risk is simply deferred to the future. One could argue that Germany’s caution protects due process, but it also avoids the harder question: Does criminal adjudication have epistemological limits beyond which automation is fundamentally incompatible?

Across these examples, a pattern emerges: AI enters criminal procedure not through legal design, but through institutional necessity, only to be challenged when its consequences become visible. This reactive approach reveals a deeper concern: evidentiary transformation might be occurring precisely in those states where legal doctrine has not yet developed the language to recognize it. The law sees outcomes but not processes. And when process becomes invisible, contestability becomes performative rather than substantive.

5.4 Institutional Motives of Adoption

AI enters criminal justice not because courts have endorsed its epistemic legitimacy, but because institutions face escalating operational burdens. Prosecutors handle vast volumes of digital evidence; law enforcement agencies face pressure to predict, not merely investigate; governments must demonstrate technological “modernization” to preserve political credibility. Under these conditions, AI does not appear as a choice, it appears as administration’s survival mechanism.

The first motive is efficiency, often framed as a neutral objective. But efficiency subtly displaces the central value of criminal justice, deliberation. When caseload pressure becomes extreme, deliberation begins to look like delay, and AI becomes not a tool but a justification. In my view, this shift is dangerous precisely because it is silent: the more AI is rationalized as operational necessity, the less its epistemological consequences are debated.

The second motive is political narrative. European states increasingly promote digital transformation as a marker of state competence. AI tools become a visible emblem of "progress," which shapes judicial practice indirectly: courts may feel pressure to keep pace with policing technologies to appear relevant and up to date. Yet this introduces a paradox. If modernization becomes symbolic, questioning AI may be seen as backward or even professionally unreasonable. That creates a cultural chill around critique, even when critique is legally required.

The third and perhaps most subtle motive is institutional liability. When AI provides outputs, risk scores, rankings, probability assessments, it diffuses responsibility. Decision-making becomes distributed across systems, developers, and officials. Accountability becomes harder to trace. Courts may claim deference to expert systems; vendors may invoke trade secrecy; investigators may rely on "best available tools" as justification. What worries me is not that responsibility is lost, but that it begins to disappear precisely at the moment when it is most needed.

Thus, AI enters trials not as a guest but as a necessity. Yet necessity is not justification. If AI is adopted because of institutional strain rather than legal reasoning, we risk creating a justice system where efficiency drives change while legitimacy tries to catch up, often too late. The danger is that contestability becomes something we promise but no longer structurally enable. And when contestability becomes symbolic, due process becomes ceremonial.

5.5 Preliminary Risks Identified

Across the various tools discussed, certain risks appear repeatedly, even if they manifest differently depending on the specific technology. The first and most visible is opacity. Machine-learning systems often rely on proprietary datasets or internal architectures that cannot be disclosed, which makes it difficult for courts, and almost impossible for defendants, to test the foundations of an algorithmic conclusion.

A second concern is the tendency for over-reliance. When a tool presents a match score or a neat probability, its apparent precision can overshadow the fact that the number rests on assumptions that may be invisible to the courtroom. Judges do not intentionally defer to algorithms, but in practice the authority of a quantified output can be hard to resist.

Finally, there is the issue of reliability in the broader epistemic sense. AI systems produce patterns and probabilities, not concrete facts about past events. If courts begin treating these patterns as direct evidence rather than interpretive aids, the process of criminal adjudication may shift from reconstructing specific acts to validating statistical inferences. That shift, even if subtle, would alter the nature of legal proof itself.

6. EVIDENTIARY EPISTEMOLOGY AND AI'S CHALLENGE

6.1 Introduction: The Epistemic Fault Line When AI Enters Legal Reasoning

Up to now, I have examined AI primarily as a procedural and regulatory challenge. Yet as I progressed with the research, it became increasingly clear to me that the deepest problem is neither regulatory nor technical, but epistemic. Criminal adjudication is fundamentally an exercise in knowledge production. Courts reconstruct past events through evidence that can be understood, contested, and ultimately justified through human reasoning. The standard of proof, the right to challenge evidence, and the requirement for reasoned judgments all presuppose a shared epistemic world where facts can be explained in language accessible to the people involved (Ashworth & Redmayne, 2021).

AI-based evaluative tools disrupt this world. They introduce forms of inference that do not operate through explicit reasoning but through statistical optimization. Complex machine-learning models generate outputs that even their designers may struggle to explain (Burrell, 2016). This creates what I see as an epistemic rupture between the traditional logic of the courtroom and the computational logic embedded in AI systems.

This tension is not merely a matter of complexity. Traditional expert evidence, however technical, remains, in principle, intelligible. Experts can describe their methods, their assumptions, and their margin of error. Courts can ask questions and assess credibility. AI tools, especially deep-learning models, often cannot provide such explanations (Doshi-Velez & Kim, 2017). Their inferential processes are the product of thousands of internal parameters and correlations that resist articulation.

This is where the professor's feedback changed how I thought about the issue: I initially framed AI as simply "opaque," but opacity alone does not make a technique legally incompatible. DNA analysis was opaque to judges at first; so were early

fingerprint comparisons. What matters is whether the reasoning behind the evidence can be made intelligible and contestable within the adversarial process.

AI challenges this because many models produce correlations without causal grounding (Pearl, 2018). In scientific terms, correlations can be useful. But in criminal adjudication, where responsibility and intent matter, correlation without explanation is insufficient. If an algorithm claims that a suspect "matches" a pattern, but cannot explain why, then the court receives information without understanding. That, in my view, threatens the epistemic foundations of a fair trial. This chapter therefore asks a simple but difficult question: can legal proof, rooted in justification, explanation, and contestability, coexist with AI systems rooted in prediction, correlation, and opacity?

To address this, I examine (1) the tension between probabilistic inference and the criminal standard of proof, (2) the ambiguous status of algorithmic outputs as "evidence," (3) the risk that machine authority subtly displaces human deliberation, and (4) the need for an epistemic model of contestability that aligns AI with the logic of the trial.

6.2 Standards of Proof vs Probabilistic Reasoning

The criminal standard of proof, "beyond reasonable doubt", is sometimes described informally as if it were a percentage threshold, but doctrinally it reflects a moral judgment rather than a numerical one. It asks whether the state is justified, in a normative sense, in imposing criminal liability. AI systems, by contrast, tend to express their conclusions through probabilistic outputs, and this tension creates practical difficulties that are not easily resolved.

A probabilistic DNA tool may report a match likelihood of 94 percent, yet that figure does not map neatly onto the moral or normative criteria embedded in criminal adjudication. Judges, moreover, often struggle with how to interpret such numbers, especially when they are presented as the product of a sophisticated computational method. There is a risk, well documented in psychological literature, that numerical precision carries more persuasive force than it deserves. When a court relies heavily on such a figure, the presumption of innocence may remain intact in theory while becoming more fragile in practice. Ultimately, the law must translate probabilistic information into binary outcomes, and that translation is a normative act that machine outputs cannot perform on their own.

6.3 The Status of Algorithmic Inference: Can Correlation Count as Evidence?

Algorithmic outputs occupy an ambiguous epistemic category. They are not observations, not testimony, not expert opinions in the traditional sense. They are predictions derived from patterns that algorithms find in large datasets. But these predictions often lack causal explanation, which poses a problem for evidentiary reasoning.

Courts often treat algorithmic classifications as if they were simply modernized expert assessments. Yet this analogy breaks down once we examine how AI systems actually operate. Experts reason they interpret data, articulate methods, justify choices, and describe uncertainty. Algorithms compute they detect statistical associations and transform them into outputs. As Pearl (2018) argues, machine-learning models excel at correlation but struggle with causation.

This distinction is crucial in criminal trials. Legal responsibility is inherently normative; it depends not only on what happened but on why it happened. A correlation cannot establish intent, motive, or culpability. But if courts mistake statistical association for evidentiary relevance, they risk replacing causal reasoning with pattern recognition.

Moreover, contestability, one of the core requirements under Article 6 ECHR, depends on the ability to interrogate evidence. Yet many algorithmic systems are shielded by trade secrecy or rely on training data that cannot be disclosed (Larsson & Heintz, 2020). In such cases, defendants face the impossible task of challenging evidence without access to its foundation. The ECtHR has repeatedly held that the defense must have the ability to examine and contest the evidence against them (Dombo Beheer B.V. v Netherlands, 1993). That principle is incompatible with evidence whose reasoning process cannot be revealed.

In my view, algorithmic outputs should therefore be treated as investigatory leads, not as evidence capable of meeting the criminal standard. Without causal explanations, interpretability, and meaningful contestability, algorithmic inference cannot bear the weight of criminal conviction.

6.4 Machine Authority vs Human Deliberation

A recurring concern in the literature is that algorithmic systems may subtly reshape human judgment, not by replacing it outright, but by becoming epistemic anchors that influence how human decision-makers interpret evidence (Citron, 2019). When an algorithm labels a defendant as high-risk or assigns a strong match of probability, that output can become a focal point of judicial reasoning, even when judges genuinely intend to exercise independent judgment.

The problem is partly psychological. Automation bias inclines humans to trust systems that appear objective, especially under time pressure (Skitka et al., 1999). But the problem is also institutional. Courts increasingly operate under constraints, limited resources, growing caseloads, and political pressure to modernize. AI promises efficiency and consistency, and these promises shape judicial culture even before judges consciously engage with specific tools.

We ought to be warned against portraying judges as passive. Courts can be skeptical. For example, R v T (2010) shows judicial hesitation toward probabilistic fingerprint analysis. But skepticism requires tools. When AI systems are proprietary, when their inferential logic is inaccessible, and when the underlying data cannot be disclosed, judicial skepticism becomes toothless. A judge may wish to challenge the algorithm but lack the epistemic leverage to do so.

This raises a more troubling possibility: human deliberation may remain formally intact but become substantively influenced by algorithmic reasoning. Judges still issue judgments; lawyers still argue; defendants still appeal. Yet the space of what counts as reasonable or plausible becomes silently structured by machine outputs. In my opinion, these risks transforming adjudication into a post-hoc justification exercise where the human decision-maker explains rather than determines the outcome.

Administrability and Epistemic Responsibility.

Criminal procedure presupposes the existence of identifiable epistemic agents who can explain, justify, and be held responsible for the evidentiary inferences presented at trial. AI-mediated evidentiary systems disrupt this assumption by distributing epistemic authority across developers, data providers, institutional users, and technical infrastructures. When no single actor can fully account for how an evidentiary output was produced, an administrability gap emerges: responsibility for explanation, error, and justification becomes structurally diffuse. This diffusion poses

6.5 Toward an Epistemic Model of Contestability

If AI is to play any legitimate role in evidentiary evaluation, courts must adopt a new epistemic model of contestability, one that goes beyond mere procedural formality and ensures that defendants can genuinely challenge algorithmic reasoning.

Contestability requires access. Defendants must be able to examine the basis of an algorithmic output. GDPR's promise of "meaningful information about the logic involved" (Art. 15) offers a starting point, but interpretations have been narrowing (Wachter et al., 2017). Without access to training data, model architecture, and error rates, contestation is symbolic. Contestability requires interpretability. Explanations must be intelligible to lawyers, judges, and defendants. Technical documentation or raw code does not constitute an explanation in the legal sense. Interpretability must be measured by epistemic accessibility, not technical completeness. Contestability also requires epistemic humility. Courts must recognize that AI systems embed assumptions, biases, and limitations. They must treat algorithmic outputs as claims requiring justification, not as quasi-scientific truths. Ultimately, an epistemic model of contestability would require courts to ask a deeper question: is the logic of this AI system compatible with the normative logic of criminal proof? If the answer is no, then regulatory safeguards cannot fix the problem; certain forms of AI will simply be incompatible with criminal adjudication.

This is where my position diverges from approaches that emphasize better regulation or improved explainability. I do not believe all AI tools can be rehabilitated for evidentiary use. Some may operate on epistemic foundations fundamentally misaligned with the requirements of justification, explanation, and human accountability. Recognizing these limits is, in my view, the only way to protect the integrity of due process in an era of accelerating automation.

7. CONCLUSION

7.1 From Technical Question to Constitutional Problem

This dissertation set out to address a question that initially appeared technical but ultimately revealed itself to be deeply constitutional: whether criminal trials can remain fair when the evaluation of evidence is mediated by artificial intelligence. What began as an inquiry into technological innovation evolved into a broader examination of the epistemic foundations of criminal proof, the structure of procedural guarantees, and the role of human judgment in adjudication. The analysis has shown that the challenge posed by AI cannot be reduced to questions of efficiency, accuracy, or innovation, but instead strikes at the normative conditions under which criminal responsibility is attributed.

7.2 Algorithmic Evidence and the Structural Tension in Criminal Procedure

The dissertation demonstrated that artificial intelligence no longer operates at the margins of criminal justice. From probabilistic DNA analysis and forensic pattern recognition to digital triage and automated risk assessment, algorithmic systems are already embedded across multiple stages of the evidentiary process. Their integration, however, has largely occurred through institutional necessity and technological availability rather than through deliberate doctrinal design. This

discrepancy between legal theory and institutional practice emerged as a central structural tension shaping contemporary criminal procedure.

7.3 Fair Trial Guarantees and Epistemic Asymmetry

At the doctrinal level, the analysis showed that European fair-trial law, anchored in Article 6 ECHR and reinforced by Articles 47 and 48 of the EU Charter, is premised on contestability, transparency, equality of arms, and reasoned judicial deliberation (Ashworth & Redmayne, 2021; ECtHR, Dombo Beheer v. the Netherlands, 1993). These guarantees presuppose that evidence can be meaningfully examined and challenged by the accused. Many AI systems currently deployed in criminal contexts, however, operate through opaque computational architectures that resist full explanation even by their designers (Burrell, 2016; Pasquale, 2015). The resulting epistemic asymmetry places defendants in a structurally disadvantaged position, not because the law explicitly weakens their rights, but because the technological conditions of proof undermine their effective exercise.

7.4 Probability, Automation Bias, and the Limits of Justification

The dissertation argued that opacity becomes normatively decisive when it intersects with evidentiary authority. Once algorithmic outputs function not merely as investigative leads but as decisive contributors to legal proof, tensions with due process intensify. Courts are increasingly confronted with probabilistic inferences, likelihood ratios, confidence scores, and risk metrics, that carry an aura of scientific objectivity while remaining detached from causal explanation (Pearl, 2018; Kaye, 2019). The translation of such probabilistic outputs into binary legal outcomes is not a scientific operation but a normative judgment. This problem is further exacerbated by automation bias, namely the human tendency to over-trust automated systems under institutional pressure for efficiency and consistency (Skitka et al., 1999; O'Neil, 2016). In such settings, human deliberation risks becoming secondary to machine-generated conclusions rather than the genuine source of legal determination (Citron, 2019).

7.5 Bias, Power, and the Procedural Environment

The analysis further demonstrated that algorithmic bias is not an incidental malfunction but a structural feature of data-driven systems trained on historically saturated datasets (Barocas & Selbst, 2016; Crawford, 2021). Within criminal justice, this bias assumes particular gravity because it directly shapes the distribution of coercive state power. When biased outputs enter evidentiary reasoning, they do not merely distort outcomes; they reshape the procedural environment itself, influencing which facts appear relevant, which suspects appear risky, and which narratives appear credible. This procedural dimension of bias emerged as one of the most constitutionally significant risks identified in the dissertation.

7.6 Regulatory Limits and Comparative Insights

From a regulatory perspective, the examination of the EU Artificial Intelligence Act revealed both progress and limitation. While the AI Act acknowledges the high-risk nature of AI systems used in criminal justice and imposes strict ex ante safeguards, it remains structurally external to criminal procedure. It governs system design and deployment, but it does not resolve the epistemic and procedural questions that arise once algorithmic outputs are introduced as evidence in court. Compliance with regulatory standards does not, in itself, ensure compatibility with Article 6 ECHR. This regulatory–procedural gap was further illustrated through the comparative analysis of Germany, the Netherlands, and the United Kingdom, where judicial responses remain fragmented, reactive, and conceptually strained.

7.7 Toward Epistemic Contestability as a Constitutional Requirement

At its deepest level, the challenge posed by AI is epistemological. Criminal trials are institutions tasked with producing justified truths under conditions of uncertainty, moral responsibility, and human fallibility. Their legitimacy depends on the ability of those affected to understand, contest, and challenge the reasons underlying judicial decisions. This dissertation has argued that not all forms of AI can be reconciled with these demands. While certain systems may function as legitimate investigatory aids, their elevation to evidentiary authorities introduces risks that cannot always be neutralized through transparency, oversight, or technical refinement alone. In such cases, exclusion from evidentiary use may be constitutionally required

At the same time, a purely prohibitionist stance would be neither realistic nor normatively coherent. The task is not to resist technological transformation as such, but to discipline it through epistemic contestability. For AI-mediated evidence to be compatible with the right to a fair trial, defendants must have genuine access to the foundations of algorithmic outputs; judges must retain substantive interpretive authority; and probabilistic inference must never be allowed to substitute for normative proof.

7.8 Final Reflection: Digital Justice and Human Judgment

What ultimately emerges is a demanding conception of digital justice. Digital justice cannot be reduced to faster courts or data-driven decision-making. It requires that technological infrastructures conform to the moral and procedural architecture of the criminal trial, rather than the trial being silently reshaped around the logic of machines. European criminal procedure stands at a constitutional crossroads. If algorithmic systems continue to enter evidentiary reasoning without a corresponding epistemic recalibration of due process, the risk is not overt injustice but a gradual hollowing-out of the right to a fair trial from within.

Bibliography

1. *Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias: There's software used across the country to predict future criminals. And it's biased against Blacks. ProPublica.* https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?utm_source

1. *Ashworth, A., & Redmayne, M. (2021). The criminal process (5th ed.). Oxford University Press.* https://search.worldcat.org/title/1025224864

1. *Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732.* https://doi.org/10.15779/Z38BG31

1. *Binns, R. (2018). Algorithmic accountability and public reason. Philosophy & Technology, 31(4), 543–556.* https://link.springer.com/article/10.1007/s13347-017-0263-5

1. *Abbott, K. W., & Snidal, D. (2000). Hard and Soft Law in International Governance. International Organization, 54(3), 421–456.*

1. *Council of Europe. (2020). Towards Regulation of Artificial Intelligence Systems (CAHAI).*

1. *De Gregorio, G. (2022). Digital Constitutionalism in Europe. Cambridge University Press.*

1. *European Commission High-Level Expert Group on AI. (2019). Ethics Guidelines for Trustworthy AI.*

1. *OECD. (2019). OECD Principles on Artificial Intelligence.*

1. *Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the EU Artificial Intelligence Act. Computer Law Review International.*

1. *Yeung, K. (2018). Algorithmic Regulation: A Critical Interrogation. Regulation & Governance.*

1. *Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 1–12.* https://journals.sagepub.com/doi/10.1177/2053951715622512

1. *Casey, E. (2011). Digital evidence and computer crime: Forensic science, computers and the internet (3rd ed.). Academic Press.* https://www.elsevier.com/books/digital-evidence-and-computer-crime/casey/9780123742681

1. *Citron, D. K. (2019). Sexual privacy. Yale Law Journal, 128(6), 1870–1960.* https://www.yalelawjournal.org/article/sexual-privacy

1. *Craig, P. (2018). Administrative law (8th ed.). Sweet & Maxwell.* https://www.wildy.com/isbn/9780414065865/administrative-law-8th-edition-hardback-sweet-and-maxwell

1. *Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.*

   https://www.researchgate.net/publication/363177193_Atlas_of_AI_Power_Politics_and_the_Planetary_Costs_of_Artificial_Intelligence

1. *Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv.* https://arxiv.org/abs/1702.08608

1. *Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1), eaao5580.*
   https://www.science.org/doi/10.1126/sciadv.aao5580

1. *Ferguson, A. G. (2017). The rise of big data policing: Surveillance, race, and the future of law enforcement. NYU Press.*
   https://nyupress.org/9781479892903/the-rise-of-big-data-policing/

1. *Garvie, C., Bedoya, A., & Frankle, J. (2016). Perpetual line-up: Unregulated police face recognition in America. Georgetown Law Center on Privacy & Technology.* https://www.perpetuallineup.org

1. *Hildebrandt, M. (2015). Smart technologies and the end(s) of law. Edward Elgar.* https://www.e-elgar.com/shop/gbp/smart-technologies-and-the-end-s-of-law-9781782540764.html

1. *Hildebrandt, M. (2020). Law for computer scientists and other folk. Oxford University Press.* https://global.oup.com/academic/product/law-for-computer-scientists-and-other-folk-9780198833718

1. *Kaye, D. H. (2019). Forensic science evidence: Science and the law. West Academic.* https://www.westacademic.com/Kaye-Forensic-Science-Evidence-Science-and-the-Law-9781683284437

1. *Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. University of Pennsylvania Law Review, 165(3), 633–705.* https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3

1. *Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. Internet Policy Review, 9(2), 1–19. https://policyreview.info/articles/analysis/transparency-artificial-intelligence*

1. *Latour, B. (2010). Technology is society made durable. In J. Law (Ed.), A sociology of monsters: Essays on power, technology and domination (pp. 103–131). Routledge. https://routledgehandbooks.com/doi/10.4324/9780203993810*

1. *Leib, E. J. (2019). Human judges in an age of artificial intelligence. Journal of Law and Courts, 7(2), 367–393. https://www.journals.uchicago.edu/doi/10.1086/702806*

1. *Mason, S., & Seng, D. (Eds.). (2017). Electronic evidence (4th ed.). Institute of Advanced Legal Studies. https://ials.sas.ac.uk/library/collections/electronic-evidence*

1. *O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown. https://weaponsofmathdestructionbook.com*

1. *Packer, H. L. (1968). The limits of the criminal sanction. Stanford University Press. https://www.sup.org/books/title/?id=2494*

1. *Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674970847*

1. *Pearl, J., & Mackenzie, D. (2018). The book of why: The new science of cause and effect. Basic Books. https://www.basicbooks.com/titles/judea-pearl/the-book-of-why/9780465097609*

1. *Skitka, L. J., Mosier, K. L., & Burdick, M. D. (1999). Does automation bias decision-making? International Journal of Human–Computer Studies, 51(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252*

1. *Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the 2019 ACM Conference on Fairness, Accountability, and Transparency (pp. 59–68). ACM. https://dl.acm.org/doi/10.1145/3287560.3287598*

1. *United Nations Office on Drugs and Crime. (2020). Education for Justice (E4J): University module series on cybercrime. https://www.unodc.org/e4j/en/teaching-hub.html*

1. *Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the EU Artificial Intelligence Act. Computer Law Review International, 22(4), 97–112. https://www.eur.nl/sites/corporate/files/2021-06/demystifying_the_eu_ai_act.pdf*

1. *Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation does not exist in the GDPR. International Data Privacy Law, 7(2), 76–99. https://academic.oup.com/idpl/article/7/2/76/3860948*

1. *Yeung, K. (2018). Algorithmic regulation: A critical interrogation. Regulation & Governance, 12(4), 505–523. https://onlinelibrary.wiley.com/doi/10.1111/rego.12158*

Legal Instruments and Case Law (APA-Formatted)

1. *Charter of Fundamental Rights of the European Union, 2012 O.J. C 326/391. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT*

1. *Council of Europe. (1950). European Convention on Human Rights. https://www.echr.coe.int/documents/d/echr/convention_ENG*

1. *Court of Justice of the European Union. (2014). Digital Rights Ireland Ltd v. Minister for Communications (Joined Cases C-293/12 & C-594/12). https://curia.europa.eu/juris/liste.jsf?num=C-293/12*

1. *Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation). https://eur-lex.europa.eu/eli/reg/2016/679/oj*

1. *Regulation (EU) 2024/1689 of the European Parliament and of the Council (Artificial Intelligence Act). https://eur-lex.europa.eu/eli/reg/2024/1689/oj*

1. *Dombo Beheer B.V. v. The Netherlands, App. No. 14448/88 (ECtHR, 1993). https://hudoc.echr.coe.int/eng?i=001-57868*

1. *Edwards and Lewis v. United Kingdom [GC], Apps. 39647/98 & 40461/98 (ECtHR, 2004). https://hudoc.echr.coe.int/eng?i=001-67052*

1. *Hertel v. Switzerland, App. No. 25181/94 (ECtHR, 1998). https://hudoc.echr.coe.int/eng?i=001-58235*

1. *Mirilashvili v. Russia, App. No. 6293/04 (ECtHR, 2008). https://hudoc.echr.coe.int/eng?i=001-90971*

1. *Schatschaschwili v. Germany [GC], App. No. 9154/10 (ECtHR, 2015). https://hudoc.echr.coe.int/eng?i=001-159566*

1. *District Court of The Hague. (2020). NJCM c.s. v. The Netherlands (SyRI), Case No. C/09/550982 / HA ZA 18-388. https://uitspraken.rechtspraak.nl/#!/details?id=ECLI:NL:RBDHA:2020:1878*

1. *R v T [2010] EWCA Crim 2439.* https://www.bailii.org/ew/cases/EWCA/Crim/2010/2439.html

1. *State v. Loomis, 881 N.W.2d 749 (Wis. 2016).* https://law.justia.com/cases/wisconsin/supreme-court/2016/2013ap1572-cr.html

1. *Roberts, P., & Zuckerman, A. (2010). Criminal Evidence.*

1. *Jackson, J., & Summers, S. (2012). The Internationalisation of Criminal Evidence.*

Disclaimer

A limited use of artificial intelligence tools was made in the course of this research, primarily for the identification of relevant case studies and case law, as well as for obtaining feedback on the overall structure of the argument. All substantive analysis, interpretation, and conclusions remain the sole work and responsibility of the author