ABA®

AI IN THE CRIMINAL COURTS | PROMISE AND PERIL OF AI IN POLICING | AI TRANSPARENCY IN FEDERAL SENTENCING

VOL 41, NO 01
SPRING 2026

CJ

A PUBLICATION OF THE AMERICAN BAR ASSOCIATION
CRIMINAL JUSTICE SECTION

AI Across the Criminal Justice System

AI in the Criminal Courts: Balancing Innovation and Justice

BY JUDGE BRIAN MACKENZIE (RET.) AND DAVID WALLACE

Artificial intelligence (AI) is not a distant technological concept emerging at the edges of the justice system. It is already an integrated part of that system, embedded in the daily work of policing, prosecution, defense, judging, and supervision. Prosecutors rely on predictive analytics to summarize large amounts of digital evidence; defense attorneys confront AI-filtered or AI-enhanced files they or their clients had no hand in generating, but that directly impact how their client is sentenced. Judges are already facing AI-related evidentiary questions, and both judges and probation staff increasingly encounter algorithmic risk scores attached to bail decisions or sentencing recommendations. AI systems are not theoretical. They shape real outcomes in real courtrooms today.

The Machine in the Courtroom

This rapid incorporation presents real challenges for the criminal justice system. It must manage the tension between (1) embracing technological progress that could improve efficiency and consistency and (2) safeguarding the constitutional guarantees that anchor fairness, accountability, and transparency. AI offers tremendous promise: the ability to synthesize data quickly, identify patterns humans might miss, and deliver consistency where human decision-making is susceptible to error or bias. Yet the same technology that promises efficiency also threatens to import bias, obscure the basis for decisions, expand surveillance, and encourage courts and counsel to defer too readily to algorithmic conclusions.

Across jurisdictions and disciplines, courts face a fundamental question: not

JUDGE BRIAN MACKENZIE (RET.) is an award-winning judicial educator who retired after nearly 27 years on the bench. Following his judicial career, he was one of the founders of the Justice Speakers Institute, where he currently serves as co-president and chief financial officer. **DAVID WALLACE** is a nationally recognized traffic safety expert and justice reform leader who also co-founded the Justice Speakers Institute, where he serves as co-president and presides over JSI's use of innovative technologies to support justice system improvement. He currently serves as the Chief Assistant Prosecutor in Huron County, Michigan.

Published in Criminal Justice, Volume 41, Number 1, Spring 2026.

phonlamaiphoto/Adobe Stock

whether AI should be used—that question has already been answered by practice—but how to use it responsibly, transparently, and in accordance with the rule of law. This article is based on the Justice Speakers Institute's (JSI) six-part series discussing *AI and the Courts* and illustrates the scope of this challenge. *See generally Artificial Intelligence and the Courts*, Just. Speakers Inst. (2025) (six-part series), https://justicespeakersinstitute.com/ai-and-the-courts-balancing-justice/. This article now examines the question of using AI from the perspective of criminal justice professionals, judges, prosecutors, and defense attorneys, through the lens of evidentiary admissibility, supervision practices, risk-assessment design, and ethical obligations. It proposes a practical, legally grounded path forward that embraces innovation while staying firmly aligned with constitutional value.

The Promise: Data-Driven Justice

AI's rapid growth in the criminal legal system reflects a core truth: When properly designed, validated, and governed, AI can enhance accuracy, reduce error, and help direct scarce resources toward the most critical tasks. Criminal cases increasingly involve mountains of digital evidence, video, audio, electronic messages, location data, sensor data, social-media activity, and more. The human ability to process these materials is limited; AI offers tools that can help ensure cases are resolved more efficiently and fairly.

Modern criminal investigations are becoming steeped in digital evidence. Technology-assisted review (TAR), a machine-learning process that trains a model to classify documents as relevant or not based on attorney-reviewed examples, has long been proven in civil e-discovery to exceed human reviewers in speed and accuracy. Its implications in the criminal context may be even more consequential. TAR tools can process millions of files, identify potentially relevant material, detect duplications, map communication patterns, and flag inconsistencies across datasets. When these systems are validated and their methods disclosed, they can accelerate the identification of exculpatory evidence, strengthen compliance with the requirements of *Brady v. Maryland*, 373 U.S. 83 (1963), and reduce the risk of wrongful arrest or charge.

AI-enhanced video analysis offers similar benefits. Algorithms can detect objects, identify individuals, or track movements across hours of footage. In a world where many criminal cases contain some form of video, the capacity to process footage quickly and accurately is invaluable. AI does not replace human judgment, but it can assist officers, analysts, and attorneys in identifying material evidence sooner and with fewer errors.

Pretrial release decisions can be inconsistent and influenced by subjective factors. AI-supported risk-assessment instruments can improve the process by providing judges with standardized information that may reduce unnecessary incarceration. Tools that evaluate objective factors from a defendant's criminal history have been shown to decrease reliance on pretrial detention, reduce failures to appear, and lower the risk of new criminal activity, including violent crime. Many of these instruments generate indicators related to court appearance, public safety, and the likelihood of violence, all intended to inform judicial discretion rather than replace it. In jurisdictions like Kentucky and in selected pilot efforts in Arizona, carefully monitored implementations of pretrial risk-assessment tools have been associated with lower jail populations and no clear increase in crime. In Kentucky, crime committed by individuals on pretrial release declined by 15 percent, even as releases increased by 70 percent. *See* Int'l Ass'n of Chiefs of Police, *Reducing Crime and Unnecessary Detention: Kentucky's Success with the New Public Safety Assessment—Court Tool* (2014), https://tinyurl.com/yave3up8.

In complex criminal cases, prosecutors and public defenders alike face discovery burdens that are vastly larger than even a decade ago. AI-supported platforms can flag delays, track disclosure obligations, and remind attorneys of Rule 16 and *Brady* responsibilities. *See* Fed. R. Crim. P. 16. These tools can assist defenders in identifying patterns, such as police officers repeatedly involved in reliability issues, and help ensure that exculpatory evidence is not overlooked. AI does not eliminate the duty to disclose or review evidence, but it can make compliance more systematic and less prone to human failures.

AI's deepest promise may lie in probation and parole. As described in the Justice Speakers Institute's sixth article, the National Institute of Justice (NIJ) is funding transformative work in this area. *See Part Six: Judging the Machine—Lessons, Guardrails, and the Path Forward*, Just. Speakers Inst. (2025), https://tinyurl.com/2vc5khbp. The Integrated Dynamic Risk Assessment for Community Supervision (IDRACS) project analyzes behavioral patterns, triggers, treatment engagement, and environmental risk in real time.

Similarly, new approaches integrate smartphone and wearable device data to detect crises early and guide officer interventions. These systems are not about harsher monitoring; they are designed to provide real-time, individualized support that can help prevent violations before they occur. Used responsibly, AI can shift community corrections toward a model that is predictive in a humane sense, identifying needs early rather than punishing failures after the fact.

The Peril: Bias, Opaqueness, and the Human Cost

AI also introduces unique dangers. Because criminal justice is grounded in due process, the right to confrontation, and the requirement of transparency, even small errors or structural flaws in AI can produce major injustices.

AI tools learn from existing data. In criminal justice, those data often reflect decades of unequal policing, prosecutorial discretion, and sentencing patterns. If tools are trained on racially skewed arrest data, they will reproduce that skew. If predictive-policing models rely on areas with historically high arrest rates rather than actual crime rates, they will reinforce cycles of over-policing. A recent investigation found that one risk assessment tool "mislabeled white defendants as low risk more often than Black defendants." *See* Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), https://tinyurl.com/34na8u45. Such instances demonstrate the risk that algorithmic tools can project a veneer of scientific legitimacy over bias embedded in the data.

Many AI tools used today are proprietary. Developers frequently claim that revealing code or training data would compromise trade secrets. But secrecy is fundamentally incompatible with adversarial proceedings. In *State v. Loomis*, the Wisconsin Supreme Court warned that proprietary algorithms could impede a defendant's ability to meaningfully challenge sentencing evidence. 881 N.W.2d 749 (Wis. 2016). Although the court ultimately permitted the use of one risk assessment tool, it did so with strong cautions, emphasizing that courts may not treat risk scores as determinative and must explicitly acknowledge their limitations. *Loomis* underscores the constitutional tension inherent in algorithmic evidence: When a defendant cannot examine or contest the basis for an AI-generated conclusion, the right to cross-examination becomes largely theoretical.

Furthermore, people tend to treat computerized outputs as inherently authoritative. That instinct is dangerous. An AI-generated facial-recognition "match" may be wrong, as demonstrated by a recent wrongful arrest in Michigan involving flawed facial-recognition technology. *See* Sharon Morioka, *Flawed Facial Recognition Technology Leads to Wrongful Arrest and Historic Settlement*, Quadrangle (Univ. of Mich. Law Sch., Winter 2024–2025), https://tinyurl.com/ye2at4ft. A risk assessment may rest on incomplete or biased data, and an AI-enhanced video frame may exaggerate artifacts or distort color. The danger is not that AI provides information; it is that courts and practitioners may treat AI-generated conclusions as conclusive.

These concerns become even more acute when AI is paired with modern surveillance technologies. AI magnifies contemporary surveillance power. The US Supreme Court recognized in *Carpenter v. United States* that aggregated cell-site location monitoring enables "near-perfect surveillance." 138 S. Ct. 2206 (2018). Today's AI-enhanced systems—facial recognition, smart-doorbell networks, predictive mapping, and real-time tracking—go even further. Without strict judicial oversight, AI risks normalizing forms of surveillance that once would have required probable cause and a warrant.

The Gatekeeping Function: Admissibility Under Rule 702

Judges have long served as guardians of scientific and technical evidence. AI tools—whether used to identify suspects, score pretrial risk, analyze forensic patterns, or enhance digital media—fit squarely within that responsibility. The legal standards already exist; the challenge is applying them to new kinds of evidence.

Relevance, Prejudice, and Foundation

AI-generated or AI-enhanced evidence must satisfy the familiar foundations:

- Rule 401 (Relevance): The evidence must make a fact more or less probable.
- Rule 403 (Unfair Prejudice): AI outputs may appear more authoritative than they are, making this balancing test essential.
- Rule 901 (Authentication): Proponent must show that AI-altered evidence (e.g., enhanced video) accurately reflects the original.

If an AI tool "cleans up" or "clarifies" digital video, counsel must demonstrate how that enhancement was performed, whether it risks adding content, and

whether its output is a fair representation of the original. These are traditional requirements, even if the tools are new.

Frye *and General Acceptance*
A minority of states, including California, Texas, and New York, continue to follow *Frye v. United States*, which admits scientific evidence only if it is "generally accepted" in the relevant field. 293 F. 1013 (D.C. Cir. 1923). Many AI systems—particularly predictive-policing software, proprietary risk tools, and novel machine-learning forensic models—may lack the long-term validation necessary to satisfy general acceptance.

In *Frye* jurisdictions, this may act as a protective brake, preventing courts from admitting untested, opaque algorithms prematurely. This suggests that *Frye*'s consensus-based approach may serve as a stabilizing check when technology evolves faster than judicial doctrine.

Daubert *and Rule 702: Scientific Validity and Application*
Most jurisdictions now follow Fed. R. Evid. 702 as interpreted by *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, which requires judges to examine

- Testability.
- Peer review.
- Known or potential error rates.
- Standards controlling the technique's operation.
- General acceptance.

509 U.S. 579 (1993). Applied to AI, the *Daubert* framework requires courts and lawyers to ask

- Has the model been tested on representative data?
- Have results been independently peer-reviewed?
- Are error rates documented and disclosed?
- Are there standards governing the model's use?
- Is the tool accepted in the relevant scientific community?

These questions must be asked even when the judge does not understand the algorithm's programming because *Daubert* demands external markers of scientific rigor.

The 2023 amendments to Rule 702 reinforce this requirement by clarifying that the proponent must demonstrate, by a preponderance of the evidence, that the expert reliably applied the methodology to the facts of the case. The amendment squarely affects criminal practice, where courts too often have treated disputes over forensic or technical evidence as matters of "weight" rather than admissibility. Under the amended rule, judges must act as true gatekeepers and require proof of reliability before allowing the jury to hear the testimony. This applies with full force to AI-generated conclusions: Courts must ensure that the technology was properly validated and properly applied, not merely that it exists.

Investigative vs. Evidentiary Use
A crucial distinction is the difference between AI used as an investigative tool, such as a facial-recognition match that gives police a lead, and AI used as evidence, such as a match introduced at trial as proof of identity. Investigative leads may be useful and appropriate, but evidentiary use requires a full *Daubert* or *Frye* analysis, appropriate disclosure, and meaningful adversarial testing.

Confrontation Clause Challenges
When AI outputs function like testimonial assertions, confrontation concerns arise. *Crawford v. Washington* holds that defendants must be able to challenge such statements through cross-examination. 541 U.S. 36 (2004). If the AI's code cannot be reviewed, the training data cannot be examined, and the method cannot be tested, then defendants cannot meaningfully confront the basis of the evidence.

Chain of Custody and Transparency
When AI processes or alters evidence, the processing itself becomes part of the chain of custody. Parties must document the original digital file, the AI tool used, the specific version or model, the parameters applied, and the output produced. *See Part Two: AI in the Courts: Ethical Challenges*, Just. Speakers Inst. (2025), https://tinyurl.com/yk2fr5wb. Each step becomes a link in the evidentiary chain, and a missing link undermines reliability.

Ethics and Advocacy: Competence in an Algorithmic Era

AI fundamentally alters the professional responsibilities of prosecutors and defense attorneys. Ethical rules drafted before deep-learning systems existed now apply to technologies that affect liberty, due process, and fairness.

Prosecutors' duties under Model Rule 3.8 extend beyond seeking convictions; they require pursuing jus-

tice with full transparency and reliability. *See* Model Rules of Pro. Conduct r. 3.8 (Am. Bar Ass'n 2024). This includes ensuring that any AI tools used in investigations are reliable, disclosing their methodologies and limitations, avoiding overreliance on automated outputs, disclosing exculpatory information surfaced through AI analysis, and refraining from presenting AI-generated evidence that lacks an adequate foundation. Failure to disclose a tool's limitations, particularly its error rates or known biases, may violate both *Brady* and Model Rule 3.8.

Model Rule 1.1 requires attorneys to maintain technological competence, a requirement that now extends to understanding how AI influences criminal evidence. *See id.* r. 1.1. For defense counsel, competent representation includes understanding risk-assessment and predictive tools; recognizing when AI has shaped or altered evidence; challenging the validity of the underlying methodology, training data, or error rates; requesting source code and technical documentation; and retaining technical experts when necessary. Without these steps, defense counsel cannot meaningfully confront AI-generated evidence. These challenges are compounded for indigent defendants, who face clear disadvantages when confronting complex AI-based evidence; in such cases, courts should consider appointing defense experts to ensure meaningful adversarial testing.

These concerns extend directly to the use of risk-assessment tools, which influence decisions at bail, sentencing, and supervision. Ethical handling requires attorneys to avoid treating risk scores as determinative; to understand the limitations of widely used tools, including those criticized in ProPublica's analysis; and to be aware of laws like Idaho's 2019 transparency statute, which mandates disclosure of algorithms' logic and data. *See* Idaho Code § 19-1910 (2019). If attorneys rely on risk scores without understanding their weaknesses, they risk misleading the court.

Similar ethical duties apply when dealing with predictive-policing data. Predictive-policing tools analyze historical arrest and enforcement records to forecast where crime is "likely" to occur, but because they rely on past policing patterns rather than verified crime rates, they often reproduce and amplify long-standing enforcement disparities. *See* Kristian Lum & William Isaac, *To Predict and Serve?*, 13 Significance, no. 5, Oct. 2016, at 14. When an arrest arises from a predictive-policing deployment in an already overpoliced neighborhood, both prosecutors and defenders must examine whether the underlying arrest is based on an actual crime or the accumulated effects of biased enforcement. Without that scrutiny, predictive tools risk reinforcing inequities rather than improving public safety.

Real-World Applications: Where AI Meets the Criminal Docket

AI now touches nearly every phase of the criminal legal process, and each domain presents unique opportunities as well as significant risks. In the pretrial stage, risk-assessment tools can help courts consider the likelihood of failure to appear, the risk of new criminal activity, and the appropriate conditions of release. Used improperly, however, they can function as automated detention triggers. The Supreme Court made clear in *United States v. Salerno* that pretrial detention requires strong procedural safeguards, meaning risk tools must serve as one factor among many and never the determinative basis for denying liberty. 481 U.S. 739 (1987).

Myimages/Adobe Stock

During discovery, AI-based review platforms are capable of categorizing millions of files and detecting relevance more accurately than traditional manual review. Courts must ensure that both sides have access to validation data, the methodology is disclosed, limitations are clearly identified, and technology-assisted review does not obscure exculpatory material. AI can assist with *Brady* obligations, but it cannot replace them.

AI's growing role in forensic evidence presents similar challenges. Tools that interpret DNA mixtures, ballistics, or voice patterns must still satisfy *Daubert*'s requirements of testability, peer review, known error rates, established standards, and general acceptance. If a forensic algorithm lacks transparency or meaningful validation, it should not be admitted. Courts must demand the same rigor that is expected in more traditional forensic sciences.

Sentencing is also vulnerable to inappropriate reliance on AI. Risk-assessment scores may exaggerate future risk, depend on socioeconomic variables, treat group characteristics as individual traits, or obscure the reasoning behind their conclusions. *Loomis* warns courts not to treat algorithmic scores as determinative; judges must articulate independent reasoning and resist substituting algorithmic predictions for individualized sentencing analysis.

AI's influence extends into community supervision as well. Recent innovations include the Risk Assessment tools for Community Supervision, which can analyze behavior, environment, and triggers in real time to help officers prioritize interventions, and the AI tools that combine smartphones and wearables to detect stress, destabilization, or relapse indicators while providing individualized support. *See, generally*, *Artificial Intelligence and the Courts*, *supra*. When used properly, these systems can reduce violations, support treatment engagement, and prevent incarceration. When misused, however, they risk transforming probation into a digital surveillance regime rather than a rehabilitative process. Courts therefore must ensure that AI-driven supervision tools incorporate strong privacy protections, avoid punitive responses to minor algorithmic alerts, maintain meaningful human oversight, and preserve clear avenues for defendants to challenge algorithmic conclusions.

Judicial Leadership

Judges occupy the central role in shaping how AI is integrated into the criminal justice system, and effective judicial leadership begins with building foundational AI literacy. Judges need not become technologists, but they must understand how AI systems generate outputs, how errors occur, how bias can enter models, and how to demand meaningful scientific validation. Much like the arrival of DNA evidence in the 1990s, AI represents a new frontier of scientific testimony that requires sustained judicial education. Leadership also involves developing the emerging "common law of AI," as courts gradually construct a body of precedent addressing the admissibility of AI-generated evidence, defense access to source code, standards for transparency in risk-assessment tools, the obligations of prosecutors under Model Rule 3.8, and the parameters of required human oversight. Each opinion contributes to a more coherent legal framework, and judges must articulate clear, principled reasoning to guide future cases. Finally, judicial leadership is essential to preserving public trust. AI systems that operate in secrecy undermine legitimacy and transparency, fairness, and explainability. These are not merely technical aspirations; they are constitutional commitments. Judges must demand disclosures, scrutinize limitations, articulate appropriate warnings, and reject opaque tools. Courts that govern AI rigorously will maintain public confidence even as technologies continue to evolve.

Policy and Governance: The Five Guardrails Framework

Courts should anchor their AI governance in a principled, practical framework. *See generally Artificial Intelligence and the Courts*, *supra*. These Five Guardrails offer a roadmap for balancing innovation with constitutional fidelity.

Guardrail One: Transparency and Explainability

No judge should rely on an algorithm they cannot understand or explain. This does not mean judges must grasp the technical details of machine learning; rather, courts must insist on clear documentation of inputs and outputs, an explanation of how variables influence results, disclosure of training data sources, and a candid articulation of known limitations. The Council of Europe's Ethical Principles for AI emphasize explainability as a core requirement for rule-of-law systems. *See* Council of Europe, European Comm'n for the Efficiency of Just (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (2018),

https://tinyurl.com/yc45f976. Without explainability, adversarial testing collapses. Opaque systems, like the one at issue in *State v. Loomis*, undermine the fairness of criminal adjudication.

Guardrail Two: Independent Validation and Peer Review

AI tools must undergo rigorous, independent validation before courts rely on them. Under *Daubert*, courts must examine testability, error rates, peer review, and general acceptance. And in jurisdictions that still apply the *Frye* standard, courts must ensure that any AI methodology reflects genuine, demonstrable acceptance within the relevant scientific community, not superficial popularity or vendor claims. Proper validation requires testing on diverse and representative datasets, documenting accuracy and error rates, assessing disparate impacts across racial and demographic groups, and publishing the findings or making them available to the defense. Without meaningful validation, AI is not evidence. It is conjecture.

Guardrail Three: Human Oversight and Accountability

AI must be advisory; people must remain accountable. The Center for Security and Emerging Technology's (CSET) *AI for Judges* framework stresses that judges must retain interpretive authority; AI cannot replace individualized analysis. *See* James E. Baker et al., Ctr. for Sec. & Emerging Tech., *AI for Judges: A Framework* (2021), https://tinyurl.com/225hw5j2. Judges must apply the tool to the facts, not allow the tool to supplant judicial reasoning.

Similarly, prosecutors and defenders must not treat risk scores or matches as conclusive. They must interpret AI evidence with caution, contextualize its limitations, and avoid automation bias.

Guardrail Four: Ethical Procurement and Vendor Transparency

Courts cannot rely on tools whose creators refuse to be transparent. Procurement contracts therefore must require disclosure of training data sources, documentation of any bias-mitigation strategies, availability of technical materials for defense review, audit rights, and disclosure of any ownership interests that could affect neutrality. Idaho's 2019 transparency law is the leading model, prohibiting vendors from invoking trade-secret defenses when their risk-assessment tools are used to make liberty decisions. Every state should follow this example.

Guardrail Five: Continuous Review and Corrective Feedback

AI systems evolve over time, and courts must ensure that their accuracy does not degrade as models are updated, datasets shift, or vendors make unannounced modifications. An algorithm that performed acceptably at initial use may behave very differently after ingesting new data or after a vendor adjusts internal parameters, changes that often occur without judicial awareness. Continuous review, therefore, requires ongoing accuracy testing, routine audits for racial or demographic bias, close monitoring of error rates, and evaluation of actual case outcomes against the tool's predictions. Courts must demand periodic validation reports and require vendors to notify them of any substantive model changes that could affect reliability. If a system begins to perform poorly or exacerbates existing disparities, it must be corrected, suspended, or withdrawn entirely. AI cannot be treated as "set and forget"; it demands active governance, continuous oversight, and a willingness to intervene when technology undermines rather than advances justice.

Conclusion: Innovation with Integrity

AI already has entered the criminal courtroom. It influences how investigations unfold, how evidence is processed, how bail is set, how forensic conclusions are reached, how sentences are determined, and how individuals are supervised. AI offers enormous promise, but it also carries profound risks. Courts must never choose between innovation and justice; they must insist on both. Harnessing AI's benefits requires transparency, independent validation, ethical use, meaningful confrontation, robust judicial oversight, and ongoing review. Rejecting these principles risks embedding discrimination, obscuring reasoning, and eroding constitutional protections. Embracing AI without scrutiny risks automating injustice. The balance is complicated but achievable. Ultimately, the future of criminal justice will depend not on whether machines can think, but on whether legal professionals will critically assess their results. The courtroom must remain a place where evidence is tested, not presumed; where technology supports truth but never replaces it. And at every stage, people, not algorithms, must remain the ones who make the decisions that determine liberty and justice.