

AI in Criminal Justice: Why Governance Matters and How to Make It Work

- March 27, 2026

- By

- Jonathan J. Wroblewski

- Related Organization(s):

- [Law & Policy Lab](#)

[Law & Policy Lab](#)

- [SUBSCRIBE](#)

- [Share on Twitter](#)

- [Share on Facebook](#)

(Originally published in the [Sentencing Matters](#) Substack on March 26, 2026)

Stanford Law School lecturer Jonathan Wroblewski

Artificial intelligence is no longer a distant or speculative technology in the criminal justice system. It is becoming part of its everyday machinery.

Police departments use algorithmic tools to analyze digital evidence, look for patterns in crime and other data, and draft police reports. Prosecutors rely on AI software to manage discovery and support charging decisions. Courts encounter algorithmic risk assessments and large language models that promise to summarize records, draft documents, and assist with legal analysis. Across the system, AI is being woven into decisions that affect liberty itself.

I had the great privilege this past fall of working with a talented student research team at Stanford Law School to examine how to close the growing gap between the pace of technological change around artificial intelligence in criminal justice and the capacity of criminal justice institutions to govern it responsibly. The work was part of the law school's [Law and Policy Lab](#) program, which gives students hands-on experience advising individuals, government agencies, and non-profit organizations about

current policy issues in real time. Our team worked in partnership with the Council on Criminal Justice's (CCJ) Task Force on Artificial Intelligence. The answer we reached on AI and criminal justice is quite simple and increasingly urgent: we need specialized AI governance entities for the criminal justice system.

Among the team's findings were that —

- AI technologies bring real promise. When they are carefully designed and responsibly governed, AI systems can help institutions use scarce resources more effectively, improve consistency, reduce backlogs, and expand access to legal and administrative services. They make it possible to process volumes of data and documentation that would be impossible to manage by hand. In a system that is chronically under-resourced and overburdened, those capabilities are understandably attractive.
- At the same time, the risks around the use of AI in criminal justice are serious. Algorithmic systems can embed and amplify bias, operate in ways that are hard to see or understand, generate unreliable or misleading outputs, and shift decision-making power in directions that are difficult to monitor and control. Errors may not be obvious or isolated; they may be subtle, systematic, and persistent — especially for users without technical expertise. And the stakes are incredibly high, as these tools increasingly influence who is stopped, searched, charged, detained, sentenced, supervised, or released. When they fail, the consequences reach beyond efficiency and accuracy to constitutional rights, democratic accountability, and the legitimacy of the justice system itself.
- As well-intentioned as policymakers and criminal justice practitioners, including lawyers, judges, and police officers, may be, without additional guidance, the lack of sufficient understanding of AI technology and its development may undermine their ability to effectively apply standards — including those developed by the CCJ Task Force — to specific use cases in a rapidly evolving AI landscape. For example, a judge may want to use a large language model (LLM) to help interpret a provision of a contract but may not fully understand the way the LLM works, how it was engineered, the risks of unreliable results, bias, and other concerns. Police departments may want to use an AI product to process, analyze, or transcribe body camera video and audio but similarly may lack the expertise to identify the risks, reliability, or security around the product.

The Policy Lab team produced a White Paper that you can find [here](#). The central problem the paper addresses is this: **Artificial intelligence**

capabilities and products are being developed and deployed in criminal justice without sufficient understanding of how they work, attention to their failure risks, or analysis of how they implicate constitutional and other values. Judges, prosecutors, police officers, probation officers, and policymakers are being drawn to adopt complex technologies without the technical background needed to assess their design, limitations, or possible failures. While some criminal justice agencies will have the capacity to put together multi-disciplinary teams to evaluate AI technology, most will not. Vendors market these systems directly to practitioners, most of whom lack both the time and expertise to conduct meaningful independent evaluation. Without governance, AI in criminal justice will be perilous.

The paper builds upon the CCJ Task Force's efforts to establish governing principles, procurement standards, and case studies for safe, ethical, and effective use of AI in criminal justice. It extends that work by examining potential structures and mechanisms through which Congress, the Executive Branch, State and local governments, and non-governmental organizations too, could institutionalize the Task Force's recommendations, and provides a foundation for assessing durable governance frameworks that can translate expert recommendations into actionable policies and implement them.

The report comes on the heels of last week's White House release of the President's *National Policy Framework for Artificial Intelligence*. That framework calls for establishing federal policy that preempts state AI laws and takes a light-touch approach to AI governance. It's based on a belief that our existing institutions will be able to manage AI's risks. In theory, this might be right. Judges can exclude unreliable evidence. Prosecutors can decline to use untrustworthy tools. Police departments can vet vendors carefully. And they all can use thoughtful standards developed by professional organizations and task forces, like the CCJ's. In practice, this is not what's happening.

Good guidelines and standards are necessary, but they are not sufficient. The CCJ Task Force is doing vital work in articulating governing principles, procurement standards, and case studies for safe and ethical AI use in criminal justice. The question now is how to deploy and implement that guidance — how to build durable structures that can translate expert recommendations into concrete policies, and support practitioners in applying them case by case, tool by tool.

That is the role an AI governance entity must play.

Such an entity — housed within the federal government, a consortium of states, or a public-interest institution with formal authority and stable funding — could serve as a hub of technical, legal, and ethical expertise. It could independently evaluate high-impact AI tools used across criminal justice, publish clear and accessible assessments, and maintain up-to-date guidance on acceptable uses, known risks, and necessary safeguards. It could develop standardized benchmarks and testing protocols, coordinate incident reporting and auditing, and provide technical assistance to agencies that cannot afford their own AI review teams.

A governance entity would bridge the gap between general principles and messy reality. It could help the judge decide whether and how a large language model might responsibly assist with reviewing a complex contract. It could help a police department understand the security, accuracy, and bias implications of an AI product designed to process body camera video. It could translate rapidly evolving technical details into actionable advice that aligns with constitutional requirements and community expectations.

AI is a set of technologies that can reconfigure where power lies in the justice system — between public and private actors, between technical experts and frontline practitioners, and between the state and the individuals whose liberty it constrains. Leaving that transformation to uncoordinated adoption and after-the-fact litigation is a recipe for invisible and visible harms and lost public trust. In the criminal justice context, light-touch governance combined with broad preemption risks creating precisely the vacuum we can least afford: a patchwork of vendor claims and ad hoc local experimentation, with little independent capacity to test, challenge, or correct.

We have already begun to articulate what responsible AI in criminal justice should look like. The next step is to build the institutions that can make those ideals real in everyday practice. A dedicated AI governance entity — designed to be expert, independent, transparent, and accountable — will not solve every problem. But without it, we are asking our criminal justice system to navigate a technological revolution blindfolded.

The tools are here. The stakes are clear. What is missing is the governance capacity to match them. It is time to build it.

— — —

I am grateful to have worked on this project with wonderful students who did the research and wrote the White Paper — Isabel Astrachan JD '27, Tejaswita Kharel LLM '26, Archit Lohani LLM '26, Anna Makena McGuire JD '27, Alexander Rivkin JD '27, Emi Sakamoto BA '28, Bianka Sedmakova, Visiting Exchange Student from University of Vienna '26 , Mirac Suzgun JD / Ph.D. (Computer Science) '26 , Victor Y. Wu JD '25 / Ph.D. (Political Science) '28 — and with Professor Robert Weisberg and Luciana Herman, Ph.D., Program Director of the Law & Policy Lab Program, both wonderful collaborators and extraordinary teachers.

The [paper](#) is worth a full read.
Here is its Executive Summary.

Executive Summary

Artificial intelligence (AI) is rapidly embedding itself into the core machinery of the criminal justice system, powering everyday decisions from police analysis of digital evidence and pattern detection in crime data to prosecutorial discovery management, charging recommendations, algorithmic risk assessments in courts, and large language models for summarizing records and drafting documents. These tools promise efficiency gains — processing vast data volumes, reducing backlogs, and optimizing scarce resources in an overburdened system — but they also carry profound risks: embedding bias, producing opaque or unreliable outputs, shifting unmonitored power to vendors, and influencing high-stakes liberty decisions like arrests, detention, sentencing, and release. The central problem is that AI capabilities are deploying without sufficient understanding of their mechanics, failure modes, or implications for constitutional rights, democratic accountability, and system legitimacy. Criminal-justice entities who encounter AI tools — thousands of under-resourced police departments, prosecutors' offices, courts, and probation units — lack the technical expertise to evaluate these tools rigorously, while vendors market directly to practitioners. This creates a governance gap: even well-intentioned actors cannot reliably apply emerging standards amid rapid technological changes, risking uneven, superficial oversight that undermines public trust.

Objectives of this White Paper

This White Paper seeks to bridge that gap by synthesizing institutional design principles for AI governance in criminal justice. It examines how other policy domains have addressed similar challenges of expertise asymmetry, evolving technologies, and high-stakes accountability. Our research formulates a twofold question: (1) What existing government entities could model — or directly assume — AI oversight roles? (2) How do those entities measure against desirable attributes, revealing strengths, weaknesses, and hybrids? Starting from the premise that fragmented agencies need pooled support to properly govern the use of AI in criminal justice, we initially leaned towards a Federal governance model. We now have reconsidered this and believe that multi-level (Federal, State, local) governance with coordination may be the most feasible in today's politics. The goal is to equip policymakers at all levels with a framework to stand up durable institutions that operationalize the CCJ Task Force on AI's standards, ensuring AI enhances rather than erodes justice.

Institutional Design Criteria

We evaluate governance models across six core institutional design criteria that are consistent with CCJ Task Force's principles, balancing creation ease with sustained impact:

- **Startup Feasibility:** Legal/political/administrative ease of launch (e.g., executive action vs. statute), funding needs, and ramp-up time, prioritizing speed for urgent AI risks without sacrificing viability.
- **Relevant Expertise:** Dual composition with decision-makers with AI/criminal justice knowledge (legal, ethical, community) along with staff for technical assessment, empirical monitoring, and evaluation.
- **Transparency and Accessibility:** Open processes (public meetings, data releases), input mechanisms (hearings, comments), and usable outputs to enable scrutiny, error-correction, and stakeholder trust.
- **Organizational Stability:** Continuity amid politics (term structures, budgets, legal basis) to build institutional memory, iterate rules, and endure political changes.
- **Policy Influence:** Impact of outputs such as binding rules, required agency responses, grants, or reputational standards that drive uptake beyond advice.
- **Responsiveness:** Agility for fast-evolving AI, including expedited reviews, paired with slow-path monitoring, such as periodic data-driven updates.

Institutional Models Examined

1. FACA-Style Federal Advisory Committees

FACA committees, created under the 1972 Federal Advisory Committee Act, are renewable advisory bodies with a tenure of at least two years that provide balanced, expert advice to the Executive Branch. Examples include the National Commission on Forensic

Sciences (NCFS) and the FBI's Criminal Justice Information Services (CJIS) Advisory Policy Board. These advisory committees generate non-binding recommendations.

2. Legislative Agencies

Legislative agencies support Congress with non-partisan analysis on complex issues. Examples are the Government Accountability Office (GAO) and the now-defunct Office of Technology Assessment (OTA, 1972–1995). GAO issues oversight reports and mandates agency responses, while the OTA produced intensive research on various technological developments. The role of legislative agencies is to inform legislation via expert reports, testimonies, and audits on emerging technologies like AI.

3. Sentencing Commissions

Sentencing commissions are independent expert bodies that develop system-wide standards for criminal justice decision-making through empirical analysis and iterative review. Examples include the U.S. Sentencing Commission and state commissions in North Carolina and Minnesota. These commissions combine pluralistic stakeholder representation with standing research staff, operate through transparent procedures, and regularly update guidance in response to observed outcomes. Although their authority is typically limited to sentencing, sentencing commissions illustrate how nonbinding or semi-binding guidelines can achieve influence and legitimacy in technically complex domains.

4. Congressional Committees

Congressional committees encompass standing committees, subcommittees, and joint committees with direct congressional oversight. Examples include House/Senate Judiciary Committees and subcommittees, along with joint committees like the Joint Committee on Taxation (JCT). Their role involves conducting hearings, investigations, and bill drafting to shape policy on legal and constitutional matters.

Comparative Evaluation of the Models

No single institutional model performs well across all six design criteria. Instead, each reflects a different tradeoff among startup feasibility, expertise, stability, influence, and responsiveness. Advisory and legislative models are easiest to activate and can rapidly elevate AI governance on the policy agenda, but they often struggle to sustain deep technical capacity or long-term oversight. More durable expert bodies, by contrast,

tend to accumulate legitimacy and analytic depth over time, but face higher political and administrative barriers to creation and may be slower to respond to rapidly evolving technologies.

FACA-style advisory committees and congressional subcommittees score highly on startup feasibility and transparency. They can be created quickly, convene diverse stakeholders, and generate public-facing guidance or hearings that shape early norms. However, their advisory nature, exposure to political turnover, and limited staffing constrain their ability to conduct sustained technical evaluation or monitoring. As the experience of the National Commission on Forensic Science illustrates, even highly effective advisory bodies may be short-lived, while congressional subcommittees remain vulnerable to shifting leadership priorities.

Legislative agencies and joint congressional committees offer deeper analytic capacity and greater institutional continuity, especially when supported by professional staff and formal reporting obligations. Entities such as the GAO demonstrate how nonbinding recommendations can nonetheless exert real influence through mandatory agency responses and reputational authority. At the same time, these models operate far upstream from day-to-day criminal justice practice and often trade speed for rigor, limiting their ability to provide timely, use-case-specific guidance in a fast-moving AI environment.

Sentencing commissions stand out as a hybrid model that balances expertise, stability, transparency, and iterative governance. Their experience shows how pluralistic decision-making bodies paired with standing research staff can translate technical analysis into durable standards that shape practice over time, even without broad enforcement authority. While their jurisdiction is narrower and their creation more demanding, sentencing commissions illustrate how AI governance institutions can earn legitimacy, adapt empirically, and persist across political cycles—suggesting that the most effective approach to AI governance in criminal justice may draw selectively from multiple models rather than replicating any single one wholesale.

The synthesis of these models reveals several consistent design principles. Durable AI governance requires permanence rather than ad hoc or temporary arrangements, “two-tier” expertise that pairs representative decision-makers with standing technical and research staff, and transparency that is structurally embedded rather than left to institutional culture. Influence need not depend on formal binding authority; mechanisms such as required agency responses, conditional funding,

procurement standards, and credible technical evaluations can drive meaningful uptake even from advisory bodies. Responsiveness should combine rapid-response mechanisms for emerging capabilities with slower, deeper processes for auditing and post-deployment monitoring.

These findings push toward a hybrid institutional approach that borrows selectively from multiple models rather than replicating any single one. Advisory committees and congressional subcommittees offer speed and agenda-setting power but struggle with durability and sustained technical depth. Legislative agencies provide analytic rigor and institutional continuity but operate too far upstream from day-to-day practice. Sentencing commissions come closest to balancing expertise, stability, transparency, and iterative governance, but their creation demands greater political investment. The most effective path forward will likely combine elements across these models—matched to the appropriate level of government—to close the growing gap between AI's pace of deployment and the justice system's capacity to govern it responsibly.