Regulating Generative AI in Criminal Courts

Criminal Justice Policy, Institute for Youth in Policy

February 2026

I. Executive Summary

This policy brief analyzes the potential of Artificial Intelligence (AI) in increasing the efficiency of the current justice system. While examining previous attempts at AI use and its now widespread adoption in the field of criminal justice, this paper evaluates areas of success as well as adverse effects. Although it has been suggested that AI can streamline administrative processes and reduce space for human error, it has the capacity to further augment pre-existing inequities within the criminal justice system. The analysis conducted considers thatAI is yet to harness the trust of the general public and that this technology's "fairness" can be ambiguous. In response to these concerns and in an attempt to maximize AI benefits, this policy brief emphasizes on the need to establish supervised and open-source AI systems. Human oversight should be considered fundamental in this process, as should education. As a whole, this brief aims to outline the complexities involved with examining the fruitfulness of AI in criminal justice and suggest remedies to its challenges.

II. Overview

The use of AI in the criminal justice system is a rapidly evolving field. While the technology has advanced significantly in recent years, its roots trace back to 1998 with the introduction of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system. Developed to prevent recidivism, the tendency of defendants to become repeat offenders, COMPAS was deployed across U.S. courts and correctional facilities by 2000, shifting risk assessment from professional human judgment toward algorithmic analysis.

In the decades since, algorithmic technology like COMPAS has processed millions of profiles in the criminal justice system. However, their widespread use has not been without controversy. A 2016 ProPublica investigation revealed that COMPAS wrongly predicted that Black defendants were almost twice as likely to be labeled as higher risk than white defendants. These racial biases sparked debate over the capability of algorithms in delivering fair, unbiased judgment and outcomes.

Despite these concerns, AI's influence in law enforcement continues to expand. Advanced machine learning now powers facial recognition, object detection, and predictive policing patterns. Digital surveillance in the U.S. continues to rise, with 90% of law enforcement agencies having adopted AI in 2025, a 55% surge from the previous year alone. This policy brief examines the potential of AI to enhance criminal court operations and how such technology can be utilized responsibly through oversight, transparency, and community-led safeguards to ensure that technology serves the cause of equity.

III. History

A. Current Stances
Proponents of regulating generative AI in criminal courts argue that in the ever-evolving digital landscape, judicial regulations around AI usage need to keep up with the times. According to the Northern District of New York Federal Court Bar Association, in recent years generative AI has been used to cite fictitious cases, inapposite cases, fake academic articles, incorrect citations of authors of academic articles, false quotations from cases, false quotations from a Restatement of the Law, use of AI to respond during a discovery conference, and more. Although there are existing legal

frameworks, these may not effectively govern generative AI's impacts in legal settings. For example, Section 230 of the 1996 Communications Decency Act, which has been heralded as a hallmark of digital protections, provides limited federal immunity to "interactive computer services," allowing them to host user-generated posts without being treated as the publisher. However, there have been numerous ongoing debates about whether these protections apply to AI generated content as well. Many point to this lack of clear legal coverage as to why increased generative AI-focused regulation is necessary.

Judicial leaders and courts are pointing to the errors caused by generative AI use as demonstrating a need for increased accountability and accuracy. Among these leaders, the 7th US Circuit Court of Appeals has emphasized litigants' responsibility to prevent AI "hallucinations" in legal documents while senior judges in the UK have warned that relying on generative AI could risk facing serious court sanctions.

Opponents of regulating generative AI in criminal justice warn that overly rigid regulation may prematurely stifle innovation or deter the benefitsl of AI tools. They argue that the full potential of generative AI hasn't yet been discovered. If too many regulations are put in place now, it may unintentionally hinder generative AI usage once it does become a more reliable tool.

Generative AI has also helped make participation in the court system more accessible. For people who can't afford expensive legal representation, generative AI helps pro se litigants better understand what they're doing. The use of generative AI also helps experienced legal practitioners improve efficiency in many areas. Clerks' offices are able to use it to streamline document processing and public defenders have used it to aid in the initial drafting of legal documents and research.

B. Tried Policy
Many states have started to address generative AI use in their courts in the past few years.

- In 2024, Arizona's state Supreme Court adopted the Use of Generative Artificial Intelligence Technology and Large Language Models rule which included limited "non-production" use for testing and evaluation following a January 2025 amendment.
- In February 2025, Nevada's state Supreme Court created an Artificial Intelligence Guide for Courts and Judicial Officers. This included a dedicated two pages examining what generative AI is, how it works, and how it impacts judicial officers.
- In March 2025, Texas' Supreme Court Advisory Committee Rule 1-14c Subcommittee issued an update regarding generative AI use. They opted to continue relying on existing evidence rules to manage reliability disputes, rather than creating generative AI-specific guidelines. They cautioned that taking specific action now would be more likely to "result in unintended consequences and likely will not significantly reduce the risk that unrepresented parties misuse AI tools in their representations and submissions to the court".
- In June 2025, Arkansas proposed an order addressing generative AI use. Their order covers awareness about generative AI usage and prohibits the Administrative Office of the Courts staff from intentionally exposing state courts' internal data to generative AI.

Internationally, the regulation of generative AI use has also been a changing landscape.

For example, in Australia, the South Australian Supreme Court issued formal guidelines for AI use by lawyers. Specifically, they said that lawyers are not required to disclose generative AI use except when asked. However, they are still fully

responsible for the accuracy of the generative AI they use and may face serious consequences for misuse. In Kerala, India, the High Court implemented explicit guidelines prohibiting judicial officers and staff from using AI tools for decision-making or legal reasoning. The guidelines apply to judicial officers, staff interns, and law clerks in Kerala. They are meant to prevent negative consequences such as violation of privacy rights, data security risks, and erosion of trust in judicial decision making.

IV. Policy Problem

A. Stakeholders
The use of artificial intelligence in criminal courts aims to improve efficiency and consistency, affecting courts, policymakers, and technology developers. These groups interact to integrate AI tools into court processes, establish oversight, and guide development that shapes how AI influences legal decision making. An effective AI system should support fair outcomes by improving accuracy and reducing administrative burdens.

In order to achieve this effectiveness, defendants must be prioritized to protect due process and maintain trust in the justice system. This is necessary not only to safeguard constitutional rights, but also to ensure individuals understand and can challenge AI influenced decisions related to bail, sentencing, or parole. So, engaging stakeholders such as public defenders, civil rights organizations, and regulators is essential to identify ethical and structural concerns that may otherwise be overlooked.

B. Risk of Indifference
The increasing use of AI systems trained on historical criminal justice data presents a significant risk of indifference within criminal courts. When judges, administrators, and developers ignore the fact that arrest records and sentencing outcomes reflect long-standing racial and socioeconomic disparities, AI systems can reinforce these inequities rather than correct them.

This indifference creates a reinforcing cycle where biased data informs algorithmic recommendations and judicial outcomes, producing new data that entrenches inequality. Indifference to transparency further threatens due process, as many AI systems operate as proprietary black boxes that prevent defendants from understanding or challenging how decisions are made.

C. Nonpartisan Reasoning
Regardless of political affiliation, fairness, accountability, and public trust in the criminal justice system are shared priorities. Ensuring AI use aligns with constitutional protections and ethical standards is a nonpartisan concern, and efficiency cannot justify practices that undermine due process.

Furthermore, judges are required to make individualized, high stakes decisions that AI cannot replace. While AI can assist by organizing information or identifying patterns, it lacks human judgment and moral reasoning. Research consistently emphasizes that AI should serve as an advisory tool rather than a determinant of outcomes.

So, with careful oversight and inclusive stakeholder involvement, AI has the potential to support justice without compromising the integrity of criminal courts, however its negatives highly outweigh positives, especially in the current political landscape.

V. Policy Options

The tension between the choice to or not to utilize AI is often due to the balance between AI's promises of efficiency and the need for human ethics, morals, or judgement in a task. This section argues that AI should most definitely be included in Criminal Courts, but be heavily restrained as to when AI can be used, who should be allowed to use it, and how the AI used in courts is made.

The efficiency that AI can provide could be incredibly necessary given that courts can become backlogged with cases easily due to inefficient planning or lack of personnel. In Palm Beach County, employees utilized a Lights-Out Document Processing program to analyze document filings while tagging and indexing them with the proper case information. The program was able to achieve a 98% to 99% accuracy rate, which "far outperformed the accuracy of their human counterparts." Further, the workload of five of these systems was equivalent to that of 19 employees. In the case of Palm Beach County, the integration of AI into their clerk's office greatly reduced the workload of employees. It should also be noted that the implementation of AI in this circumstance would not be taking jobs away from humans. Shortage of staff have caused significant backlogs in cases in multiple U.S. states including Nebraska, California, and New York, leading to delayed justice or resolution for thousands of people. The court system is already swamped with more tasks than they can handle, and the help of AI would reduce the stress on employees and avoid backlogs.

Furthermore, as AI analysis can go through much more information than the average human, using AI could enable agencies to spend more time addressing more important issues. Further, the discovery of issues can be difficult in itself, requiring frequent and deep analysis of data. "[We have so] much information [that] is sitting there that we're not using. [T]he possibilities of problems we could be solving if we just understood our data is crazy," the Executive Director of the Oklahoma Department of Corrections Steven Harpe said. Thus, AI provides the opportunity to criminal courts to identify and focus their time and energy on more important matters.

However, the usage of a robot in the research, building, or argumentation of a case in criminal court stands out as problematic in many respects. First, AI integration in courts could cause the perpetuation or amplification of existing problems and unequal access in the justice system. AI models are costly. The cost of a deep learning system necessary for complex issues and lots of data holds a starting price tag of $50,000 and can rise to $500,000. These sky high prices are posed to extend the unequal access to resources already prevalent in justice systems as modestly funded courts and agencies are likely to struggle to obtain access to models.

There is always the potential for AI systems to be manipulated or misused. Though this could exist with any other tool of the law, it is much more difficult to hold a machine accountable for a mistake and to fix said mistake. Many of the most popular AI models today operate in a "black box," meaning that the model is considered proprietary data and the entire decision making process of the system is unavailable to the public. This "black box" leaves no clues as to why the machine made a decision or how the machine came to that decision, making it difficult for people overseeing the model to determine flaws in the reasoning preemptively and to locate where exactly the model's reasoning went wrong.

The benefits and downsides of AI focus on vastly different areas of the function of criminal courts. There are obvious upsides to it in the clerical side, and obvious cons to AI in the actual argumentation of a case. Therefore, AI should be integrated into administrative work with a few key pieces of information in mind. First, to address the black box scenario and ensure maximum transparency with the public, only open source AI systems in criminal courts. Open source systems allow users to view the code of a system and sometimes make a copy and edit it. They provide key advantages that will help AI in criminal courts improve. As others are allowed to view the code of open source models, allowing for easy auditing and supervision of the models to avoid biases, locate flaws, and ensure accountability.

Second, AI training should be mandated for all personnel to ensure that human oversight of systems are meaningful and reliable, so that when issues do arise humans are able to spot and quickly block or correct a mistake. Employees should be able to understand the basics of machine learning, AI concepts, and robotics. They should know about the ethical implications of AI regarding biases and privacy and learn, as well as what is key to look for when reviewing AI.

Lastly, a large part of developing "good" AI is securing "good" data to train it with. AI is "born" knowing nothing and learns everything from the data that it is fed. Therefore, if people show AI data including human biases, erroneous judgements, or unethical decisions, AI will also learn to adopt those same behaviors and utilize and amplify them in its responses. Recognizing innate human biases and eliminating them from data sets is vital to ensuring that AI systems remain unbiased.

In conclusion, AI holds potential, but its many flaws make it unfit for usage in the argumentation of cases and AI should always be supervised by humans. AI integration should go far beyond the step of implementation, otherwise it may simply serve as a false solution, where AI implementation creates an illusion of progress.

VI. Conclusions

In recent years, there is substantial growing interest in utilizing artificial intelligence in the criminal justice system especially to enhance justice system operations in criminal courts. Artificial intelligence can accurately identify patterns, mitigate human errors in decision making processes, and increasingly help agencies allocate resources in the most effective way. These open AI systems can create transparent and clear audit trails that allow the public and the decision makers to understand how the decisions are made. However, in order to ensure successful implementation and deployment, there is an imperative need for meaningful human oversight, training, and careful attention to data quality to prevent the misuse of ai and existing bias. In addition, collaboration between agencies, technical experts, and community members helps to build trust and a sense of security. Together, this balanced approach harnesses AI potential in criminal courts while protecting the fairness, due process, and fundamental principles of the criminal justice system.