Principles for the Use of AI in Criminal Justice

October 2025

As artificial intelligence (AI) capabilities rapidly advance and tools using AI become more widely available, policymakers and leaders face decisions about deploying systems that can have significant direct and indirect effects on individual liberty, due process, and public safety. **AI systems offer substantial potential to improve criminal justice outcomes** by enhancing process reliability, reducing human bias, increasing efficiency, improving resource allocation, and enabling data-driven insights that can support more effective and just practices. When thoughtfully implemented, these technologies can potentially improve both public safety and justice.

At the same time, the use of AI in criminal justice contexts demands careful attention to potential risks and unintended consequences. The stakes in criminal justice require that we pay attention to how AI systems can malfunction or even undermine democratic control. The complexity and opacity of many AI systems, combined with their potential to operate at unprecedented scale and speed, mean that errors, biases, and misallocations can cause significant harm and have widespread and lasting impacts on individuals and communities. This necessitates awareness of and attention to the limitations and risks of AI technologies, robust safeguards, and ongoing vigilance to ensure that the promise of AI is realized without compromising foundational principles.

Trade-offs are fundamental features of the criminal justice system. AI deployments are no different, and such tradeoffs should be explicitly acknowledged and actively managed. Unlike technical domains where optimization might focus on a single metric, criminal justice AI systems operate within competing demands for public safety, individual rights, efficiency, fairness, democratic accountability, and other criteria. However, while some considerations can be balanced against each other, certain fundamental principles remain inviolable and cannot be sacrificed regardless of potential benefits. These include basic constitutional protections, due process rights, human dignity, and prohibitions against unlawful discrimination. **The goal should be to make deliberate, transparent choices about how to balance legitimate interests with constitutional and human rights protections, not to eliminate tensions between competing values.**

AI Principles at a Glance

AI systems that are deployed in the criminal justice system must be:

1. Safe and Reliable
2. Confidential and Secure
3. Effective and Helpful
4. Fair and Just
5. Democratic and Accountable

The purpose of these principles is to establish a framework for responsible integration of artificial intelligence in the criminal justice system. This framework is intended to support evidence-based decision-making about the use of AI in criminal justice that serves both public safety and individual rights while building public trust and remaining within the bounds of constitutional and human rights.

Scope and Definition

For the purposes of this framework, we define artificial intelligence (AI) as machine-based systems that operate with varying levels of autonomy, may exhibit adaptiveness after deployment, and infer from inputs how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Consistent with both American legislative trends and operative international frameworks, this definition accommodates both existing tools and prospective technologies.

While maintaining this broad scope, we distinguish between algorithmic tools that follow predetermined decision trees or checklists and more complex AI systems that process ambiguous inputs, learn from data, or generate novel outputs. Simple checklist algorithms that operate through rule-based logic can be easily audited and understood. In contrast, more complex **AI systems—such as machine learning models, neural networks, or generative AI—often involve opaque processes, can produce unexpected outputs, can be challenging or impossible to definitively validate, and may exhibit**

emergent behaviors that can be difficult to predict or explain. For instance, simple algorithmic tools used in criminal justice include automated warrant checks, basic case routing systems, and crime mapping that identifies hotspots. More complex AI systems include automated report writing tools, machine learning-based risk assessments, and facial recognition technologies that process ambiguous visual data. The systems that involve complex, ambiguous inputs and outputs require proportionally greater oversight and safeguards due to their potency, complexity, and lack of transparency.

Core Principles

1. Safe and Reliable

The rapid advancement of AI capabilities means that systems deployed today may behave differently than anticipated as they encounter novel situations. Unlike traditional software, many AI systems can be unpredictable or produce outputs that seem plausible but are incorrect. Criminal justice applications demand exceptional reliability because errors can result in wrongful detention, inappropriate sanctions, or public safety failures. **Safe and reliable deployment requires comprehensive testing procedures, clear protocols for management of system failures, explicit allocation of responsibilities between human operators and AI systems, and robust monitoring that can detect when systems are performing outside acceptable parameters.**

2. Confidential and Secure

AI systems in criminal justice may require access to vast quantities of sensitive personal data, including criminal histories, behavioral patterns, biometric information, and law enforcement interaction records. Advanced AI can derive intimate details about a person from seemingly innocuous data points, while interconnected systems create new vulnerabilities for data breaches, unauthorized access, or malicious manipulation. Criminal justice data carries particular sensitivity due to its lasting impact on individuals' opportunities and rights, involvement of vulnerable populations, and potential for misuse by internal actors and external parties. **Confidential and secure deployment requires robust data stewardship practices, comprehensive security measures protecting against technical vulnerabilities and insider threats, clear data sharing protocols that**

balance inter-agency cooperation with individual privacy rights, and clear and transparent policies informing individuals about AI data use. Implementation of AI in criminal justice must include regular security audits, vendor oversight ensuring third-party compliance with criminal justice data standards, and mechanisms enabling people to understand and contest the use of their personal information in AI decision-making processes.

3. Effective and Helpful

The increasing availability and marketing of AI solutions create pressure to adopt systems based on technological novelty even before the products are sufficiently tested and evaluated across critical criteria. In criminal justice, where resources are often constrained and decisions can have profound consequences on many parties involved, **AI systems must demonstrably outperform existing alternatives or achieve equivalent effectiveness at reduced cost. Effective and helpful deployment requires robust validation, real-world evaluation of benefits and limitations in practice, comparison with non-AI alternatives, and recognition that the most sophisticated existing technology is not necessarily the most appropriate solution for a given problem.**

4. Fair and Just

AI systems can potentially perpetuate, amplify, or obscure biases in criminal justice while creating new forms of discrimination and unfair treatment. As AI capabilities expand and influence more decisions, ensuring these tools are fair and just becomes both more important and more challenging. **Fair and just deployment requires regular assessment of both potential and actual impacts across demographic groups, investigation of ways to reduce disparities, rigorous comparison of AI bias with existing human decision-making processes, and recognition that pursuing improvement is valuable even in those cases where perfect fairness is unattainable.**

5. Democratic and Accountable

Complex AI systems pose unique challenges for democratic governance and public accountability. AI systems may operate through processes that are difficult and sometimes

impossible to explain even to technical experts, their capabilities may evolve rapidly beyond public understanding, and their speed and scale may outpace human ability to monitor every decision. As AI systems become more capable and autonomous, maintaining meaningful human and democratic control becomes increasingly important, while system complexity may make it difficult for human operators to understand when and how to intervene. **Democratic and accountable deployment requires public documentation of AI capabilities, limitations, and the engineering decisions underlying the systems. It also requires clear chains of responsibility for AI decisions, binding policy guardrails to ensure democratic governance before and during deployment, access to validation data, and mechanisms for ongoing feedback that can influence system development and deployment decisions.** Additionally, operators must receive adequate training to understand and appropriately interact with AI systems, and institutions adopting AI must have clear procedures for identifying, learning from, and redressing harm from AI-related issues.

Conclusion

The rapid advancement of AI capabilities creates both opportunities and risks for criminal justice. The principles outlined here provide a structure for navigating the complexities involved in AI deployment while emphasizing that responsible governance requires transparency, evaluation, and honest acknowledgment of limitations. By focusing on evidence-based decision-making and democratic accountability, these principles can support AI deployment that enhances both public safety and individual rights while building the public trust necessary for legitimate governance in the AI age.

Acknowledgements

These principles are from the Council on Criminal Justice Task Force on Artificial Intelligence, a national, nonpartisan initiative to develop standards and evidence-based recommendations to guide the safe, ethical, and effective use of AI in the criminal justice system. They are the product of its members, who graciously shared their time and expertise.

Support for the Task Force on Artificial Intelligence comes from the Heising-Simons Foundation, The Just Trust, Microsoft, Southern Company Foundation, and The Tow Foundation, as well as the John D. and Catherine T. MacArthur Foundation and other CCJ

general operating contributors.

Suggested Citation

Council on Criminal Justice. (2025). *Principles for the use of AI in criminal justice*. https://councilоncj.org/principles-for-the-use-of-ai-in-criminal-justice/