The Implications of AI for Criminal Justice

Key Takeaways From a Convening of Leading Stakeholders

October 2024

Introduction

The rapid advancement of artificial intelligence (AI) technologies has implications for every sector of society, including the criminal justice system. As AI tools for investigation, adjudication, prioritization, analysis, and decision-making proliferate and evolve, understanding their potential benefits and risks becomes increasingly important.

In June 2024, the Council on Criminal Justice (CCJ) convened a group of experts and stakeholders to discuss the implications of AI for the U.S. criminal justice system. The meeting brought together a diverse group of three dozen leading stakeholders from across ideologies, disciplines, and sectors of the system—policymakers, practitioners, researchers, technologists, and advocates—for two days of discussion and the examination of three use cases. The event was hosted by the Stanford Criminal Justice Center at the Stanford University School of Law.

Participants

Hassan Aden
Founder, The Aden Group

James Anderson
Director of Justice Policy, RAND Corporation

Kirk Arthur
Worldwide Government Solutions Lead, Microsoft

Chiraag Bains
Senior Fellow, Democracy Fund, Brookings Institution

Veronica Ballard Cunningham
Executive Director, American Probation and Parole Association

Richard Berk
Emeritus Distinguished Professor of Statistics, UCLA; Emeritus Professor of Criminology and Statistics, University of Pennsylvania

Michael Buenger
Executive Vice President and COO, National Center for State Courts

Pamela M. Casey
Vice President, Research and Design, National Center for State Courts

Ed Chung
Vice President of Initiatives, Vera Institute of Justice

Brandon del Pozo
Assistant Professor of Medicine, The Warren Alpert Medical School of Brown University

Joshua Essex
Co-Founder and Chief Technology Officer, Recidiviz

Steven Harpe
Executive Director, Oklahoma Department of Corrections

John Hollywood
Senior Operations Researcher, RAND Corporation

Yasser Ibrahim
Senior Vice President of Research and Development, Axon

Max Isaacs

Senior Staff Attorney, NYU Policing Project

Katie Kinsey
Chief of Staff, NYU Policing Project

Nancy La Vigne
Director, National Institute of Justice

Glenn E. Martin
Founder and President, GEMtrainers, LLC

Carlos J. Martinez
Public Defender, 11th Judicial Circuit of Florida

Gabriele Mazzini
Team Leader, AI Act, European Commission

Ben Packer
AI Lead and Data Scientist, Recidiviz

Rory Pulvino
Director of Analytics, Justice Innovation Lab

Meilani Santillán
Program Director, Code for America

Cornelia Sigworth
Senior Director, Axon

Radhika Singh
Vice President, Civil Legal Services and Strategic Policy Initiatives, National Legal Aid & Defender Association

Danyelle Solomon
Senior Director of U.S. Government Affairs, Microsoft

Andre Stancil
Executive Director, Colorado Department of Corrections

Rebecca Wexler

Faculty Co-Director, Berkeley Center for Law and Technology; Assistant Professor of Law, UC Berkeley School of Law

Dane Worthington
Director of Mobility, Center for Employment Opportunities

Jonathan Wroblewski
Lecturer, Harvard Law School

The key goal of the convening was to jump-start a national conversation about how to integrate AI into criminal justice in ways that promote justice, efficiency, and effectiveness and avoid exacerbating existing problems or creating new ones. This report summarizes key themes from the convening.

Seven Areas of Inquiry

CCJ framed the convening around seven key areas of inquiry:

- **Effectiveness and Efficiency**: How can AI enhance justice system operations while protecting rights? AI offers potential to improve efficiency and accuracy, but raises concerns about over-reliance and bias.
- **User Training**: What knowledge, skills, and other preparation do practitioners need to use AI effectively and ethically? Proper training is crucial to maximize benefits and avoid misuse.
- **Benchmarks and Performance Standards**: How should the performance of AI tools be evaluated? Establishing reliable standards is a critical and complex task given the difficulty of measuring AI performance.
- **Data Quality and Algorithmic Reliability**: How can we ensure data integrity and algorithmic reliability in AI systems, especially given concerns around the accuracy, implicit bias, and comprehensiveness of criminal justice data?
- **Privacy, Bias, and Fairness**: How can AI be used to reduce rather than amplify discrimination? AI has potential to identify and mitigate biases, but also risks reinforcing or creating new discriminatory practices.
- **Transparency, Explainability, and Accountability**: How can AI systems be made

more transparent and accountable? Balancing the need for explainable AI with proprietary interests and technical complexity is a key challenge.

- **Governance and Enforcement**: What regulatory frameworks and oversight mechanisms are needed? Establishing robust governance structures is essential to ensure ethical use of AI in criminal justice.

Balancing Goals

During the convening, participants placed a consistent emphasis on striking the delicate balance between harnessing the potential benefits of AI to enhance public safety and operational efficiency and the imperative to protect individual rights, ensure procedural fairness, and address ethical and safety concerns.

Across the convening, participants highlighted several promising upsides to AI integration. Examples include:

- identifying and mitigating human errors in decision-making processes, leading to more consistent and just outcomes across cases;
- enhancing system transparency by providing clear audit trails and facilitating more comprehensive oversight of decision-making processes;
- improving public trust and legitimacy by increasing transparency, access to information, and reliability; and
- helping agencies allocate resources more efficiently, potentially reducing response times in emergencies and focusing preventative efforts where they're most needed.

"In terms of AI … the thing I'm most excited about is that [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 … if we can get all states doing this to some level where you modernize your system, put AI technology on top of it to start teaching you about your data sets, about your population, about the problems they solve, and also your employees. I mean, there's all sorts of things we could be addressing."

STEVEN HARPE
Executive Director, Oklahoma Department of Corrections

Participants also discussed risks from AI integration, such as:

- the perpetuation or amplification of existing problems and unequal access to justice (such as imbalanced availability of tools between prosecution and defense, or between well-funded agencies and those with modest budgets);
- the potential for AI systems to be manipulated or misused;
- undermining due process when AI systems influence decisions that involve liberty interests, such as those related to arrest, bail, sentencing, and parole;
- degrading public trust and legitimacy by reducing transparency, adding complexity loops, and diminishing the public's ability to understand and critique decisions; and
- serving as a false solution, where AI implementation creates an illusion of progress while actually reinforcing or exacerbating various existing systemic problems.

Key Themes and Takeaways

The discussion produced four key themes and takeaways that could help guide future work: (1) Values- and Goals-Driven Adoption; (2) Critical Engagement with and Understanding of AI Tools; (3) Governance, Guardrails, and Regulatory Frameworks; and (4) Stakeholder Involvement.

Values- and Goals-Driven Adoption

Criminal justice agencies should articulate clear, values-driven goals when considering the adoption of AI tools. These could include increasing efficiency by automating routine tasks to free up staff time for higher-value work or advancing equity by using AI to identify and mitigate human biases or errors in decision-making. Agencies should be explicit about the intended purposes upfront and assess any potential AI application against them.

"All social science theories are informed by our values … The point is… when we're looking at how to bring AI into criminal justice, we have to foreground those values, and all our theories should be evaluated accordingly. That has to be an explicit … way of framing the process."

BRANDON DEL POZO

Assistant Professor of Medicine and Health Services, Policy, and Practice at Brown University's Warren Alpert Medical School; retired police chief

Participants discussed key actions to ensure that AI adoption is driven by intentional goals and values, and the prioritization of those actions:

- explicitly state intended purposes upfront and assess AI applications against those objectives;
- ground AI adoption in core democratic values such as transparency, accountability, non-discrimination, privacy, and equal rights under the law;
- prioritize AI tools whose inner workings can be examined and validated, with protocols for democratic oversight;
- study and design effective human-machine teaming processes to ensure that human oversight is meaningful and reliable;
- proactively communicate with stakeholders about how AI is being used; and
- establish mechanisms for redress if errors or harms occur.

Critical Engagement With and Understanding of AI Tools

Criminal justice agencies should proactively engage with AI systems (and the private companies that are producing AI tools and technologies) to shape their development and implementation in ways that align with public safety goals and ethical considerations. This engagement requires a multifaceted approach that encompasses a solid understanding of AI fundamentals, proactive strategies, and rigorous evaluation of AI capabilities and weaknesses.

"If you are moving to the implementation phase for AI tools, have we provided the documentation that the folks are going to need? Have we provided the education they're going to need? Have we provided the money for the staff?"

RICHARD BERK
Emeritus Professor of Criminology and Statistics, University of Pennsylvania

Understanding AI Fundamentals

- Providing the necessary training and preparation for the thousands of criminal justice leaders and independent agencies to develop a solid grasp of AI fundamentals to make informed decisions about adopting and governing these tools is a key goal and significant challenge.
- Understanding AI fundamentals is crucial for implementing and communicating clearly with stakeholders about AI use, validation, and oversight. This knowledge base should include:
    - recognition of the strengths and constraints of AI systems;
    - awareness of potential harms if AI systems are not well-designed or deployed;
    - appreciation that AI models learn patterns from historical data, and that these data can contain biases and have limited reliability;
    - an understanding that AI performance can degrade when applied to new contexts; and
    - acknowledgment of the “black box” nature of many current AI models.

Proactive Engagement Strategies

- Criminal justice agencies should recognize that AI systems will almost certainly become more sophisticated and common over time. Rather than adopting a reactive posture, agencies should proactively engage with AI technologies. Strategies for such engagement could include:
    - horizon scanning: regularly assessing the current state of AI technology and its potential applications;
    - scenario planning: envisioning various ways to deploy AI;
    - policy development: creating standards and regulations for AI adoption;
    - interdisciplinary collaboration: fostering partnerships between criminal justice professionals, AI researchers, ethicists, and community stakeholders;
    - pilot programs: implementing small-scale, monitored AI initiatives to gain practical experience; and

    - knowledge sharing: establishing networks for agencies to exchange lessons learned, best practices, and rigorous research.

Assessing Value and Trustworthiness

- Rigorous, independent evaluation is critical to assess the impacts of AI tools in real-world settings. Criminal justice agencies should advocate for, execute, and/or learn from empirical studies that evaluate AI capabilities and real-world performance. These evaluations should:
    - compare metrics such as crime rates, response times, and community satisfaction for AI-optimized methods versus conventional methods;
    - holistically examine impacts on issues of concern, from due process protections to quality of service, racial equity, and public trust;
    - periodically be repeated to assess how effectiveness may change over time; and
    - establish thresholds and protocols for halting or amending AI use based on evaluation results.

Governance, Guardrails, and Regulatory Frameworks

Participants discussed the importance of establishing the proper regulatory guardrails and governance frameworks to help ensure that the integration of AI goes as well as possible and that the right balance is struck between maintaining appropriate oversight and encouraging innovation.

Participants discussed lessons from the development of the European Union AI Act and its risk-based approach to regulating AI systems. Core elements for criminal justice AI governance could include:

- conducting impact assessments for proposed AI applications;
- applying more stringent controls to higher-risk use cases;
- ensuring transparency at multiple levels (development, procurement, use);

- preserving meaningful human oversight;
- building in ongoing monitoring and evaluation;
- establishing protocols for investigating and resolving errors or adverse impacts; and
- creating AI oversight functions that include broad and diverse representation by people and communities affected by AI outputs.

"There's a debate around whether we should regulate AI foundation models or just their applications. A number of companies and some academics have taken the position that we should regulate only the tools through which AI is used and deployed, not the models underlying them. Others think it makes sense to go upstream, because that's where we can have greater impact on the technology that the applications rely on.

Other questions include: Should we regulate AI in criminal justice broadly or by use case? Should we regulate high-risk cases as a class — including law enforcement uses — or should we have bills or rules tailored to particular applications like facial recognition technology or predictive policing? Should we set up an external auditing regime, or should we rely on self-policing and self-reporting by companies? Do we need any new liability regimes in order to make AI deployment safe and fair?"

CHIRAAG BAINS
Senior Fellow, Brookings Institution

Other conversations throughout the convening focused on the proper regulatory approaches and critical considerations for oversight, including:

Balancing Innovation and Oversight

- How do we ensure that the promise of AI systems can be accomplished and avoid strangling potential upsides, while safeguarding against present and potential harms?

Areas of Focus for Regulation

- **Model Development:** Should regulations target the creation and training of AI

models?

- **Tech Deployment:** Is it more effective to regulate the specific applications and implementations of AI, or should regulation encompass broader sectors and settings?
- **Risk Levels:** Should regulation be tiered based on the potential harm or impact of AI systems?
- **Civil and Human Rights:** How can regulations be designed to protect fundamental rights across all AI applications?

Centralization vs. Decentralization

- **Centralized Approach:** Would a single, overarching regulatory body be most effective, should individual criminal justice agencies or municipalities develop their own AI governance frameworks, or would a hybrid approach work best?
- **Cross-Jurisdictional Considerations:** How can regulations be harmonized across different jurisdictions to ensure consistency and, where appropriate, interoperability?

Enforcement and Accountability

- What penalties or consequences should be established for non-compliance with AI regulations?
- How can oversight bodies be empowered to effectively monitor and enforce AI governance?

Historial Analogues for Regulation

- How can agencies and policymakers making AI regulatory decisions learn from existing regulatory frameworks for governing research, therapeutics, and dual-use technologies, such as the Belmont Principles and the FDA's risk-based approval processes?

Stakeholder Involvement

Cultivating public trust in criminal justice AI will require proactive, multi-directional communication with all relevant stakeholders, including criminal justice system staff and administrators, crime victims and survivors, incarcerated and formerly incarcerated individuals and communities, technologists, and the general public.

"We – both at NIJ and DOJ writ large – have been doing a lot of convenings with AI developers, with technologists, with civil rights advocates. NIJ also has a seat at DOJ's newly established emerging technology board and so far it's been all AI all the time… We're doing a lot to listen and learn."

NANCY LA VIGNE
Director, National Institute of Justice

Core points of this discussion included:

- **Consultation and Input**
  - conduct listening sessions and surveys
  - hold public forums for feedback on:
    - rationale for AI adoption
    - potential risks
    - metrics for assessing effectiveness
  - include all stakeholders in helping define:
    - acceptable use policies
    - governance protocols
- **Building Public Trust and Legitimacy**
  - develop plain-language explanatory materials
  - conduct sustained, deep, multi-stakeholder engagement
  - hold public forums for discussion and feedback

    - give stakeholders a voice in policy and governance
    - provide hands-on training
- **Personnel Training**
    - communicate goals of AI adoption
    - provide hands-on training
    - actively engage frontline workers in implementation to:
        - tap domain knowledge
        - identify opportunities
        - troubleshoot issues and pain points
        - maximize feasibility and acceptability

Stakeholders won't automatically accept the use of AI or perceive it as legitimate. With that in mind, criminal justice leaders should make affirmative efforts to sincerely listen to and address people's concerns and create avenues for dialogue to oversee and guide the implementation of new technology tools.

Conclusion

The convening highlighted both the potential of AI to improve criminal justice outcomes and the complex challenges that must be proactively addressed. Participants expressed hope that AI tools could help increase the accuracy, consistency, and efficiency of tasks like service delivery and resource allocation. At the same time, they emphasized the importance of validating AI performance, monitoring for disparate impacts, and preserving human judgment in high-stakes decisions.

Ongoing multi-stakeholder engagement and collaboration will be essential as the field continues to evolve. With AI technology already in use in large and small ways across the criminal justice field, system leaders, researchers, technology developers, civil society organizations, and community representatives should engage in deliberate and collaborative efforts to chart a path forward for responsible expansion and innovation. Regular dialogue and information-sharing can help flag emerging opportunities, identify and mitigate risks preemptively, and build shared standards and best practices. Responsibly managed, AI has

the potential to enhance both safety and justice.

Acknowledgements

This event was organized with the support of hosts Debbie Mukamal, David Sklansky, and Robert Weisberg of Stanford Law School. In addition, this convening was made possible by funding from The Just Trust, Microsoft, Open Philanthropy, and CCJ's general operating contributors.