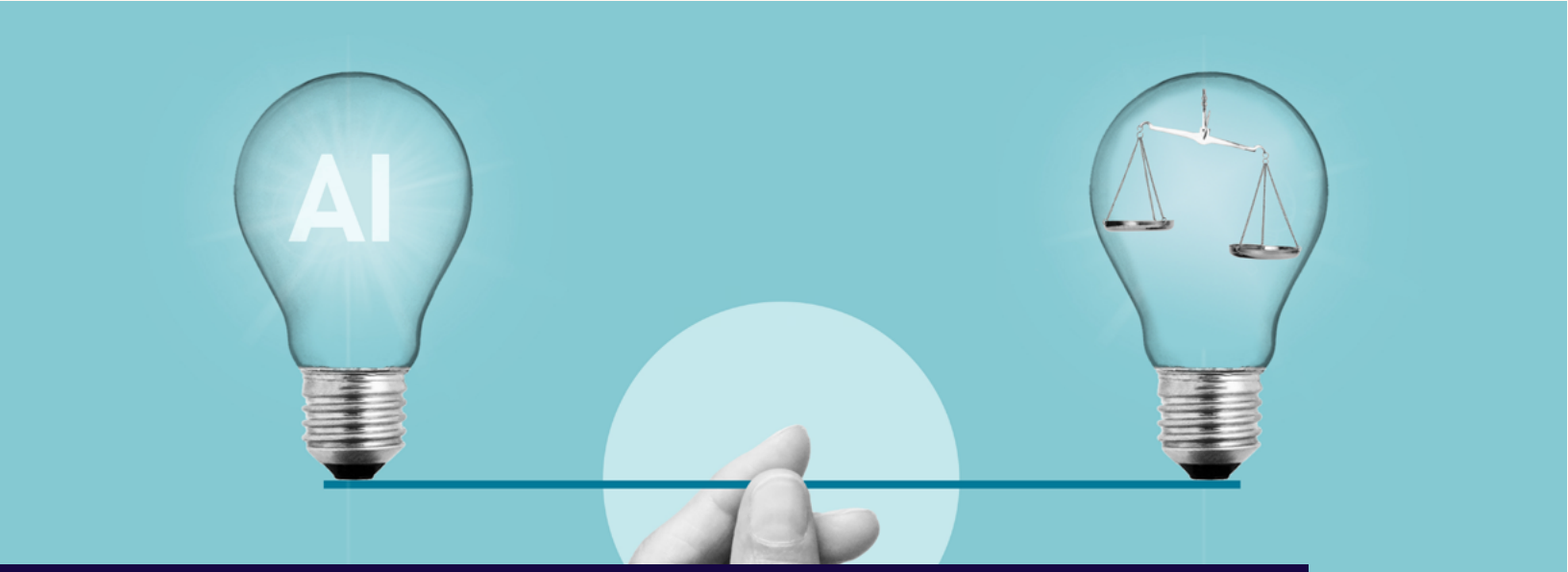

Accountability Principles for Adopting Artificial Intelligence Within the Criminal Justice and Immigration Systems

The development of widely available artificial intelligence[1] (AI) models has presented businesses, organizations, and governments with the opportunity to employ tools with the unprecedented capability to assist learning, analysis, and decision-making. At the same time, these tools pose novel risks and threats, in addition to requiring extensive resources to develop and maintain. As AI becomes increasingly used in the criminal justice and immigration systems, the Vera Institute of Justice (Vera) is committed to promoting the accountable use of AI: systems that produce fair, just, and equitable outcomes that help end the criminalization and mass incarceration of people of color, immigrants, and people experiencing poverty.

To achieve this, Vera recommends the following accountability principles to serve as a guiding framework and overarching lens for decision-making related to AI. These principles are designed to apply Vera's mission and values in this space, offer a tool for critical thinking and ethical consideration, and serve as parameters to foster responsible AI

adoption in the systems we work to change. Vera intends to revisit and revise these principles as technology and adoption practices evolve.

1. Only engage AI systems that shrink the footprint of mass incarceration and reduce criminalization.

AI used in the criminal justice and immigration systems context must have an explicit purpose to address the harms these systems create and minimize unnecessary interactions with them. AI that either intentionally or unintentionally expands the reach of the criminal justice or immigration systems, including AI that increases public surveillance or law enforcement deployment, should be avoided.

2. Define clear objectives before adopting AI systems, and explore whether those objectives can be achieved with fewer risks by using alternative tools.

AI systems used by actors in the criminal justice and immigration systems pose novel and unique risks that must be weighed against the potential benefits of their use. Risks may include biased or incorrect outcomes, user errors, lack of transparency in how the outcome was generated, and system failure. Before adopting any AI system, operators should define the specific purpose for the application, document resources required, and conduct an internal assessment to determine whether the need requires an AI solution or if there could be an alternative option. To further minimize risks, those adopting AI should determine whether the current infrastructure can support an AI system, including ensuring data quality, user training resources, and long-term maintenance capabilities.

3. Ensure that all uses of AI include rigorous human oversight, standards, and checks.

AI systems must have a corresponding entity that is responsible and accountable for their use and outputs. That entity must maintain appropriately rigorous standards for the AI system and be responsible for its oversight, including regular checks for errors and biases, compliance with recognized data ethics practices, and feedback loops for continuous improvement. AI system operators should be trained to use the AI system, interpret the outputs, and understand the limitations of the system; identify and mitigate the risks of harm from privacy breaches, biases, and inaccuracies; and stop or alter the system to address feedback and fix errors.

4. Center community-defined priorities for public safety as the basis for any AI adoption.

When used, AI should always serve human needs. In the criminal justice and immigration contexts, AI adoption must center the people and communities most impacted by these systems. Communities should be included in the decision-making process to ensure that resources are directed toward safety and justice, the needs of community members are met, and there are pathways to elevate risks and concerns.

5. Disclose the use of AI and related data protection measures to the public.

The disclosure of the use of AI and related data regulations enables the public, impacted communities, community organizations, and academics to monitor AI functions, contest system output, and address bias or error. This disclosure must include when and how an AI system is used and for what; how the system is being audited/reviewed; and what data and modeling techniques are used. System operators must provide this information so that it is comprehensible and accessible to both technical and non-technical audiences and include a point of contact for the public to submit feedback on the outputs.

The Vera Institute of Justice is powered by hundreds of advocates, researchers, and policy experts working to transform the criminal justice and immigration systems until they're fair for all. Vera's headquarters is in Brooklyn, New York, with offices in Washington, DC, New Orleans, and Los Angeles.

Endnotes

1 For this document, we are using the following definition for artificial intelligence: "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments." White House Executive Order 14110, "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," 88 Fed.Reg. 75191, October 30, 2023, https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

Cover image credit: Shutterstock