AI in Corrections: The Basics and a Way to Experiment[1]

Matthew G. Rowland[2]
Amit Shah[3]
Ashit Chandra[4]

THERE IS AN ongoing debate concerning the appropriate role of Artificial Intelligence (AI). This is particularly so in politically charged domains such as criminal justice, where the balance between societal and individual interests is already inherently contentious. As the use of AI moves beyond routine activities and closer to substantive decision-making, such as influencing sentences imposed for criminal violations and how those sentences are enforced, the debate becomes more intense. Often missing in the debate, however, is the perspective of criminal justice officials, including those in community corrections. This omission could be costly, because such officials are uniquely situated to recognize where the technology can enhance operational effectiveness and where it can pose risks to desired outcomes.

Criminal justice officials need to become more familiar with the technology to contribute meaningfully to the debate and its resolution. This can be best achieved through pilots and experimentation specifically designed to surface the technology's strengths, weaknesses, and costs. Without such firsthand experience, criminal justice officials can get lost in the technojargon, hyperbole, and IT industry's self-interest that often hinder a true understanding of AI.

This article aims to provide a basic understanding of AI and the surrounding issues to officials in the criminal justice field, who may not already have this understanding. Also, we advocate for a specific use case for criminal justice agencies to begin their own AI journey. That use case was selected because of its ability to expose the potential and challenges of AI in a relatively safe—yet important—environment. In addition, a successful prototype exists that leverages scalable and economical AI tools, with a criminal justice-oriented thought process already applied. The prototype provides a solid foundation on which to build while reducing time and financial burdens on agencies.

The proposed use case focuses on staff training, which has been identified as a top-tier need of criminal justice agencies (Russo, 2019). At the same time, a training application offers a more controllable environment and involves less sensitive data than would a direct case management use, and allows for extensive human supervision of the AI outputs.

[1] This article is the product of human thought and articulation. Generative AI, specifically OpenAI (2021), was used for the quotes in highlighted text boxes. Also, please note Generative AI was provided with a draft of this article and prompted with the question: "Will readers of the Federal Probation Journal find this article interesting." The AI responded: "[. . .] The article covers a critical and emerging topic with a focus on practical application, making it valuable and engaging for readers interested in the intersection of AI and corrections." We'll leave the accuracy of that AI assessment to you, the human reader.

[2] Matthew G. Rowland has more than 35 years of leadership and operational experience in the criminal justice field. He retired from government service as a senior executive and chief of the Probation and Pretrial Services Office for the Administrative Office of the United States Courts. He has since been a principal in Maloney, Rowland, & Associates LLC, providing diverse consulting services to government agencies, think tanks, and global technology firms focused on improving criminal justice outcomes.

[3] Dr. Amit Shah earned his doctorate in Biomedical Engineering at the University of Illinois at Chicago. He led clinical studies on the application of virtual reality and adaptive feedback algorithms on neurorehabilitation of stroke and traumatic brain injury survivors. After earning his Ph.D., Amit has served as a leader of data science initiatives at several companies and has led and won hackathons in predictive analytics. He is an expert on building generative AI solutions, predictive analytics, and What-If modeling. Amit was the Data Science Manager at Abbott Laboratories before leaving to found his own Data and AI consultancy, GNS-AI. He now serves as the President and Chief Data Officer of GNS-AI and is working with various companies and government agencies to deliver safe, cost-effective, and innovative solutions.

[4] Ashit Chandra is Vice President of Infobahn Solutions, specializing in Data Engineering, Data Analytics, and Enterprise Data Management. He has a successful track record designing and implementing Enterprise Data Lakes, Enterprise Data Warehouses, Business intelligence, and Big Data Analytics solutions that empower organizations to harness their data assets and meet mission goals.

An Introduction to AI

AI does not lend itself to easy definition. As noted in a journal from one of the United States' leading technical institutions, "artificial intelligence is constantly evolving, and the term often gets mangled" (Hao, 2018). In an attempt to offer at least something to help readers comprehend the technology, the same article suggests viewing AI as mimicking human intelligence, an analogy frequently employed by others to explain AI as well (Frankenfield, 2023). Unfortunately, the analogy is flawed.

One reason is that human intelligence itself is notoriously hard to define and conceptualize (Weder, 2020). So, the comparison is of one riddle to another. Also, there are clearly fundamental differences in how and why AI and humans operate "cognitively." AI relies on a digital framework, using binary code (a series of zeros and ones) to operate. Further, the technology is directed exclusively by algorithms (programmed step-by-step instructions) that are encoded by humans. Science fiction accounts to the contrary: AI is devoid of consciousness and self-initiative. It does nothing that cannot ultimately be traced back to its human-designed software and hardware. It is an inanimate tool, and its value and impact, whether positive or negative, hinges on the individuals who develop and wield it.

Conversely, humans function within a sensory framework, drawing upon our observations, auditory input, and tactile sensations, all interwoven with our wealth of experience, emotions, and intuition. Human agency and choice are inherent in our decision-making processes. Our collective choice to collaborate, complemented by shared creativity and innovation, gives rise to technologies like AI. It is so sophisticated that it is likely impossible for one person alone to understand it fully. But a full understanding of AI is not what is needed by criminal justice officials.

There are countless examples where we, as individuals, use technology effectively without an in-depth understanding of how it works. Most people would be at a loss to explain the inner workings of their car, cell phones, or personal computers, yet they skillfully use the devices with awareness of what constitutes legal and proper use.

Similarly, when it comes to AI, a conceptual grasp and recognition of appropriate and inappropriate uses are well within the realm of common understanding. While the development and proper use of this technology may necessitate a collaborative effort involving technical, operational, and administrative experts, individuals can ultimately assess the value and correct operation of AI.

What makes AI particularly exciting and disconcerting at the same time is the rate at which it is improving. The fundamental building blocks of AI, including computational speed, the amount of digital data, and sophisticated algorithms, are increasing exponentially (Henshall, 2023)—as are the number and expertise of developers to harness those growing resources. Combining that progress with increased investments in things like robotics conjures images of a dystopian future where computers dominate society, rendering human involvement unnecessary.

However, it is vital to differentiate between the speculative future of AI and its current state and near-term trajectory. The remote and imaginable should not eclipse the present and tangible. As discussed further below in relation to the proposed use case, AI can more than quickly process large stores of digitized text. Through a functionality called Natural Language Programming, it can understand and use human language. Additionally, image analysis functionality can identify objects, faces, scenes, and anomalies in digital photos and videos. It can display the results of its analysis in the manner and context that individual users find most valuable. And it can do all that when designed and resourced correctly in close to real-time, and with a precision and discipline that humans would find hard to match. Think how efficient and effective probation officers and other criminal justice officials would be with AI assistants to take in, process, and display information when and how officials need it. [5]

That alone warrants agencies undertaking greater testing of, and experimentation with, the technology. Furthermore, understanding AI today will better inform us about its future potential and risks. And if those arguments were not enough, then there is the reality (which the field of corrections is a testament to) that others will experiment and use the technology regardless of ethical considerations because of its lucrative potential. If AI is only available to such people, the dystopian future is all but assured.

Nevertheless, there are significant concerns surrounding AI even in its current state, particularly within the criminal justice domain. These concerns include: (1) AI's reliance on historical data and input from existing criminal justice personnel, potentially perpetuating biases, inequities, and inefficiencies attributed to the current system, and (2) the risk of undue deference to technology by future personnel within the criminal justice system (Burns, 2022).

The fear, somewhat paradoxically, is that AI will be overly influenced by poor and bad actors from the past. Conversely, future actors in the criminal justice system will subjugate their own good judgment and overly rely on AI. Fortunately, good technology implementation practices in relation to AI can mitigate those concerns. These practices include transparency and diligent human examination and oversight of AI inputs, outputs, and uses. In addition, the technology itself offers ways to combat faulty inputs and distorted outputs, such as AI-based tests and techniques that can proactively expose potential biases in data analyzed and conclusions drawn (Feast, 2020).

So, well-designed AI applications can actually reduce bias, offering "a number of advantages [over human judgment alone], including the speed at which they process information. Also, because they do not have feelings, they are more objective and predictable than people in their decision-making. They are a core component of overcoming the pervasive bias and discrimination that exists in the criminal justice system" (Rizer & Watney, 2018).

AI Already in Use

It's highly probable that you've interacted with AI today without even realizing it. Whether you used GPS to navigate around traffic, enjoyed personalized music, or engaged with social media, AI played a pivotal role in those experiences. If you're reading this article online, your device, network, and the search engine likely leveraged AI.

While not mainstream, some criminal justice agencies have started using AI in their operations. A consortium organized by the National Institute of Justice found that AI is currently employed to screen prison visitors and incoming mail for contraband. It's also used for analyzing inmate telephone conversations to identify threats and potential

[5] A practical example is that the national and local policies and procedures of the Federal Probation and Pretrial Services System are collectively thousands of pages long. It is impractical to expect individual probation and pretrial staff members to memorize and adhere to so many policy and procedural provisions without support. There are several ways AI can help the officers navigate to the relevant policy guidance when, where, and how they need it.

criminal activity. Additionally, some agencies use chatbots[6] to remind pretrial defendants of court dates and to provide probationers with relevant information to help them comply with their supervision conditions. AI is also sometimes employed for actuarial prediction of recidivist risk presented by inmates and individuals under community supervision.

Although reports on the effectiveness of AI in those instances is not yet publicly available, the consortium concluded that "AI is here to stay" and that "AI-enabled tools have the potential to improve efficiency, reduce costs, and expand capabilities across many criminal justice use cases[.]" To achieve those benefits, the consortium added, "will require intentional investment, careful consideration, and sustained efforts from criminal justice decision makers" (Criminal Justice Testing and Evaluation Consortium, 2020).

A similar conclusion was reached by officials from the Administrative Office of the United States Courts (AO). The AO is a federal judicial agency responsible for overseeing, supporting, and reporting on the Federal Probation and Pretrial Services System (FPPS).

Both the AO and FPPS face the challenge of processing a vast amount of information related to court-involved individuals and the strategies and activities probation and pretrial services officers use in investigating and supervising those individuals.

For probation and pretrial services officers, the task involves sifting through the mass of information provided by clients themselves, the community, and other agencies to identify what is relevant and actionable for effective case management.

For AO administrators, the challenge is identifying systemic patterns and best practices from literally tons of data (if it were printed out) that officers enter into case management systems regarding their clients and the strategies employed to achieve positive case outcomes.

To determine if AI could help meet the challenge of efficiently managing the vast amounts of data to answer operationally important questions, the AO undertook a *proof of concept*. The effort is described in more detail in an *Irish Probation Journal* article (Rowland, Beatty-Gregoire, & Fitzgerald, 2019) but, in short, involved forming two teams, each with a handful of probation officer specialists and computer engineers.[7] Each team was also equipped with open-source[8] AI tools.

One team was provided with the case notes, known as chronological entries, typed by probation officers on 133,000 post-conviction supervision cases; the other group given scanned copies of 11,243 presentence reports. [9]

Each team was then tasked with using the AI tools to answer specific questions.

The team with the supervision case notes was asked to identify specific references in the notes that would justify concluding that the person supervised had ties to violent extremist groups. The team handling the presentence reports had to determine how many defendants, as described in the reports, were suffering from mental illness and the nature of their condition.

Both teams first conducted quality control checks on the documents they were given, standardized the data format to facilitate AI analysis, and developed algorithms to categorize relevant information hierarchically based on the posed questions. They also created output reports that allowed probation officers in the courts, who were familiar with the cases being analyzed, to verify the accuracy of the results. Importantly, the outputs allowed the officers in the courts to see the exact data upon which the AI relied to classify the case as involving persons with ties to violent extremist groups or suffering from mental illness.

Upon reviewing the output reports, the officers in the courts familiar with the cases provided feedback on the reports' accuracy and offered insights into why the results were correct or incorrect in each case. That input led to modifications to the algorithms, and the process was repeated, accuracy improving with each iteration to the point that the outputs were considered highly reliable based on the data analyzed.[10]

The ultimate finding from the proof of concept was that "at roughly 3% of the price of doing it manually and at a fraction of the time, the AI [. . .] revealed insights into violent extremists under supervision and the mental health condition of persons being sentenced in federal court." Further, it was concluded that "[AI] offers unprecedented opportunities to learn from past cases, to make [corrections] more efficient, and to further several public interests."

Consequently, the AO project evaluators recommended additional experimentation with AI—but with the caveat that the agency and those like it considering AI "invest in the front end to ensure business needs are clear and that the AI is properly 'educated' about the data it will be processing. Again, there is strong support for the 'supervised model' of AI with the technology and subject-matter experts working together, rather than independently."

A Proposed New Use Case

Deciding where to begin experimenting with AI can be daunting for any corrections agency. However, it is important to remember that lessons learned from adopting other technological tools can be applied to AI. This includes following generally accepted change-management principles, conducting legal and ethical reviews, developing cost-benefit measures, and not operationalizing anything that could affect actual case management without sufficient testing and vetting. Moreover, technology seldom operates perfectly out of the box and requires ongoing configuration and adjustments for optimization.

As mentioned previously, it is an established best practice to use cross-cutting teams when developing AI applications and to ensure ongoing human supervision of the AI. Such a cross-cutting team can also assist on the front end in defining project goals and thinking about how outcomes can assist in shaping an ultimate vision for AI, assuming the technology proves useful. To that end, to develop a new use case for this article, the first step taken was discussion with federal, state, and local corrections officials and their technology teams.

A vision for the technology that emerged

[6] Chatbots are AI applications that enable technology to engage with humans through speech or text, answering questions, directing queries, and furnishing necessary information.

[7] The computer engineers were a mix of judiciary employees and contract vendors, most with only recent exposure to the AI tools to be used in the project.

[8] Open source refers to software that is publicly available and free to use.

[9] To protect against inappropriate secondary use or disclosure of the case data collected as part of the proof of concept, only government-approved environments were used to store and analyze the data. In addition, judiciary data retention and disposal rules were applied, and all staff involved in the project were subject to confidentiality agreements and government security regulations.

[10] The subject-matter experts and reviewing officers noted that output reports were only accurate to the degree the data analyzed was up-to-date and complete. They noted such analysis is not a "launch and forget" endeavor but rather that "ongoing review of data dictionaries, expansion of data sources and a strong feedback loop with users are needed for the technology to achieve its full potential."

from those discussions was AI as a "digital assistant": not replacing humans but helping improve human decision-making. With AI memory capacity and speedy recall, AI could prove to be a valuable repository of policies, procedures, and best practices and a conveyer of institutional knowledge. With a well-designed interface, AI can give criminal justice officials the information they need, when, where, and how they need it.

As illustration, envision a case that transfers between probation officers. It may not be easy for the new officer to detect changes in the client's appearance or living conditions compared to what occurred before the transfer. In contrast, AI technology can easily compare digitized photos of the client and residence, taken before and after the transfer. The AI could also point out the differences in a relevant way based on what has been learned from other cases, such as distinguishing where weight loss may be a sign of improving health in the client as opposed to resumed drug use or a mental or physical medical problem.

In terms of detecting changes in the client's living conditions, one of the most notorious cases in community corrections history involved a person under supervision who transferred repeatedly among officers and agencies. The transfers contributed to new officers not detecting signs that the person under supervision had modified his house and property to conceal the presence of persons he had kidnapped and whom he repeatedly assaulted during the period of supervision. It is admittedly speculative but interesting to think that AI might have helped the officers detect the changes in the residence over time and led to quicker detection, or ideally deterred the client's criminal behavior.

The vision aside, those in corrections consulted for the identification of a new use case had questions regarding the ethical use of AI. The United States is only beginning to consider regulations pertaining to the technology, with a first-of-its-kind executive order being recently signed by President Biden (White House Briefing Room, 2023). The full impact of that order is not yet known, but it appears in substantive areas to be consistent with regulations a little further along promulgated by the European Union (EU). The EU approach establishes categories of risk based on how and by whom the AI is used, and sets out limitations and requirements commensurate with that risk level.

Under the EU system, correctional agencies' operational use of AI would likely be deemed "high risk." The regulations call for careful thought and documentation related to the data selection, algorithm development, and other inputs into the AI, as well as the technical workings of the technology design itself. The regulations also call for ongoing and rigorous testing and human supervision of the AI outputs and proactive steps to avoid any impermissible biases from influencing the AI and its outputs. An overall requirement under the EU regulations is transparency (European Union, n.d.). Consequently, corrections agencies should keep direct stakeholders informed and consider publishing papers in professional and academic journals about their AI use as well. This will have the added benefit of allowing corrections agencies to learn from each other regarding AI utility and best practices.

With those considerations in mind, the specific use case that we suggest corrections agencies explore to gain familiarity with AI relates to staff training—specifically, using AI to interpret audio recordings of mock interviews between staff and "clients." The training context offers a controlled environment, enabling limitations on sensitive information and identities and plenty of human supervision of the AI outputs. Another factor for the recommendation is that we have successfully developed a prototype that analyzed audio-recorded conversations between probation officers and anonymized or mock clients. For those of you interested in the technical specifications of the prototype, see the end notes.[i]

Beyond its ability to transcribe conversations and identify speakers, the prototype offers both descriptive and qualitative insights into the dialogue. Notably, the ChatGPT-style interface is a compelling feature for extracting this kind of information.

By simply inputting a question like *Who spoke more, Speaker One or Speaker Two?,* the AI promptly responds with the answer. It also allows for more in-depth inquiries, such as "Did any of the speakers use profanity?" or "Did one speaker talk over the other?" or "Did the speakers discuss the facts and circumstances surrounding the client's recent drug use?"

The possibilities are virtually limitless, with the only constraint being the need to develop training materials for the AI. It's worth noting that crafting these training materials demands careful consideration and testing, as their quality significantly impacts the AI's output. For instance, if you intend to assess whether officers in a recorded conversation are using specific techniques, like cognitive-behavioral or motivational interviewing methods, you must carefully define the words, phrases, and even the tone of voice associated with these techniques. Similarly, when gauging the "client's" response to the officers' use of these techniques, you'll need to specify the words, phrases, and tone that the AI should identify.

On the back end, effort was required to provide the AI with feedback regarding the accuracy of its determinations and the rationale behind its decisions. For instance, it was essential to ascertain whether the AI accurately classified the officer's conversational approach as "directive" or "instructional," when in fact, it was more "collaborative" and "emphatic." This feedback loop involved querying the AI through the interface to understand the data it relied on to reach its conclusions. Human supervisors of the application then had to evaluate whether new training material for the technology was necessary or if modifications to the application's algorithm were warranted.

The prototype illustrated the significant potential of audio recording analysis. Words, phrases, their arrangement, and the nuances of tone hold the key to correctly understanding and categorizing a conversation. However, if a jurisdiction so desires, the option of adding visual analysis is available. Given the substantial portion of communication that is non-verbal, supplementing verbal cues with facial expressions, body language, and other non-verbal signals can render the analysis even more comprehensive.

Capturing people's voices and images, however, creates risk. Although in our modern digital world many of our voices and images are floating somewhere in the public domain, what makes certain AI uses, like Deep Fakes,[11] disconcerting is that they can manipulate our voices and images to make it appear we have said and done things we have not. There are some defenses to that, but restricting the data made publicly available helps as well. Consequently, there are privacy considerations that should go into the development of agencies' AI environments. For a brief discussion of such considerations, please see the end notes following the Bibliography.[ii]

Bibliography

Burns, G. (2022, February 1). *The use and future*

[11] Deep Fakes refer to images created or manipulated using AI technologies for purposes of deceiving those that view it. See, https://www.forbes.com/sites/alexandralevine/2023/10/12/in-a-new-era-of-deepfakes-ai-makes-real-news-anchors-report-fake-stories/?sh=36960f2957af

of artificial intelligence monitoring in prisons. Retrieved from The Reasons Foundation: https://reason.org/commentary/the-use-and-future-of-artificial-intelligence-monitoring-in-prisons/

Criminal Justice Testing and Evaluation Consortium (2020, August). *Artificial intelligence applications in corrections.* Retrieved from National Institute of Justice, U.S. Department of Justice: https://cjtec.org/files/64bfb2359c420

European Union. (n.d.). *Shaping Europe's digital future.* Retrieved March 17, 2024, from European Commision: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai#:~:text=The%20AI%20Act%20allows%20the,EU%20fall%20into%20this%20category.

Feast, J. (2020, October). *Root out bias at every stage of your AI-development process.* Retrieved from Harvard Business Review: https://hbr.org/2020/10/root-out-bias-at-every-stage-of-your-ai-development-process

Frankenfield, J. (2023, April 24). *Artificial Intelligence: What it is and how it is used.* Retrieved from Investopedia: https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp

Hao, K. (2018, November 10). *What is AI? We drew you a flowchart to work it out.* Retrieved from MIT Technology Review: https://www.technologyreview.com/2018/11/10/139137/is-this-ai-we-drew-you-a-flowchart-to-work-it-out/

Henshall, W. (2023, August 2). *4 charts that show why AI progress is unlikely to slow down.* Retrieved from Time: https://time.com/6300942/ai-progress-charts/

Rizer, A., & Watney, C. (2018). Artificial intelligence can make our jail system more efficient, equitable, and just. *Texas Review of Law and Politics*, 181-227, 183.

Rowland, M. G., Beatty-Gregoire, N., & Fitzgerald, J. J. (2019, October). Testing artificial intelligence in the United States Probation and Pretrial Services System. *Irish Probation Journal*, 107-117, 116. Retrieved from https://www.probation.ie/EN/PB/0/D9C80060AC8ED236802584C100510922/$File/Testomg%20Artificial%20Intelligence%20in%20the%20United%20States%20Probation%20and%20Pre-Trial%20Services%20System.pdf

Russo, J. (2019, December 1). *Workforce issues in corrections.* Retrieved from National Institute of Justice: https://nij.ojp.gov/topics/articles/workforce-issues-corrections

Tim Fountaine, B. M. (2019, July). *Building the AI-Powered organization.* Retrieved from Harvard Business Review: https://hbr.org/2019/07/building-the-ai-powered-organization

Vynck, G. D. (2023, August 29). *AI images are getting harder to spot. Google thinks it has a solution.* Retrieved from *Washington Post*: https://www.washingtonpost.com/technology/2023/08/29/google-wants-watermark-ai-generated-images-stop-deepfakes/

Weder, A. (2020, October 5). *Q&A – What Is Intelligence?* Retrieved from News & Publications: John Hopkins School of Medicine: https://www.hopkinsmedicine.org/news/articles/2020/10/qa--what-is-intelligence

White House Briefing Room. (2023, October 30). *White House.* Retrieved from Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

End Notes

i. For the proof of concept centered around interview analysis via generative AI, we employed a carefully curated tech stack to maximize efficiency and performance. Streamlit was our choice for web app development due to its interactive and straightforward nature, making it ideal for rapid prototyping. We harnessed the power of OpenAI's GPT-3.5 Turbo model through their API for sophisticated language modeling, analysis, and summarization, ensuring in-depth insights from the interview data. We also leveraged OpenAI's Whisper model to create a transcript that GPT-3.5 turbo could process. The ChromaDB, an open-source vector database, facilitated efficient data management and indexing to allow for inferencing on the transcript, while LangChain was invaluable for creating a seamless chain of prompts, enhancing the user interaction and data input process, which resulted in a chatbot we could use to dynamically query the resulting transcript. All of this was adeptly put together using Visual Studio Code as the integrated development environment (IDE) for its versatility and extensive developer support.

ii.To protect sensitive audio data and transcripts, for example, a multi-layered on-premise security approach should be taken. The audio files should be transcribed using private voice-to-text models that are served locally and private large language models that are trained internally using the organization's data. This prevents exposing the raw audio to external cloud services. The audio and resulting transcripts should be encrypted and anonymized to remove identifiers. Any additional natural language processing or machine learning inferencing on the transcripts should occur locally on private edge servers, not in the cloud. Strong access controls and audit trails should track all data access, with logs monitored for unauthorized usage. The data should be stored on local servers with hardened security including firewalls, intrusion prevention, and minimal ports exposed. Regular pen testing should check for vulnerabilities. With proper encryption, private models, access controls, on-premise infrastructure, and auditing, the confidentiality of the audio data can be maintained from transcription through usage.