



Data Governance and AI in State Courts

Accurate and timely data are essential for managing cases, measuring court performance, and administering courts. The rapid increase in the use of AI and advanced data analytics increases the urgency and importance of having strong data governance.

Diane Robinson

Principal Court Research Associate, National Center for State Courts

Shay Cleary

Managing Director, Court Consulting Services, National Center for State Courts

Sarah Gibson

Data Scientist, National Center for State Courts

Andrea L. Miller

Senior Court Research Associate, National Center for State Courts

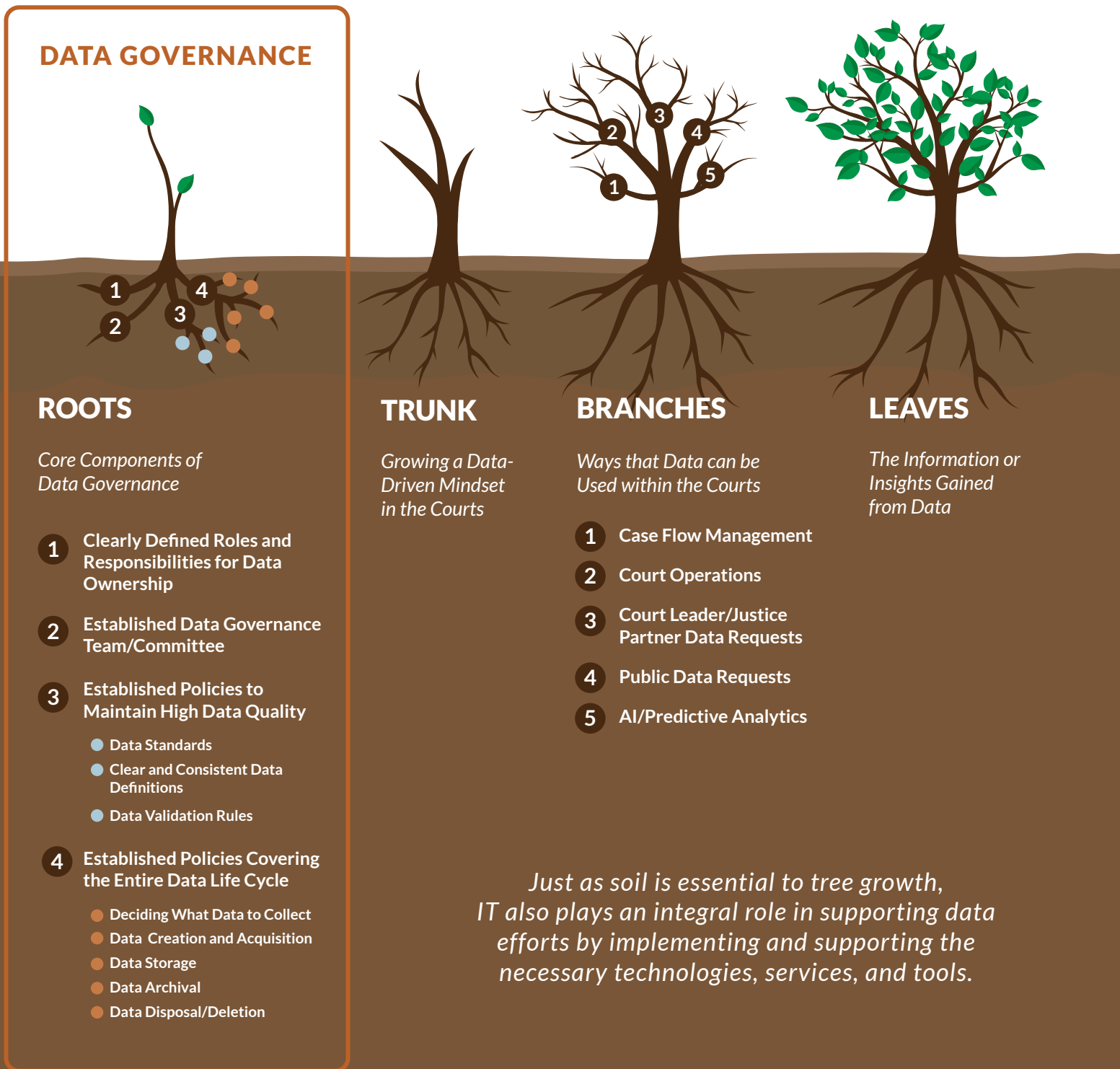
Data Governance

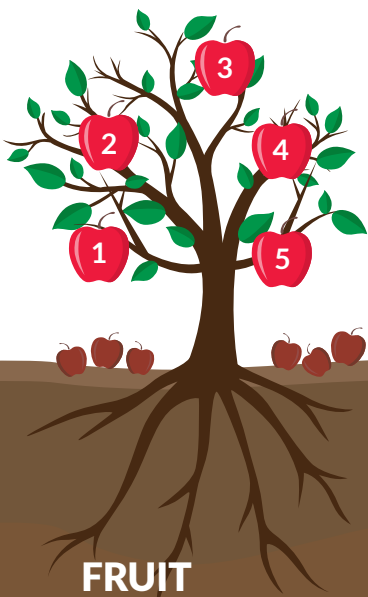
Data governance is a framework encompassing the people, policies, processes, and technology that ensure high-quality data, data management, and data security. Although strong data governance has always been an essential component of high court performance, the increasing use of artificial or augmented intelligence (AI) in court operations makes data governance even more critical. This article explores best practices for data governance and the intersection of data governance and AI. Data governance can be considered the “roots” of data use in the courts, including clearly defined roles and responsibilities, an established data governance team, documented policies to maintain consistent and high-quality data, and policies that cover the entire data life cycle (*see Figure 1, next page*).

Figure 1 The Roots of Data Use

THE FIRST STEP FOR GROWTH: DATA GOVERNANCE

Successful Data Use Starts with Strong Roots



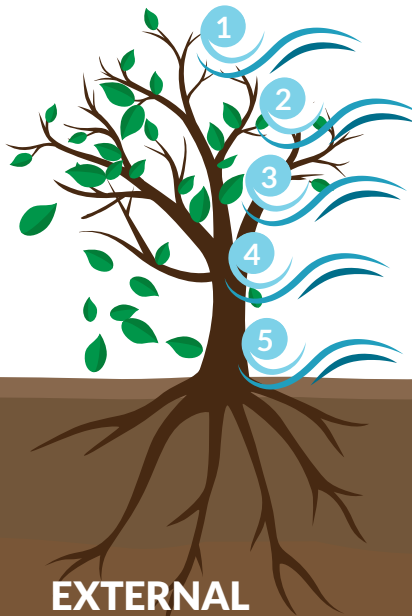


FRUIT

The Outcomes of Effective Data Use

- 1 Improved Case Processing
- 2 Effective Resource Management
- 3 Consistent and Timely Response to Data Requests
- 4 Increased Trust and Confidence in the Judiciary
- 5 Improved Case Outcomes

Note: Not all fruit will be viable. Some will fall to the ground but will break down and enrich the soil (i.e., the lessons learned will be invaluable).



EXTERNAL FORCES (WIND)

Strong Data Governance Roots Mitigate Damage and Decrease Impact from External Events

- 1 Inaccurate Narratives about Courts/Court Data
- 2 Data Breaches
- 3 Misuse of Court Data
- 4 Legislative or Rule Changes that Impact How Data are Shared/Handled
- 5 Advancements in Technology



PRUNING AND MAINTENANCE

Data Governance Needs to be Maintained and Revised

Prune dead branches to make room for new areas of data use that can also result in new root growth (aka data governance policies).

Water soil with lessons learned and feedback from court staff and impacted groups like court users, stakeholders, and community members to encourage continued growth.

Figure 2 Data Governance and AI

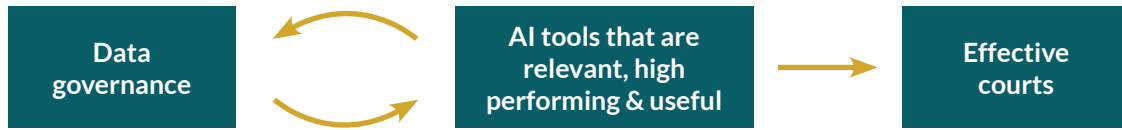
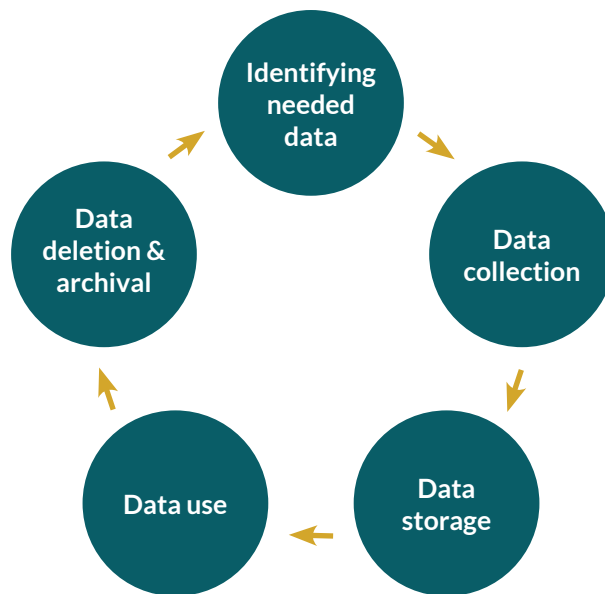


Figure 3 Life Cycle of Data



Source: Robinson and Gibson, 2019.

The relationship between court data governance and AI moves in both directions (see *Figure 2*). The quality of the courts' current data will determine what kinds of AI innovations are possible and how successfully new AI tools will be integrated into the courts. Conversely, the introduction of AI technologies can improve data governance while also creating new kinds of data and corresponding data-governance-related issues, considerations, and ethical concerns.

A hallmark of courts with strong data governance is that everyone working with court data understands the importance of data to the court and to court customers. This includes judicial officers, the court administrator, the newest employee in the clerk's office, and even the attorney e-filing a case. Data governance requires attention to data quality throughout the entire data life cycle (see *Figure 3*).

Data governance begins with identifying needed data. No court has unlimited time or resources to collect data and ensure data quality. “Nice to know” is not a sufficient justification for collecting data. This means looking with a critical eye at data collected to ensure that they are being used effectively. However, effective use of AI may make data collection more efficient, expanding the scope at little cost.

Data collection must be done in a way that maximizes the data’s quality. Case management systems and other court IT systems should be configured to maximize data quality. For example, entering a date should have validation to ensure the date is in the correct format and is plausible. Courts should also streamline codes to encourage selection of the most accurate choice. Data collection is an area with real opportunity for AI to improve the speed and accuracy of data entry from documents, while also being able to learn patterns and provide alerts and potentially automatically adjust incorrectly entered data.

Data storage must ensure that the necessary data are available to the right individuals at the right time and in the right amount. For example, having a production database separate from that used for reports or research helps manage traffic while keeping data accessible. Data-use policies address who has access to what information and what resources or information are used to check and improve data quality. Finally, data governance policies should be consistent with state law and court rules about how long data are kept. For courts working on clean-slate initiatives or other decriminalization steps, careful consideration is needed of how to protect individuals’ rights while maintaining accurate information about caseloads.

AI

Merriam-Webster defines artificial intelligence as “the capability of computer systems or algorithms to imitate intelligent human behavior.” AI includes a wide variety of technologies that are built using machine learning, such as natural language processing, facial recognition, and generative AI. *Merriam-Webster* defines generative AI as “artificial intelligence . . . that is capable of generating new content (such as images or text) in response to a submitted prompt (such as a query) by learning from a large reference database of examples.” The promise of more effective state courts through the use of AI is no longer a characteristic of a hypothetical future court—it is the work of state courts today.

A core component of effective courts has always been collecting and storing information and data. Although the mechanisms have evolved from physical papers and filing boxes to digital documents, databases, and cloud storage, information gathering has always been central to the functioning of courts. What *has* changed is both the widely accepted view of data as a strategic asset, instead of merely a byproduct of court operations, and the rapidly expanding capabilities of AI-based tools and technologies to enhance data-driven decision-making. As the pace of improvements in AI accelerates, effective and responsible use of data is critical. Data provide the foundation for the use of AI, and to successfully navigate and use these new AI-based tools, strong and effective data governance is an absolute necessity.

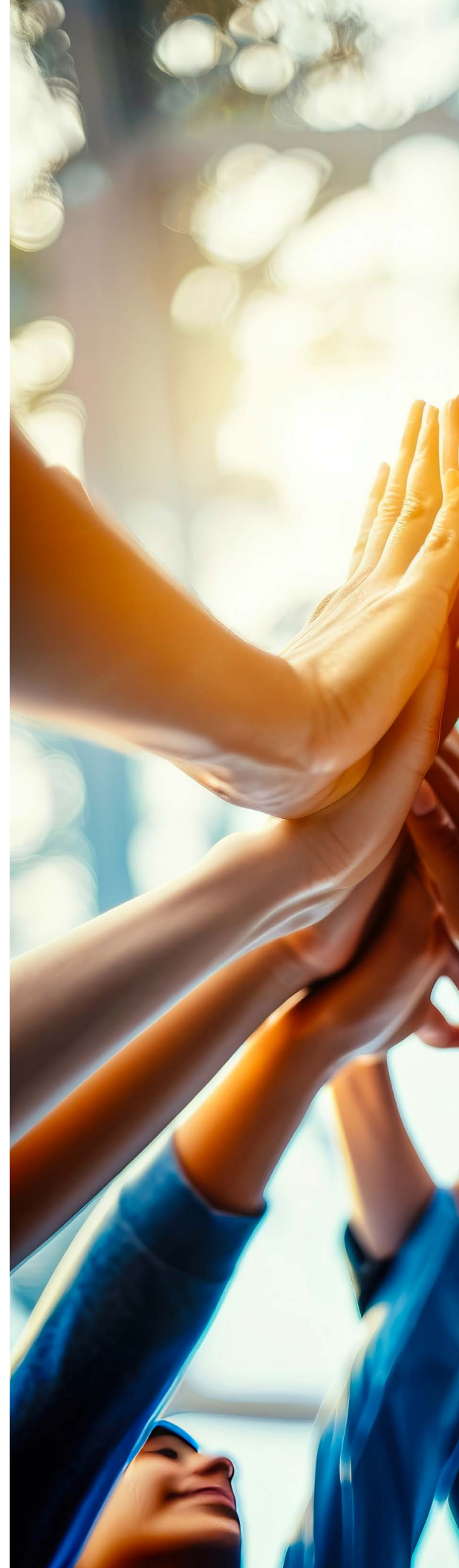
Early examples of AI applications in the state courts illustrate the potential breadth of roles that AI might play in this sector. Some courts, such as Mohave County Superior


Court, Maricopa County Superior Court, and Miami-Dade County Courts, are using bots to help answer court users' questions and assist court users with wayfinding.¹ Others, including Palm Beach County and Orange County, are using AI to increase their efficiency in processing new filings and managing caseflow (see Joint Technology Committee, 2024, and Reinkensmeyer and Billotte, 2019, for more examples of early AI applications).² As technologies improve over time, AI could be used for a wide variety of tasks involved in case processing and caseflow management, in interfacing with court users and helping them navigate court systems, and in substantive legal decision-making that shapes case outcomes. If AI is integrated thoughtfully and with care, these technologies could improve access to the courts, promote equitable justice outcomes, improve equity in the workforce, increase efficiency in case processing, and promote public trust and confidence in the courts.

Because AI technologies are built on data, data quality must be addressed before implementing AI. When inaccurate or biased data are used to design or train a new AI tool, the AI will perpetuate and magnify those inaccuracies and biases. Consider Amazon's recruiting tool that was created to screen résumés and identify top talent. On its surface, the tool seemed like an efficient use of AI technologies to help automate a time-consuming process of screening applicants. The tool was trained on data Amazon had collected from previously submitted job applications. What Amazon did not consider is the quality and diversity of those data and the inherent bias that was present. In the tech field, most job applicants and eventual hires have historically been men. The AI model noticed this pattern and penalized résumés that had indicators that the applicant was a woman. Now imagine a cyclical process of using

¹ See Mohave County at <https://perma.cc/2QEL-HZLT>; Maricopa County at <https://perma.cc/S7SW-8BUA>; and Miami-Dade County at <https://perma.cc/C5MP-UPU6>.

² See Palm Beach County at <https://perma.cc/83UG-SL4G> and Orange County at <https://perma.cc/H87S-E8DU>.





the output from the screening tool as additional data to train the model, and it becomes clear how the initial data-quality issues and bias are not only perpetuated, but made worse every time new data are supplied to the model (Dastin, 2018).

Examples like Amazon's recruiting tool provide a poignant reminder of the necessity for strong data governance to ensure that questions of bias and overall data quality are addressed. Because the potential impacts of AI on society are so great, the European Union and other governmental bodies have already implemented AI regulation ("EU AI Act: First Regulation on Artificial Intelligence," 2023). Having established definitions and policies in place at every point in the data life cycle ensures that there are accountability and methods for identifying potential risks and issues. Strong data governance creates transparency and accountability that will enhance trust among court customers, stakeholders, and the public.

To the extent that a court's data measure the right things and are accurate, complete, and compatible across jurisdictions, the court will be equipped to:

- Understand where AI technologies are most needed, will have the greatest benefits, and will do the least harm;
- Identify specific court operations and tasks to enhance with AI and which tasks to leave to humans;
- Develop better-performing AI tools;
- Measure the performance of AI tools to ensure they meet the courts' needs before launch and make adjustments and improvements as needed;
- Make sound, data-driven decisions about court policies and practices; and
- Share technology and knowledge across jurisdictions, coordinate with justice partners, and learn from other courts.

AI Can Support Strong Data Governance

AI has the potential to support data governance by improving data entry and data quality. By leveraging AI to automate and streamline data entry, AI can reduce human error, resulting in greater accuracy and reliability of data. AI can also improve overall data quality by using machine learning to analyze, categorize, and validate data against data governance policies and flag anomalies or potentially incorrect information for further review. The following are examples of how AI may improve data entry and data quality.

Contextual Understanding and Validation

AI does not just read text; it can compare it to its context. This means it can validate data as they are entered. For instance, if an AI system is processing an efiled or scanned judgment or order, it can cross-reference elements to check for consistency and accuracy.

Automated Data Categorization and Tagging

AI can automatically categorize and tag data as they are entered. For example, in an efiled or scanned document, AI can identify and categorize important elements such as case numbers, dollar amounts, party names, dates, and the type of document being filed, streamlining subsequent data entry. AI can even identify a case type based on the content of the text.

Real-Time Correction

When incorporated into a workflow, AI can provide real-time suggestions for error correction during data entry, thereby preventing errors from entering the system in the first place.

Data Imputation

AI can intelligently fill in missing values based on the patterns and relationships it discerns in the data. For example, it might identify missing values in a motion based on the known values of similar records (or even the same case) and alert the submitter/user to the missing value, make suggestions, or input what is missing.

Automated Data Governance

AI can enforce data governance policies automatically, ensuring data quality standards are consistently met across the organization. Furthermore, AI can continuously learn from new data and feedback and improve its ability to cleanse, correct, and augment data over time.

Scalability

AI can manage large volumes of data efficiently, a task that is not feasible for manual processes, especially while maintaining high data quality.

AI Will Complicate Data Governance

Although new AI technologies have the potential to improve data quality and strengthen data governance practices in the courts, AI may also complicate court data governance. AI will create new data and new types of data, and it will require additions or revisions to data governance policies. The following are a few examples of critical issues for courts to consider.

Court Leadership and Administration

AI will affect how courts engage in strategic planning, goal setting, performance evaluation, and data-driven decision-making. AI tools will create new data and help the courts draw new insights out of complex data. If AI is implemented well, AI-produced data might enable courts to continuously and seamlessly assess effectiveness and efficiency and to update and improve practices. However, the proliferation of data may also make it more challenging for court leaders to extract useful insights from what may feel like an ocean of data—in other words, separating the signal from the noise. Courts will need to increase their data literacy to confront these new complexities.

Court Workforces and Job Design

AI will change how court personnel interact with data. As AI technologies change how data are created, stored, protected, and used, the types of court roles involved in data governance may change. New court jobs may emerge as new types of skills are needed to develop, operate, and maintain AI tools. Existing court roles may expand to have a bigger role in data governance, requiring a higher level of data literacy. Courts' practices relating to job design, recruitment and hiring, and continuing education will need to evolve to meet these demands.

Security and Privacy

AI will affect data security and privacy concerns. As part of the creation, use, and maintenance of AI technologies, court data may be shared with more organizations and entities outside of the court system. For example, court data may be shared with technology vendors for building and maintaining AI technologies or shared with justice partners for collaborative case processing. While greater coordination with outside entities may enable the court to provide better and more efficient services, it will also increase the need for strong security practices that protect court users from inappropriate uses of their information.

Technology Vendors

AI will change the courts' relationships with private technology vendors. Courts will need to establish and monitor appropriate roles of private for-profit entities in court operations. They will need to make important decisions surrounding who will control algorithms, data, quality control, performance monitoring, and maintenance of AI technologies. Courts should especially be wary of "black box" AI tools, where the AI's algorithms and decision-making processes are not transparent.

Action Items for Courts

Strong data governance provides the guardrails needed to effectively use AI. AI has the potential to improve data quality and data-driven decision-making, but rapid change also means that policies need to be carefully considered before a court reaches a crisis point. Courts need to lay the groundwork to meet the opportunities and challenges that new changes will bring *before* AI technologies are integrated into the courts on a large scale. Strong data governance will be a vital component of successfully adopting these technologies, and courts can begin to take important steps now to build the foundation for these changes.

Improve Data Quality

Currently, much of the court data that will be used to build AI tools using machine learning are rife with missing or inaccurate data or are collected by another entity outside of the courts. Addressing these issues will be crucial, and NCSC's Data Governance Policy Guide is a useful starting point.³

Standardize Across Jurisdictions

Software platforms and data standards are often incompatible across jurisdictions. The courts' increased reliance on future technologies will make it vital to reconcile technology and data formats, increase information sharing, and build systems that can share and merge mismatching data. Adopting the National Open Court Data Standards (NODS) is one step in this direction.⁴

Build Data Literacy

To meet the coming changes, court leaders and staff need a better understanding of the fundamental principles of data governance and data-driven decision-making. Other court stakeholders, including justice partner organizations, attorneys, and the public, also need a better understanding of how courts collect and use data. To this end, NCSC is developing an online course on data literacy.

Develop New Principles and Standards Related to AI Governance

NCSC is gathering court leaders, experts, and stakeholders to begin developing these principles.⁵ We are providing guidance on issues such as weighing the risks and benefits of AI for specific court tasks, appropriate roles for private vendors, defining and measuring acceptable AI performance, transparency in AI-assisted decision-making, data privacy and security, using a human-centered approach to integrate AI into the court workforce, identifying and eliminating disparate impacts of AI-assisted decisions, and ensuring full and equal access to the courts, among other considerations.

³ See Data Governance at <https://perma.cc/2WXW-2R8N>.

⁴ See NODS at <https://perma.cc/8ND7-A4XR>.

⁵ See Artificial Intelligence at <https://perma.cc/6MF6-YE9M>.

References

- Dastin, J. (2018). “Insight—Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women.” Reuters, October 10.
- “EU AI Act: First Regulation on Artificial Intelligence” (2023). Topics, European Parliament, December 19. <https://perma.cc/D485-RX5H>
- Joint Technology Committee (2024). “Introduction to AI for Courts.” *JTC Resource Bulletin*, March 5. <https://perma.cc/6G5T-QT8D>
- National Center for State Courts (2023). “5 Things to Know about Data Governance.” *Data Dives*, November. <https://perma.cc/C3K3-KPBK>
- Reinkensmeyer, M., and R. Billotte (2019). “Artificial Intelligence (AI): Early Court Project Implementations and Emerging Issues.” *Court Manager*, vol. 34, no. 3. <https://perma.cc/V9H7-H35K>
- Robinson, D., and S. Gibson (2019). “Data Governance Policy Guide.” National Center for State Courts. <https://perma.cc/U6AW-Z3DT>