# Principles and Practices for Using AI Responsibly and Effectively in Courts

## A Guide for Court Administrators, Judges, and Legal Professionals

*From the Thomson Reuters Institute/National Center for State Courts AI Policy Consortium for Law and Courts*

# Introduction

Generative artificial intelligence (GenAI) is transforming how we work, including within our state court systems. These tools can generate text, analyze documents, and assist with various tasks - but they must be used responsibly. This guide provides fundamental principles for the ethical use of generative AI in court settings. Central to this guide is the concept that judges, court administrators, and legal professionals should be empowered to use technology competently and consistent with their ethical obligations to best serve the public and the people who appear in their courts.

## What is Generative AI?

Generative AI: "The class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content" (Booth et al., 2024).

Large Language Model (LLM): "a class of language models that use deep-learning algorithms and are trained on extremely large textual datasets that can be multiple terabytes in size. LLMs can be classed into two types: generative or discriminatory. ("Large Language Model (LLM)," 2024).

> Generative LLMs: "are models that output text, such as the answer to a question or even writing an essay on a specific topic. They are typically unsupervised or semi-supervised learning models that predict what the response is for a given task" ("Large Language Model (LLM)," 2024).
>
> Discriminatory LLMs: "supervised learning models that usually focus on classifying text, such as determining whether a text was made by a human or AI" ("Large Language Model (LLM)," 2024).

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" (Booth et al., 2024)

## Levels of Risk

Courts must consider relative levels of risk in the broad spectrum of AI use. Using tools in ways in which they are not intended increases risk.

*Minimal risk:* Some uses of AI are minimal risk, such as the AI incorporated in common word processing tools. AI may be used to summarize meetings, to create a first draft of an email or a presentation, or to retrieve data from routine court filings. These low-risk uses of AI need a *human-on-the-loop*, providing supervisory oversight of the output of the AI.

*Moderate risk:* Some uses of AI are moderate risk. These include using AI to draft an opinion or to do research on a reliable platform. These moderate-risk uses of AI need a *human-in-the-loop* to verify any citations and ensure the accuracy and quality of the output – this is a higher level of scrutiny and involvement than human-on-the-loop.

*High risk*: Some uses of AI are high risk. These are uses of AI that may impact a person's rights. Examples include any use of AI to predict risk or recidivism in pre-trial release, sentencing, or other legal decisions. Uses of AI that are high risk must have a *human-in-the-loop*, requiring meaningful human review, revision, and ultimate decision-making.

*Unacceptable risk*: Uses of AI that are unacceptable risk are those that automate decisions about life and death, family relations, incarceration, health, and housing – in these cases, AI should not be used.

# Core Ethical Principles

### 1. Human Oversight and Responsibility

While AI can provide valuable assistance in researching, drafting, and analyzing information, it should never be the final arbiter of any court decision. Court staff and legal professionals maintain full responsibility for reviewing and verifying all AI-generated content, treating these tools as assistants rather than replacements for human judgment. This ensures that the essential human elements of justice remain at the center of our court system.

The level of human oversight required depends upon the specific use of AI. Any high-risk use requires a human-in-the-loop. Minimal risk uses of AI only require that there be a human-on-the-loop, monitoring the processes and outcomes.

### 2. Accuracy and Verification

Every piece of information generated by AI systems must have appropriate oversight before being relied upon in court processes or case management. This includes checking legal citations, verifying factual claims, and critically analyzing any conclusions or recommendations provided by AI tools. For minimal-risk uses of AI, establishing systematic verification processes helps maintain the high standards of accuracy required in legal proceedings.

## 3. Confidentiality and Privacy

Court systems often handle sensitive information that requires protection. When using AI tools, courts must implement robust security measures and data protection protocols based upon the risk level of the information. Public GenAI tools, while readily available, may not offer sufficient privacy guarantees for court-related information, but may be acceptable for minimal risk uses that do not expose sensitive data.

Courts should carefully evaluate AI platforms before use, ensuring they meet all necessary security requirements and comply with relevant privacy regulations. This may require working with specialized vendors who can provide appropriate safeguards for sensitive court data. Courts should develop clear protocols for what information can and cannot be shared with GenAI systems. When implementing a new AI tool, contracts or agreements should be clear on the ownership of, access to, and use of information entered into, used by, and generated from the AI tool. For example, courts should ensure that sensitive information entered into the GenAI tool is not used to train the tool itself.

## 4. Transparency and Community Support

Courts must maintain open and clear communication about their use of high-risk AI tools. This includes transparency regarding AI's use in court processes, disclosing AI assistance to relevant persons, when appropriate, and being prepared to explain AI-supported processes to the public. For high-risk uses of AI, it includes seeking community input in the adoption of AI tools, particularly from the portions of the community most directly affected. Any public-facing GenAI application must include a clear user interface with acknowledgement of the use of GenAI.

This transparency includes keeping detailed records regarding what tasks and purposes AI is used for in court processes. Transparency builds trust and ensures accountability in the integration of AI technologies.

## 5. Fairness and Bias Prevention

While AI systems could be implemented in ways to reduce bias and discrimination, AI systems can inadvertently perpetuate or amplify existing biases, making regular assessment of these tools crucial for maintaining equitable justice. An AI tool might produce outputs that reflect bias in training data or from other sources. Courts must actively monitor AI outputs for biased patterns or language, ensuring that the use of AI does not disadvantage any groups or individuals. This requires ongoing evaluation of AI systems' impact on different communities and maintaining strong protocols for equal access to justice. Courts should establish diverse oversight committees to review GenAI implementations and their effects on various stakeholder groups.

Any use of AI to predict risk or recidivism in pre-trial release, sentencing, or other legal decisions is a high-risk use that must be evaluated and monitored especially closely given that the algorithms used for these applications may have been developed and trained on biased data.

### 6. Competence and Training

Effective use of AI tools requires ongoing education and skill development. Courts must provide training programs that help judicial officers and staff develop a functional understanding of AI, including the capabilities, benefits, and risks of using AI systems. Understanding how AI tools work helps judicial officers and staff to use AI appropriately and to understand common pitfalls, such as "hallucinations: where GenAI confidently presents incorrect information. This training may include:

- regular updates on AI developments;
- information on how AI will impact workers;
- reskilling/upskilling, including practical training in using specific tools;
- best practices when using AI tools; and
- guidance on ethical considerations.

Courts should maintain detailed documentation of their AI systems and usage protocols, ensuring that staff can access clear guidelines when needed. Regular refresher training on AI helps maintain high standards of competence across the court system.

### 7. Inadvertent Plagiarism

Users should be aware that LLMs can reproduce verbatim text from their training data, a phenomenon known as "regurgitation." Judges and court staff who use AI need to be sensitive to potential plagiarism. Output of AI should be cited, just as any other source is. The citation will typically include the author, year of the version used, the title, source and the prompt used (see, for example, https://apastyle.apa.org/blog/how-to-cite-chatgpt and https://www.chicagomanualofstyle.org/qanda/data/faq/topics/Documentation/faq0422.html).

## Best Practices for Implementation

### 1. Start Small

Courts should begin their AI implementation journey with carefully selected low-risk tasks. This measured approach allows for learning and adjustment with minimal risk to core court functions. As successful outcomes are demonstrated and documented, courts can gradually expand their use of AI tools into other appropriate areas. Regular evaluation of these initial implementations provides valuable insights for future expansion.

### 2. Establish Clear Policies

Every court system needs written guidelines governing the use of AI tools, whether such tools are standalone or embedded in other products. These policies should clearly define appropriate and inappropriate uses of AI, establish approval processes for new AI implementations, create robust oversight mechanisms. The policies should include a risk framework, distinguishing between high-risk and low risk uses. Clear policies help ensure consistent and ethical use of AI across all court operations while providing staff with concrete guidance for daily decision-making.

Good GenAI policies might include guidance around:

- The level of scrutiny and review required, depending upon the risk of the use.
- Whether, when, and how individuals should disclose their use of AI tools.
- How private information should be protected and secured while using AI tools.
- How to respond when private information might inadvertently have been leaked to a model.

Review may include validating and measuring the accuracy of machine-generated information. If a tool is provided by a vendor, the vendor should provide information on how validity and accuracy were assessed.

## 3. Evaluate any potential use of AI

The use of AI should be based on solving a problem of the court or of litigants, whether such use is internal or public facing. Any potential use of AI should be evaluated with the following questions:

- What problem am I trying to solve?
- What is the level of risk of this use?
- What internal or external stakeholders will be affected?
- Is this a usage for which we should engage the affected stakeholders/community?
- If things go wrong, who will be harmed and at what potential cost?
- What data privacy and security protections are in place, and do they meet the needs of this community and my court?
- What evidence do we have that this AI tool solves the problem?
- Has there been sufficient testing to ensure that the tool is accurate, valid, and reliable?
- How is this a sustainable tool? What evidence do we have that the tool will be durable in its effectiveness? Is the tool likely to receive improvements/updates? If not, could the tool's usefulness decrease in the future?

## 4. Regular Review

The rapid evolution of AI technology necessitates ongoing assessment and policy updates. Courts should establish regular review cycles to evaluate AI tool performance, assess security and privacy measures, and identify any new ethical considerations that arise. These reviews should include feedback from various stakeholders in an iterative process and lead to policy updates when needed. This process ensures that AI use remains aligned with court values and objectives while adapting to technological advances.

# Common Risk Areas

## 1. Overreliance

There is a natural tendency to place too much trust in AI-generated content, particularly given its often polished and authoritative presentation. Courts must actively guard against this by maintaining robust human review processes and encouraging critical thinking at all levels. Staff should be trained to approach AI outputs with appropriate skepticism.

## 2. Privacy Breaches

The sensitive nature of some court information makes privacy protection paramount for that information. Courts must be especially vigilant in preventing unauthorized sharing or exposure of sensitive information through AI tools. This requires careful vetting of AI platforms, regular security audits, and clear protocols for handling sensitive information. Staff should receive ongoing training in privacy protection and data handling practices.

## 3. Bias Amplification

AI systems can unknowingly amplify existing societal biases, potentially affecting court decisions and processes. Courts must maintain active oversight of AI outputs for potential bias, regularly assessing the impact on different communities and demographic groups. This requires maintaining diverse perspectives in oversight roles and establishing clear protocols for identifying and addressing potential bias in AI systems.

# Conclusion

AI offers significant potential benefits for state courts but must be implemented thoughtfully and ethically. Success requires ongoing attention to these principles, regular training, and consistent oversight. As AI technology evolves, courts should regularly review and update their ethical guidelines while maintaining their fundamental commitment to justice, fairness, and human judgment.

Ethics is at the heart of the legal profession and requires that we innovate and adopt technology to improve access to law and justice. We hope this analysis helps to provide the beginning of a framework for making good decisions in your use of AI. The checklist below can help you implement these ethical considerations in your work.

# AI Ethics Self-Assessment Checklist

For Court Administrators, Judges, and Legal Professionals Using AI Tools

## Instructions

Answer each question with a Yes or No. Any "No" answer indicates an area that may need attention to ensure ethical and effective use of AI tools. Use the references after each section to find relevant guidance for improvement.

The self-assessment checklist should be conducted on a regular basis, especially as tools and uses change.

### Competence and Training

- [ ] Have I completed training on the AI tools I use?

- [ ] Do I understand both the capabilities and limitations of my AI tools?

- [ ] Do I stay informed about updates and changes to AI systems?

- [ ] Can I recognize common AI errors and problems?

- [ ] Do I know who to contact if I have questions about AI use?

- [ ] Have I documented my AI-related training and certifications?

### Human Oversight and Decision-Making

- [ ] Do I personally review and verify all AI-generated content before using it in any court-related work?

- [ ] Do I maintain final decision-making authority rather than deferring to AI recommendations?

- [ ] Have I established clear boundaries between low-, moderate-, and high-risk tasks to help ensure that I am using AI appropriately?

- [ ] Do I document when and how I use AI tools in my work, particularly when they are high-risk uses of AI?

### Data Security and Privacy

- [ ] Have I confirmed that the AI tools I use are approved by my court system?

- [ ] Do I know the privacy policy and terms of service for each AI tool I use?

- [ ] Do I have a clear understanding of what types of court information can and cannot be input into AI tools?

- [ ] Have I avoided entering confidential information into public AI tools?

### Accuracy and Verification

- [ ] Do I have a systematic process for fact-checking AI-generated content?
- [ ] Do I verify all legal citations and references provided by AI to avoid use of AI “hallucinations” ?
- [ ] Can I identify the source materials that inform AI-generated content?

### Transparency and Communication

- [ ] Can I readily explain to others how I use AI in my work?
- [ ] Do I consider whether, when, and how to inform relevant parties when AI tools have been used in their matters?
- [ ] Have I disclosed AI use to supervisors or other stakeholders as required?
- [ ] Can I demonstrate the role AI played in any particular task or decision?
- [ ] Have I considered whether and how to engage affected stakeholders?
- [ ] Do I cite AI as I would any other source?

### Bias and Fairness

- [ ] Do I regularly review AI outputs for potential bias?
- [ ] Have I considered how AI might affect different demographic groups?
- [ ] Do I compare AI-generated content across similar cases to check for consistency?
- [ ] Have I received training on recognizing AI bias?
- [ ] Do I have a process for reporting potential bias in AI outputs when identified?

### Risk Management

- [ ] Do I have a backup plan for when AI tools are unavailable?
- [ ] Have I identified high-risk areas where AI use should be limited?
- [ ] Do I maintain detailed records of AI-related issues or concerns?
- [ ] Have I established boundaries for appropriate AI use in my role?
- [ ] Do I know the procedure for reporting AI-related incidents?

## Action Planning

For any "No" answers above:

1. Prioritize areas needing immediate attention
2. Document specific concerns or gaps
3. Identify resources needed for improvement

4. Set timeline for addressing each issue

5. Schedule regular reassessment

# Scoring Guide

Count your "Yes/No" answers in each section:

- All "Yes": Excellent compliance

- 1-2 "No": Areas need attention

- 3+ "No": Significant improvements needed

# Follow-Up Actions

For each "No" answer:

1. Review relevant policies and guidelines

2. Seek additional training if needed

3. Consult with supervisors or AI specialists

4. Develop specific plan to address gap

5. Set deadline for resolution

## Resources Needed

Document what you need to improve compliance:

- [ ] Additional training

- [ ] Written policies

- [ ] Technical support

- [ ] Supervisor guidance

- [ ] Expert consultation

## Regular Review

Schedule to complete this checklist:

- Initial assessment: [Date]

- First review: [30 days]

- Regular review: [Quarterly]

- Annual comprehensive review: [Date]