

Key Considerations for the Use of Generative AI Tools in Legal Practice and Courts

Guidelines for Safe and Effective Use of Legal AI

AI Policy Consortium
for Law & Courts

NCSC



August 2025

Overview

As generative Artificial Intelligence (GenAI) tools increasingly support legal work, legal professionals and courts considering their adoption must ensure these tools serve as intended and do not undermine public trust and confidence. It can be difficult to understand how GenAI tools work and how they have been trained, but at least a basic understanding of these tools is necessary to ensure that legal professionals and judicial officers meet standards requiring competency, ethical integrity, and public trust. This document recommends foundational best practices, emphasizing that the rigor of application for these practices – from tool qualification to ongoing oversight – must be directly proportional to the intended use case and the inherent risk level associated with that use. Our recommendations are grounded in principles derived from legal education, professional evaluation, ethical oversight, and practical implementation pathways.

Intended Audience

- Court professionals considering AI tools for use by legal professionals
- Legal professionals considering AI tools for use by legal professionals

Key Definitions and Tool Typology

Legal GenAI Tool: Any software application using generative AI to perform, assist with, or automate tasks traditionally done by legal professionals or judicial officers.

Human Supervision (Human-on-the-loop): A supervisory oversight model where a human user monitors the output of an AI system, typically for minimal-risk applications. The user’s role is to provide general supervision and ensure the AI's work is acceptable for its intended purpose.

Active Human Engagement (Human-in-the-loop): An active and engaged oversight model where a human user actively interacts with, monitors, and potentially revises the AI's output before use. This level of scrutiny and involvement is higher than “human-on-the-loop” and is required for moderate to high-risk applications, ensuring meaningful human review and fact-checking, revision, and ultimate decision-making. In these scenarios, the AI serves only as an assistive tool, and the final decision rests solely with a human.

Understanding Risk Levels in Legal Generative AI Adoption

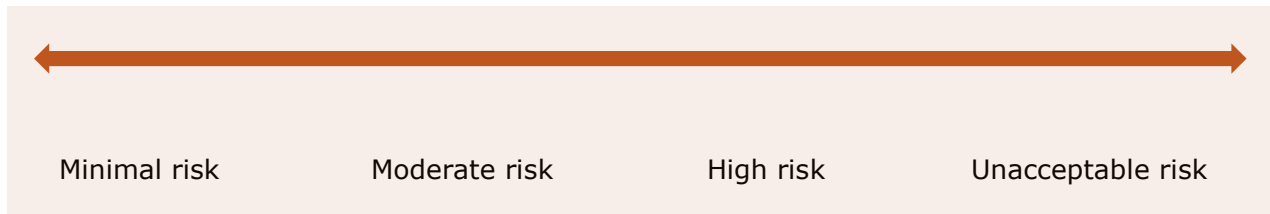
Legal professionals and judicial officers should adopt a risk-informed approach when considering the use of Legal GenAI tools. Not all uses carry the same potential for harm or error, and thus the level of scrutiny and human involvement must be proportionate to the risk. It is important to acknowledge that risk exists on a continuum, requiring nuanced consideration. It is also important to note that as technology develops, the perceived risk level of specific tasks may change.

- **Minimal risk:** This category encompasses use cases where the potential for direct legal harm, significant error, or impact on individual rights is extremely low. Outputs are typically administrative or organizational in nature and are easily verifiable by a human with general oversight. Examples include the AI incorporated in common word processing tools (such as spellcheck) or using AI to summarize meetings, to create a first draft of an email or a presentation, or to retrieve data from routine court filings for internal use. Other examples include the automation of reports where AI updates routine dashboards and the use of courthouse wayfinding kiosks. These uses of AI need human supervision, providing supervisory oversight of the output of the AI.
- **Moderate risk:** This level applies to uses where AI outputs directly inform professional work, requiring factual accuracy and verification, but do not involve independent legal interpretation or direct, substantive impact on individual rights without substantial human review. While errors could lead to procedural delays or minor inaccuracies, they are typically discoverable and rectifiable through diligent human engagement. These include using AI to perform research on a reliable platform, synthesize legal information for internal drafting, transcribe a recording, or assist with routine procedural filings. These moderate-risk uses of AI need a *human-in-the-loop* to verify content (e.g., citations and characterizations) and ensure the accuracy and quality of the output.
- **High risk:** This category covers uses where AI outputs involve legal interpretation, analysis, or application of law that could directly and substantively impact individual rights, due process, or the outcome of a legal proceeding. The consequences of error are significant, requiring sophisticated human legal judgment and active engagement to prevent adverse outcomes. The AI must serve strictly as an assistive tool, and final legal determination or official document generation remains solely with a human. Examples include any use of AI to draft a substantive legal brief or opinion, predict risk or recidivism in

pretrial release, sentencing, or other legal decisions. Uses of AI that are high risk must have active human engagement, or *human-in-the-loop*.

- **Unacceptable risk:** These are applications where automated decision-making technology (“ADMT”) for final decisions is deemed inappropriate, irrespective of human oversight, because they fundamentally undermine human discretion, accountability, or core principles of justice. The AI’s role should be entirely precluded from making the ultimate determination. Uses of AI that are unacceptable risk include those that automate decisions, such as those related to life and death, family relations, incarceration, health, housing, and other areas with the potential for material harm to an individual – in these cases, decisions should not be made by AI, even if humans are in the loop. They also include situations where human judgment is essential, such as assessing the credibility of witnesses. This applies as well in cases where the use of AI for even an initial draft or assessment has the potential to improperly influence a final decision, even if subconsciously. There may also be cases where the nature of the information input into the system, as with the identity of a law enforcement informant, creates an unacceptable risk.

Risks to privacy and confidentiality are increased for courts and for legal professionals when AI tools across a court or firm are connected through a Model Context Protocol (MCP), which enables two-way connections between data sources and AI-powered tools. While these tools can increase efficiency, they also increase the risk of privacy or confidentiality breaches.



In order to assess the risk, legal professionals must either understand how a given tool works or understand the empirical evaluation of the risks by an independent, trusted source. Based on the assessed risk, legal professionals need to take appropriate actions (such as human review) to sufficiently address that risk.

Types of Legal GenAI Tools (and associated general risk profiles)

- **Productivity Tools (Minimal to Moderate Risk):** Assist with routine administrative or document management tasks. Examples include summarizing internal meeting notes, creating first drafts of internal emails, generating checklists, or retrieving data from routine court filings for internal use. These generally require a human-on-the-loop.
- **Research Tools (Moderate Risk):** Retrieve, synthesize, and summarize legal information (e.g., case law, statutes, regulations, secondary sources). Examples include generating initial case summaries or statutory analyses. These typically require a human-in-the-loop to verify sources and accuracy.
- **Drafting Tools (Moderate to High Risk, depending on the document type):** Generate initial drafts of legal documents such as routine orders, notices, motions, pleadings, or general legal arguments. While drafting administrative documents may fall into the moderate-to-high spectrum, the drafting of substantive judicial opinions is considered High Risk due to its direct impact on legal precedent and individual rights. These tools necessitate a human-in-the-loop for thorough review, verification, and potentially substantial revision.
- **Decision-Support Tools (High Risk):** Provide analysis to assist in judicial decision-making, such as suggesting litigation outcomes, assessing legal risks, and assisting in pretrial release assessments. These tools have direct implications for individuals' rights and due process. As discussed above, there are many circumstances where such tools should not be used at all. Where their use is appropriate, they demand rigorous human-in-the-loop oversight and critical review.
- **Client-Facing/Public-Facing Tools (Moderate to High Risk):** Chatbots, intake screeners, or form fillers that interact directly with the public or litigants. These carry significant risks related to transparency, misinformation, privacy, and accessibility. The level of human oversight will vary depending on the tool's function and potential impact on access to justice.

Guiding Principles: A Risk- and Use-Case-Based Approach

The specific considerations for legal professionals or judicial officers in choosing and deploying a particular GenAI tool will depend fundamentally on what the tool is being used for and the associated level of risk. The more critical the task, the higher the potential impact on judicial processes or individual rights, and thus the greater the required scrutiny in qualification, evaluation, ethical oversight, and human supervision.

Core Considerations for GenAI Tool Adoption

1. Qualification Framework for Legal AI Tools

The foundation of a reliable GenAI tool lies, in large part, in its training. The scope and depth of this training must align with the tool's intended use and the sensitivity of the tasks it performs.

1.1 Structured Legal Curriculum Training

GenAI tools specifically designed for legal use cases must be trained on comprehensive, curated legal datasets. The *rigor*, *specificity*, and *documentation* of this training should increase with the tool's risk level and complexity of its intended use.

For **high-risk (such as drafting opinions, decision support)** or specialized applications, training must include:

- Foundational areas of law (such as contracts, civil procedure, constitutional law);
- Jurisdiction-specific statutes, procedural rules, and local court practices;
- Model Rules of Professional Conduct (as applicable to legal reasoning); and
- Practical legal skills such as legal writing, statutory interpretation, case analysis, client interaction, and advocacy.

Training datasets should be domain-specific (such as IP law, corporate law) where relevant to the intended use by the legal professional or judicial officer and clearly documented to ensure transparency and traceability of sources.

Even with such training, hallucinations and other errors remain a risk, and the burden is always on the user of AI to ensure that the output is accurate. Providers and users of Legal GenAI tools should be aligned on the measures taken to address anticipated legal use cases, such as “hard-coding” instructions to avoid hallucinations (i.e., “don’t make up case law”) and linking to underlying sources for human verification.

1.2 Jurisdictional and Cultural Sensitivity

Legal GenAI systems should be optimized for specific jurisdictions relevant to the court's operations, while also incorporating multi-jurisdictional knowledge for comparative and contextual purposes where applicable. This dual competency mirrors how legal professionals and judicial officers interpret and apply varying legal standards and is necessary for research tools, drafting tools, and decision-support tools.

2. Evaluation and Certification Standards

Prior to deployment, upon deployment, and throughout their use, Legal GenAI tools, depending on their risk profile, must undergo rigorous and ongoing evaluation to ensure performance, accuracy, and reliability, with the intensity of such evaluation scaling proportionately with the tool’s risk profile. In some instances, a built-in reporting mechanism could enable users to easily flag issues for feedback and corrective action.

2.1 Rigorous Evaluation Sets

Before deployment, Legal GenAI tools should be benchmarked using standardized legal evaluation sets. The *breadth*, *difficulty*, and *human vetting* of these benchmarks must be commensurate with the tool's risk profile and intended use.

For moderate or *high-risk* tools, such as decision support and opinion drafting, evaluations should:

- Be tailored to legal tasks such as brief drafting, statutory interpretation, case reasoning, and client advisory work;
- Include inputs and answer keys vetted by multiple legal professionals knowledgeable about the specific subject area and jurisdiction;
- Support consistent performance comparisons across systems (such as bar-exam-style testing, case reasoning exercises); and
- Be thoroughly vetted by legal professionals.

Ongoing evaluation must monitor performance degradation or data drift, similar to the way that continuing legal education (CLE) requirements help maintain proficiency of attorneys. The frequency and depth of ongoing evaluation should be higher for moderate to high-risk tools. Standardized testing scenarios could be used to measure performance across products and over time, provided that test inputs continue to evolve so Legal GenAI providers cannot pre-program products for known test inputs.

2.2 Global and National Standards Compliance

Developers should design legal AI tools that meet global and U.S. AI frameworks such as:

- International standards (e.g., ISO/IEC 24029-1, ISO/IEC 38507); and
- National frameworks (e.g., NIST AI RMF 1.0, NIST GenAI RMF 600-1).

Compliance with these standards is essential for all tools to ensure these systems meet accepted thresholds for safety, fairness, transparency, and accountability.

3. Ethical and Professional Oversight

Just as legal professionals and judicial officers are held to high ethical standards, so too must the tools they use be subject to robust ethical and professional oversight. The nature and intensity of this oversight are directly linked to the risk level and the potential impact of the GenAI tool.

3.1 Competency, Confidentiality, and Supervision

Under principles mirroring ABA Model Rules 1.1, 1.6, 5.1, and 5.3 (requiring competence, confidentiality, and supervision), and the Model Code of Judicial Conduct 1.2, 2.2, 2.3, 2.4, 2.5, 2.9, 2.12, and 3.5 (addressing confidence in the judiciary, impartiality and fairness, bias, external influences, competence, ex parte communications, supervisor duties, and use of nonpublic information), courts and legal professionals are responsible for the tools they deploy. Legal AI tools should be:

- Evaluated for hallucination rates, with specific thresholds and safeguards established, particularly for tools handling sensitive data or producing critical outputs (moderate to high risk);
- Evaluated for data retention, use, and privacy policies;

- Evaluated for their potential disclosure of confidential information, with strict data security protocols in place for all tools handling non-public court or client data;
- Transparent about their training data;
- Transparent about their limitations and the probabilistic nature of their outputs, especially for tools used in judicial processes (moderate to high risk); and
- Subject to appropriate human supervision: a *human-on-the-loop* for minimal risk tools and a mandatory *human-in-the-loop* for moderate to high-risk tools.

Legal professionals and judicial officers using these tools, particularly for moderate to high-risk applications, must have sufficient understanding of GenAI systems to critically evaluate and verify their outputs, identify and correct hallucinations and other inaccuracies, and ensure compliance with legal and ethical standards.

Transparency for Legal GenAI tools means relevant information should be available both to licensed users of the tools as well as recipients of work product (i.e., output) generated by such tools.

3.2 Continual Learning and System Updates

Just as many legal professionals must complete CLE credits, legal AI tools must evolve with the law and technology.

- Tools handling dynamic legal information (such as research, drafting, and decision support) must demonstrate a capacity for near real-time updating of statutes, regulations, and case law, reflecting continuous changes in the legal landscape (including an understanding when legal information is no longer valid or otherwise outdated);
- Incorporation of emerging ethical guidelines and procedural updates is essential for all tools, particularly those interacting with sensitive data or influencing court processes (moderate to high risk); and
- Regular retraining based on domain-specific changes and societal feedback is critical to prevent degradation of performance and/or fairness

Implementation Pathways (Strategies for Risk-Managed Adoption)

- **Pilot AI Credentialing Programs:** Partner with bar associations, legal educators, and technology providers to test AI systems against simulated, practical legal tasks relevant to court operations. Prioritize pilot programs for *moderate to high-risk* use cases to establish robust validation processes before broader deployment.
- **Develop Open Legal Evaluation Sets:** Collaborate with academic institutions, open-source repositories, and industry to create standardized benchmark tests across practice areas and jurisdictions. These open sets are particularly valuable for transparently assessing tools intended for *moderate to high-risk* applications in the legal and judicial systems.
- **Adopt Hybrid Governance Models:** Combine technical evaluation of performance on legal benchmarks with comprehensive ethical oversight including audit trails and fairness assessments. This blend is crucial for ensuring accountability, especially for tools handling sensitive court data or influencing judicial processes (*moderate to high-risk use cases*). Establish cross-disciplinary AI legal oversight boards comprising judges, legal scholars, technologists, and ethics experts to guide adoption strategies and risk mitigation for *all* Legal GenAI tools in courts.
- **Transition Support and Workforce Development:** Acknowledge workforce impacts and encourage upskilling, both in judicial and court administration education programs and in law schools. Active legal professionals and law students require education on AI literacy and oversight, focusing on critical evaluation, appropriate supervision (human-on-the-loop vs. human-in-the-loop), and ethical oversight of Legal GenAI outputs relevant to different risk levels.

Conclusion

To ensure that Legal GenAI tools contribute meaningfully and ethically to the legal system, legal professionals and judicial officers must establish qualification pathways tailored by risk and use case. Transparent training, rigorous testing proportionate to each Legal GenAI tool’s risk profile, and continuous oversight that prioritizes human involvement will fortify trust in legal AI, improve judicial efficiency, expand access to justice, and protect the integrity of the legal profession and the judiciary.

For more details on what to look for in Legal GenAI tools, please see our [due diligence tool](#), found at ncsc.org/ai.