Assessing AI in Action

A Case Study on Public Defender Use of General-Purpose AI Tools

September 2026

Task Force and Case Studies Context

The Council on Criminal Justice Task Force on Artificial Intelligence is a national, nonpartisan initiative to develop standards and evidence-based recommendations to guide the safe, ethical, and effective use of AI in the criminal justice system. In March 2026, the Task Force released a User Decision Framework to provide guidance for responsible AI procurement and deployment in the criminal justice field.

This case study is one of three that extends the work of that framework by demonstrating its guidance in action across different AI tools and hypothetical use cases, with a focus on substantive decision-making considerations.

These case studies follow the logic and structure of the framework and draw on information from practitioners, technology vendors, and publicly available research and product documentation. The assessments also reflect the insights and expertise of Task Force members, whose perspectives helped identify key questions, risks, opportunities, and implementation considerations. Together, these sources informed the case studies' deliberations and conclusions.

More Case Studies

Executive Summary

Many public defender offices are facing resource constraints and an overflow of cases, which can lead defense attorneys to consider using AI to support case management and reduce administrative burden. Ethical and legal considerations, along with limited budget flexibility,

often mean that offices have elected not to invest in AI tools for office-wide use. But some individual public defenders may circumvent these restrictions by relying on accessible, general-purpose AI tools to support their work, creating risks for legal work.

This case study examines general-purpose AI tools used by public defenders, including for research, document review, motion drafting, discovery analysis, and case preparation. The hypothetical use case considered here is a large, urban public defender office weighing the risks and opportunities associated with allowing its attorneys to use this category of tool for their work. A distinct consideration for users in this scenario is comparing AI-assisted defense work to the existing baseline under which public defenders manage their time, attention, and capacity.

Applying the five decision phases outlined in the User Decision Framework to this case study, the key takeaways are:

Phase 1: Foundation and Readiness

The goal is to reduce administrative burden and resource constraints without jeopardizing professional ethics and the quality of representation. Organizational readiness largely depends on the office's capacity to create and enforce rules and limitations around AI tool use.

Phase 2: Classification

The context and use case do not trigger limitations from the prohibited-use screening, though the relevant AI systems are often less transparent, less predictable, and harder to validate. The use case presents a low risk and substantial opportunity, as long as privacy preconditions are met, which triggers "**Standard Deploy**."

Phase 3: Procurement

Resources and processes for thoughtful and ethical implementation—such as disclosure considerations, device or network management, and use case restrictions—should be prepared in advance.

Phase 4: Implementation

The standard deploy label resulting from the use case's low risk and substantial opportunity classifications only requires the standard implementation, including data protections and human-centered design.

Phase 5: Ongoing Management and Reassessment

Scheduled reassessment for systems classified as low risk should take place before contract renewal. Reassessment beyond this could be triggered by developments such as expanding relevant use cases and changing ethics concerns.

Some questions raised by this case study can only be resolved by considering the specific product and jurisdiction in question. This is especially true for questions concerning confidentiality, device management, verification, accuracy, and cost.

Glossary

Air-gapped AI tool: An AI system run entirely on on-site, physically isolated hardware systems with no connection to the internet, cloud networks, or off-site servers.

Automation Bias: The tendency to over-rely on AI-generated outputs without sufficient critical evaluation.

Compliance Application Programming Interface: Automates regulatory compliance verification.

Consumer-tier AI tool: A publicly available AI product intended for individual users, typically accessed through a personal account and governed by standard terms of service rather than an organization-specific contract.

eDiscovery: The process of managing or producing Electronically Stored Information (ESI) to be used as evidence in legal procedures such as lawsuits, audits, or government investigations.

Enterprise-tier AI tool: An AI product provided under an organizational contract that includes administrative controls and negotiated protections.

FedRAMP (Federal Risk and Authorization Management Program): A program that ensures cybersecurity standards for digital products and services used by federal agencies.

Hallucination: When an LLM fabricates inputs or makes incorrect or unsupported claims.

LLM (Large Language Model): A generative AI system designed to understand and

generate human-like language.

Retrieval Augmented Generation (RAG): The AI process of sourcing information from large, unstructured sources of data outside of its training database to improve output accuracy.

Workflow Walkthrough

Phase 1: Foundation and Readiness

Define the Problem

The Framework at a Glance: *What is the agency? What problem does it face? Why is this product being considered? What alternatives to AI exist? What would success look like?*

The Agency and the Problem

The agency considered in this case study is a large, urban public defender office facing high caseloads and serious resource constraints. In this context, the office is not comparing AI-assisted defense to an ideal, fully resourced model of defense practice. It is comparing AI-assisted defense to the existing limitations under which attorneys triage time, attention, and capacity across too many cases.

The Opportunity

The office is considering acquisition of general-purpose AI tools because they may help reduce attorney time spent on lower-level work while preserving or improving the quality of representation, and because defenders may already be using consumer-tier models. Possible use cases include legal research, discovery summarization, motion drafting, case preparation, evidence and body-camera review, case prioritization, translation, and client communication drafts. These uses are not intended to replace attorney judgment, but they may help defenders begin tasks more quickly, organize large volumes of information, and

focus their time on strategy, client counseling, negotiation, investigation, and courtroom work

Defining Success

The relevant alternatives include paralegal hours, contract attorneys, manual research and document review, declining cases, and law-student fellows. These alternatives should be evaluated as comparators because the practical question is not whether AI is better than perfect representation, but whether it improves on the realistic options available to an under-resourced office.

Success would mean equal or higher-quality representation at lower attorney time per case. It would also mean no professional ethics issues, including no client-confidentiality breaches, no loss of privilege, minimal implicit offloading of defense decisions to AI, and no hallucination-induced errors reaching the court. In this setting, efficiency gains matter only if they are achieved without compromising attorney-client confidentiality, professional judgment, candor toward the tribunal, or the client's right to effective counsel.

Organizational Readiness

The Framework at a Glance: *What capacities does this product compel beyond generic technological readiness?*

Preserving Professional Ethics

The most important readiness consideration is the office's capacity to create and enforce rules and implement systems around the use of AI that preserve the professional ethics code that governs legal work. Public defenders, like all lawyers, have ethical duties that have specific application related to the use of technologies. These duties are set forth in the American Bar Association's (ABA) Model Rules of Professional Conduct, which have been adopted, in whole or in part, in all states.

In July 2024, the ABA issued Formal Opinion 512 regarding the use of AI tools.[1] The opinion focuses on the implications of such technology for several ethical duties, including the duties of competence, confidentiality, candor, and the proper supervision of attorneys and non-

attorneys. Several states have issued similar opinions or guidance on the use of AI. The office should have the capacity to ensure compliance with these ethical duties before seriously considering any AI tool or system.

One key risk is staff bypassing the office's approved, enterprise-tier or air-gapped AI tools by using consumer-tier AI accounts on personal devices for client work, which could route client information through products whose terms do not preserve confidentiality or privilege. Most defenders' offices have limited capacity to prevent this without new technical, policy, and training controls. Concrete controls could include mobile-device management on office-issued devices, network-level blocking of consumer AI domains on office Wi-Fi, a written policy with named consequences for personal-account use on client work, and training that explicitly names this prohibition.

Output Verification

The office should also have the capacity to verify relevant AI outputs. Verification is not incidental; it is the principal control for hallucination, inaccurate citations, flawed legal reasoning, and state or local law errors. If the effort required to verify outputs negates or outweighs the time saved by AI assistance, the opportunity case becomes weaker.

Office Policy Infrastructure

Finally, the office should be able to maintain and enforce an acceptable-use policy (see Appendix J). Many offices do not currently have a policy infrastructure that can define permitted uses, prohibited uses, data-entry restrictions, review requirements, attribution expectations, documentation practices, and training obligations for general-purpose AI. Without that infrastructure, safe deployment will be difficult and ethically risky, even if the product itself has strong contractual protections.

Phase 2: Classification

Assessing Specific Products

AI tools relevant for this case study vary by architecture, security posture, public-sector availability, training policy, legal specialization, and integration profile. Some tools are broad productivity suites embedded in office software. Others are general-purpose chat systems, or legal-specific products designed for research, drafting, or document review. These distinctions matter because the risk profile changes substantially depending on whether the office is using an enterprise or government product with contractual protections, or a consumer-tier product whose terms may permit training on user inputs or disclosure of prompts.

Offices should consider key product-specific questions. These include whether the product has FedRAMP authorization or comparable security certification, whether customer data are used by default to train models, whether public-sector or legal aid pricing is available, and whether any notable legal, operational, or implementation issues have been documented. Across AI tools, the following questions should be treated as central procurement and implementation issues rather than minor technical details:

- **Privacy and data handling:** What are the vendor's practices for handling case-related and sensitive legal information?
- **Hallucination:** How has the vendor measured hallucination rates and what have the tests shown?
- **Training on customer data:** What is the vendor's policy on using customer data for training the tool? How is the tool trained?
- **Model updates:** How are customers notified of migration to new models?
- **Verification:** What independent audits, red-teaming, or third-party evaluations exist or are underway? What are the legal-task benchmarks?

Prohibited Use Screening

The Framework at a Glance: *The prohibited-use screening asks whether the AI system poses an unacceptable risk to fundamental rights and should be prohibited in the criminal justice context. If the answer is yes to any screening question, the system is presumptively prohibited unless mitigation is documented.*

This screen is deliberately categorical, so each answer records only whether the use crosses a prohibited-use red line, not how risky it is overall. “No” means the use does not trigger that category; “no at screening” means it does not trigger the category but raises a concern that is not waived and instead carries forward into the graded risk-and-opportunity analysis later in this phase.

In this case, the screen for general-purpose AI use by public defenders does not produce a direct prohibited-use finding, but it does identify several issues that should be carried forward into the sector context and risk-opportunity analysis, especially those related to attorney-client privilege, effectiveness, bias, and accuracy. Given the lack of a prohibited-use finding, the agency should proceed to complexity, sector context, and risk-opportunity classification.

Screening Question	Answer	Reasoning
Q1: Does the system make autonomous decisions about liberty (e.g., detention, sentencing) without the possibility of substantial human review?	NO	Defenders do not delegate strategy or filing decisions to the model.
Q2: Does the system eliminate or impair a person’s right to contest a pending decision, or appeal a decision that’s already been made affecting their rights?	NO at screening	AI use does not by itself impair the client’s right to contest, though hallucinated citations or dropped issues could materially harm representation if unreviewed.
Q3: Does the system circumvent or undermine established legal or constitutional protections (e.g., due process, equal protection)?	NO at screening	Attorney-client privilege and effective assistance of counsel are the key concerns.
Q4: Does the system perform individualized tracking and surveillance of or otherwise have a chilling effect on a group engaging in lawful, constitutionally protected activities (e.g., First Amendment-related activities)?	NO	Same platforms can be used by other actors (prosecutors, surveillance vendors) to surveil protected activity—separate concern from this case.

Screening Question	Answer	Reasoning
Q5: Does the system target or select people based on protected characteristics and create unjustified discriminatory effects because of race, gender, religion, national origin, disability, or another legally prohibited ground?	NO at screening	Bias risk in legal-AI output (training-data bias and over-correction) is a low-risk-with-mitigation consideration.
Q6: Does the system systematically undermine human dignity or value (e.g., by publicly shaming or humiliating people or stripping them of all agency)?	NO	Does not undermine core liberty and human rights nor degrade the potential for individual human agency.

System Complexity and Interpretability

The Framework at a Glance: *How does this system score across Appendix C's four dimensions of decision transparency, predictability, validation difficulty, and adaptability?*

General-purpose AI tools should be treated as higher-complexity systems under the framework's four dimensions of decision transparency, predictability, validation difficulty, and adaptability. Even when tools are used for drafting or research rather than final decision-making, they generate outputs through large language models whose internal process cannot be traced step by step.

The tools' higher complexity triggers enhanced oversight under Appendix C, requiring case-specific decision rationales, independent technical validation, and ongoing validation after deployment.

Dimension	Assessment	Reasoning
Decision transparency	Less Transparent	Large Language Model (LLM) outputs are not traceable step-by-step; RAG-based legal tools produce citations but can still hallucinate.
Predictability	Less Predictable	Same prompt can produce different outputs across runs and model versions.

Dimension	Assessment	Reasoning
Validation difficulty	Harder to Validate	Legal correctness is partially contestable; ground truth varies by jurisdiction; state/local-law performance is particularly weak.
Adaptability	Adaptive	Underlying foundation models migrated frequently by vendors; prompts and behavior shift with updates.

Sector Context

The Framework at a Glance: *What is the current-practice baseline, and how does AI change risk and opportunity at the margin?*

The Baseline

The current baseline is that public defenders manage the caseload they have with the tools that are available to them. The AI comparison is not against a well-resourced ideal; it is against the existing landscape, which features resource and capacity limitations. That baseline matters because the cost of over-caution is not neutral. If AI can ethically and responsibly reduce or reallocate attorney workload, it may improve legal representation offered by offices that are under severe strain.

Operational Considerations

At the operational level, AI may enable a shift of some attorney attention and capacity toward higher-level work. For example, summarization, first-draft generation, research scoping, discovery review, and routine correspondence can consume significant time. If AI helps attorneys begin or complete those tasks faster, attorneys may be able to devote more time to client counseling, strategy, investigation, negotiation, and courtroom preparation.

Professional Ethics

At the legal level, professional ethics are a core concern, with variables like confidentiality, disclosure to clients, and attorney-client privilege being especially important considerations. With firm contractual guarantees, the legal regime that already governs lawyers' use of third-party tools—privilege, confidentiality, candor, competence, supervision, and communication—can absorb AI, eliminating the need to create a novel framework. Without those guarantees, the same tools can become privilege-destroying disclosures to a third party. *United States v. Heppner*[2] waived privilege and work-product protection due to defendant LLM-usage, though the court left enterprise-grade tools open, and the law here is still developing.

Court Disclosure

Furthermore, the court-disclosure landscape is evolving and jurisdiction-specific. As of March 2026, roughly 300 state and federal judges have standing orders addressing AI use, so implementation should track the rules of each relevant court.[3] The sanctions backdrop is also significant, with more than 1,700 documented cases worldwide involving AI-related legal errors as of July 2026.[4] Additionally, state and local law is where regulation of general-purpose large language models is weakest, but state and local law is where criminal defense most often operates. The office should not assume that strong performance on federal or general legal benchmarks translates into reliable state-and-local-law performance.

Evolving Obligations

State Bar ethical obligations and the ABA recommendations provide guidance for a professional-responsibility baseline, addressing topics like competence, confidentiality, communication, fees, supervision, and candor. Guidance may evolve to oblige lawyers to proactively advise clients about loss of privilege from clients' own AI use for case-related work. That possibility should be tracked as a developing duty, and the office should consider adopting a proactive-advice posture before any formal rule change.

Risk, Opportunity, and Classification

The Framework at a Glance: *With sector context and use case in mind, assess the proper classification level and implicated next steps.*

All things considered, the classification for this use case is **low risk** and **substantial opportunity**.

Low Risk

The low-risk classification only holds if firm privacy, confidentiality, and other ethical guarantees are negotiated and operational. With enterprise-tier contractual confidentiality or air-gapped systems, no-training-on-customer-data terms, operational controls preventing staff from using personal AI accounts for client work, and disclosure of any data sharing with third-party model providers bound to equivalent privacy terms, the confidentiality-breach pathway is closed. Hallucination and bias risks are then managed through lawyer-in-the-loop verification.

Substantial Opportunity

The substantial opportunity classification relies heavily on the potential for AI to help maintain or improve the quality of representation while reducing attorney time per case,[5] potential that is supported by some reports of time savings and broader efficiency.[6] The opportunity is especially significant because the access-to-justice gap is large and the baseline is an under-resourced defense system, rather than ideal representation.

Under the user decision framework, low risk combined with substantial opportunity triggers the Standard Deploy path, which means the agency proceeds with only the Level 1 implementation requirements.

However, this recommendation only holds if the privacy and other ethical preconditions are met. If preconditions are not met, the recommendation shifts to Generally Avoid. This classification is binary rather than sliding-scale because the absence of confidentiality changes the nature of the system. Without privacy protections, the tool creates an active privilege-loss pathway that the opportunity case cannot overcome.

Phase 3: Procurement

Budget and Resources

The Framework at a Glance: *What costs exist across procurement, deployment, and ongoing management?*

Procurement should account for the full cost of acquisition, integration, training, and ongoing monitoring. Although an agency's ability to proactively prepare for procurement depends on vendor transparency and readily available information, key considerations for public defender use of general-purpose AI tools include:

- The office should seek public-sector, legal-aid, or defender-specific pricing when possible; products designed for large law firms may not be affordable for indigent defense offices without reduced pricing or public funding.
- Integration costs will include enterprise-tier or air-gapped IT work, single sign-on, document-management integration, mobile-device management on office-issued devices, and network-level blocking of consumer AI domains. These costs are not secondary; they are part of the privacy and confidentiality architecture that makes Standard Deploy possible.
- Training and change management should address lawyer-in-the-loop, hallucination-spotting, jurisdiction-specific disclosure rules, and ethics refreshers grounded in best practices and applicable state rules. Attorneys and staff should understand both what the tools can do and what they cannot do, and they should also understand what information may and may not be entered into the tool. Continuing legal education (CLE) requirements offer a natural vehicle for keeping this training current.
- Ongoing monitoring should track the rate of AI-citation flags caught in attorney review, the rate of staff using prohibited tools, the rate of staff using approved tools for prohibited uses, and the volume of court disclosures filed. These measures will help the office determine whether the system is producing efficiency gains without creating confidentiality, accuracy, or compliance failures.

Contract Negotiation and Essential Terms

The Framework at a Glance: *Which Appendix F categories apply, and what does each look like for this specific product?*

The applicable contract terms should be drawn from Appendix F's Level 1 requirements. Level 2 (enhanced) terms are not required at the Standard Deploy classification.

Appendix F Category	Product-specific term that matters most	Notes
L1 System Requirements	Enterprise tier or air-gapped only; no consumer accounts on personal devices used for client work (post-*Heppner*)	Vendor pushback unlikely
L1 Performance Standards	Hallucination-rate disclosure and methodology; cross-vendor standardized question set	Vendors may resist disclosure; competition arguments
L1 Evidence and Validation	External legal-task benchmarks; criminal defense-specific accuracy data if available	Limited public benchmarks for criminal defense specifically
L1 Transparency and Explainability	Disclosure of model updates and migration notification; citation-grounding for legal-research outputs (Retrieval Augmented Generation, or RAG, style) where available	Vendors may prefer unilateral update authority
L1 Data Governance and Privacy	Contractual confidentiality; no training on customer data; Compliance Application Programming Interface (API) or equivalent for eDiscovery; data-residency where applicable; disclosure of and binding on third-party model providers	Enterprise vendors likely to accept. Vendors for consumer-tier products may not, which is the binary.
L1 Bias Testing and Fairness	Disclosure of any known bias issues and mitigation strategies	Vendors may resist standardized testing
L1 Human Oversight and Override	Lawyer-in-the-loop verification required for all AI-assisted work; no "approve all" user interface shortcuts in legal contexts	Vendor pushback unlikely; this is a product-design question
L1 Training and Support	Vendor training on hallucination behavior and verification workflows	Vendor pushback unlikely

Appendix F Category	Product-specific term that matters most	Notes
L1 System Changes and Updates	Mandatory notification for model changes; right to refuse updates that change the risk profile (a material change to the privacy preconditions is itself a Phase 5 trigger)	Vendors may prefer unilateral update authority
L1 Liability and Indemnification	Vendor accepts liability for system errors; indemnifies for civil rights violations caused by the system	Vendor pushback likely; where it is not possible the office should fall back to liability caps and audit rights instead of treating this clause as a dealbreaker
L1 Termination	Termination for convenience, performance failure, or rights concerns	Vendor pushback likely to mirror standard termination power pushback

Phase 4: Implementation

The Framework at a Glance: *How will the system function in your environment? How will you ensure it performs as intended?*

Level 1 Requirements For All Systems

These baseline recommendations apply to every AI system deployed in criminal justice settings.

Human-Centered Design and Training

The risk of automation bias is heightened by caseload pressure. In an overburdened office, “approve as written” can become the path of least resistance, and training, supervision, and protocol will be needed to counter that tendency. Implementation should begin by framing AI outputs as drafts, research starting points, or internal work aids, never as final filings. The

office should require attorney verification and sign-off of AI-generated products before any external use. Attorneys should be trained to treat AI output as a starting point that may be useful but must be checked, challenged, and adapted to the client's facts, jurisdiction, and strategy.

Training should cover what hallucinations look like and how to catch them, jurisdiction-specific disclosure rules, State Bar ethics and obligations, ABA guidance, confidentiality rules governing what may and may not be entered into the tool, and automation-bias risks and mitigations. The training should be practical and role-specific, with examples drawn from legal research, motion drafting, discovery review, client communication, and filing review.

Transparency, Accountability, and Privacy

Implementation should address court disclosure, accountable officials, and privacy and data-security measures. Court disclosure is jurisdiction-specific, so the office should maintain a current matrix of applicable standing orders, local rules, and judge-specific requirements. Disclosure should be built into the filing-review checklist rather than left to individual attorney memory.

Verification time is a key consideration. If the effort required to verify AI-generated citations, factual summaries, legal standards, and draft arguments negates or outweighs the time saved by the tool, the opportunity case is significantly harder to make. Independent expert input is needed on the verification-time tradeoff, and the office should measure verification time during any pilot.

Personal AI accounts used for client work are a primary privacy concern. Attorneys or staff using consumer-tier ChatGPT, Gemini, or Claude accounts on personal devices could undermine the office's enterprise-tier or air-gapped system procurement and create privilege or confidentiality exposure. The controls are mobile-device management on office-issued devices, network-level blocking of consumer AI domains, a written policy with named consequences, and training that names the failure mode directly. This is the key operational control for maintaining the Standard Deploy classification.

Client communication should also be considered. ABA Opinion 512 may be revised to urge lawyers to proactively advise clients about loss of privilege from the client's own AI use. The office should consider adopting that proactive-advice posture before any formal revision. As a concrete implementation requirement, the office should allow only enterprise-tier accounts to

be used for any work touching client information.

Pilot Program

The pilot should define scope, duration, success criteria, comparison methodology, and a formal decision-making process. Because the system is classified as Standard Deploy if privacy preconditions are met, pilot duration is at the office's discretion, rather than a six-month minimum. The pilot should nevertheless run long enough to evaluate the success criteria meaningfully. Consider partnering with academic researchers to design and execute a robust pilot.

The scope should consider which use cases are included, such as research, drafting, discovery review, and case prioritization, as well as which divisions participate, such as felony, misdemeanor, or juvenile. Success criteria should include time savings net of verification, case-outcome quality, attorney satisfaction, client satisfaction, and the absence of sanctions, ineffective-assistance claims, or confidentiality incidents.

The comparison methodology should ideally use a matched-attorney or matched-case design. The office should compare AI-assisted matters to baseline rates of motion success, time per matter, and other quality indicators. The framework recommends a formal go/no-go decision after the pilot. If the pilot reveals confidentiality incidents, sanctions, or plausible ineffective-assistance concerns, then the privacy preconditions are not operational in practice, and the recommendation shifts to Generally Avoid.

Acceptable-Use Policy

Appendix J makes the case for an Acceptable-Use Policy for general-purpose AI in case-related work, regardless of risk classification. This obligation is parallel to the Standard Deploy or Generally Avoid call in that it is not a substitute for the Phase 1 through Phase 5 analysis.

The policy should define permitted uses, prohibited uses, data-entry restrictions, review and attribution requirements, documentation practices, and training obligations. Permitted uses may include legal research drafting, discovery summarization for internal review, motion drafting starts, and routine correspondence. Prohibited uses should include communicating with clients through AI without review, entering sealed records or privileged client material

into non-enterprise tools, and presenting AI output as original analysis without substantive review.

Data-entry restrictions are especially important. Personally identifiable information, case-specific details, confidential informant data, and sealed records should not be entered into external AI systems unless the office has confirmed the provider's data handling through formal review and negotiated a contract that supports it. All AI-generated content used in official work products, including citation verification, should be substantively reviewed by the responsible attorney before filing. Operators should be able to identify that AI was used, the purpose of its use, the review conducted, and who performed the review.

Training under the Acceptable-Use Policy should address capabilities and limitations, the tendency of AI systems to generate confident but inaccurate outputs, potential bias, and professional obligations under ABA recommendations and applicable state rules. The policy applies regardless of which side of the privacy-precondition line the office ultimately falls on.

Phase 5: Ongoing Management and Reasses

The Framework at a Glance: *How should ongoing monitoring and periodic reassessment be leveraged to ensure AI tools continue to function as intended?*

Ongoing management should define the scheduled reassessment cadence, triggered reassessment events, and product-specific indicators that would require review. For a low-risk system, scheduled reassessment should occur before contract renewal. If the system is reclassified to substantial risk but not a General Avoid recommendation, annual reassessment should apply. The agency should monitor indicators that reveal accuracy, privacy, efficiency, and legal-system impact, including but not limited to:

- The rate of bad or misleading citations generated by AI that are caught by attorneys in review
- The rate of staff using consumer-tier AI accounts for client work, detected through mobile-device management logs, network monitoring of consumer AI domains, and self-report
- The volume of court disclosures filed and court sanctions
- Verification time per matter

- Client complaints
- Outcome differentials between AI-assisted and non-AI-assisted matters

Triggered reassessment should occur after underlying model migration if key product or context changes occur, such as:

- A vendor swap from one foundation model to another
- Material policy changes by the vendor, including training-on-customer-data terms or data retention practices
- Scope expansion to new use cases, such as client-facing AI or case prioritization
- Performance problems, such as citation-flag rates moving outside agreed thresholds
- Rights concerns such as sanctions, ineffective-assistance claims, or confidentiality incidents
- Environmental changes such as ABA Opinion 512 revisions, state-court rule changes, or *Heppner*-line case law developments

Acknowledgments

This case study is from the Council on Criminal Justice Task Force on Artificial Intelligence, a national, nonpartisan initiative developing standards and evidence-based recommendations to guide the safe, ethical, and effective use of AI in the criminal justice system. The case study is a product of its members, who graciously shared their time and expertise.

Jesse Rothman produced this report, with support from Cameryn Farrow, Andrew Page, Jonathan Wroblewski, and others from the Council on Criminal Justice team.

James Anderson and RAND serve as research partners to the Task Force.

Support for the Task Force on Artificial Intelligence comes from the Georgia Power Foundation, Heising-Simons Foundation, The Just Trust, Microsoft, and The Tow Foundation, as well as the John D. and Catherine T. MacArthur Foundation and other CCJ general operating contributors.

Suggested Citation

Council on Criminal Justice. (2026). *Assessing AI in action: A case study on public defender use of general-purpose AI tools.* https://counciloncj.org/assessing-ai-in-action-a-case-study-on-public-defender-use-of-general-purpose-ai-tools/

Endnotes

[1] American Bar Association Standing Committee on Ethics and Professional Responsibility. (July 2024). *Generative artificial intelligence tools* (Formal Opinion 512). https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf

[2] *United States v. Heppner*, No. 25 Cr. 503 (JSR), 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026). https://storage.courtlistener.com/recap/gov.uscourts.nysd.652138/gov.uscourts.nysd.652138.27.0.pdf

[3] Cohan, A. (March 2026). AI disclosure in court: 300+ rules you need to track. Hintyr. https://www.hintyr.com/blog/ai-disclosure-court-rules-guide

[4] Charlotin, D. (June 2026). AI hallucination cases. Damien Charlotin. https://www.damiencharlotin.com/hallucinations/

[5] Ikeh, M. (March 2025). Bringing AI to the District Attorney's office: A policy framework for innovation in criminal justice. SSRN. http://dx.doi.org/10.2139/ssrn.5207487

[6] UC Berkeley Law. Existing AI tools for criminal defense. https://www.law.berkeley.edu/research/criminal-law-and-justice-center/focus-areas/ai-for-public-defenders/existing-ai-tools/