Assessing AI in Action

A Case Study on Police Use of AI-Generated Report Writing 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

Police departments across the country are beginning to use artificial intelligence tools to draft incident reports from body-worn camera audio and other officer inputs. These products often hold promise to reduce administrative burden, improve report quality, and allow officers to

spend more time in the field. Police reports are often regarded as simple administrative documents, but they can be highly consequential, shaping charging decisions, plea negotiations, sentencing outcomes, civil liability, and the ability of defendants to challenge the evidence against them.

This case study examines AI-generated police report writing tools that produce draft incident narratives from body-worn camera data, including audio or video. Currently, these products use large language models to convert recorded interactions, transcripts, officer voice notes, or related inputs into draft reports for officer use.

The hypothetical use case considered here is a 200-officer, mid-size, ex-urban police department weighing whether to adopt this category of tool. The department is operating in a context familiar to many similarly situated agencies: Report writing consumes substantial officer time, delays in report completion can slow investigations and prosecutions, and subpar report quality can create downstream consequences for prosecutors, defendants, supervisors, and the agency itself. A key consideration for agencies in this position is how to define the conditions under which use is permissible and the conditions under which the tradeoffs outweigh the potential benefits.

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, improve report quality, or both more effectively than non-AI alternatives would allow. Organizational readiness largely depends on the department's ability to navigate variables like automation bias and legal expectations.

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 both substantial risk and substantial opportunity, which triggers "**Careful Implementation**."

Phase 3: Procurement

Resources and processes for thoughtful implementation, such as officer training on treating AI products as suggestions for review, should be prepared in advance.

Enhanced contractual protections should be employed to address the substantial risk of the use case and tools, including additional terms around auditability.

Phase 4: Implementation

The careful implementation label resulting from the use case's substantial risk and opportunity classifications requires enhanced implementation, including incorporation of relevant due process and Sixth Amendment protections.

Phase 5: Ongoing Management and Reassessment

Reassessment should take place annually, as is recommended for systems and use cases that are classified as substantial risk. Reassessment in addition to the annual schedule could be triggered by developments such as vendor changes to core prompts and expansion into new input types beyond audio or video.

Some questions raised by this case study can be resolved only by considering the specific product and jurisdiction in question. This is especially true for questions concerning error rates (i.e., hallucinations), validation, bias, auditability, and prosecutorial acceptance.

Glossary

Access Mode: How the AI product accesses and uses the underlying foundational model, such as through a third-party programming interface, a vendor-hosted service, or a dedicated or self-hosted deployment.

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

Brady: Under the Supreme Court's holding in *Brady v. Maryland*, prosecutors must disclose material evidence held by the government that is favorable to the defense. This includes information that could lessen a sentence, undermine a witness's credibility, or cast doubt on guilt.

CJIS (Criminal Justice Information Services): The largest division of the FBI and a centralized hub for law enforcement and national security data.

Domain Adaptation: Adjusting an AI system so it works better in a particular field or setting

by training or fine-tuning it with relevant examples or data. For example, a general AI model may be adapted to better understand police-report language, legal terminology, or correctional records.

Due Process: A concept in the U.S. Constitution, largely rooted in the Fifth and Fourteenth Amendments, that guarantees that no one may be deprived of life, liberty, or property without proper legal procedures.

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.

Migration: A change from one underlying AI model or version to another. For example, a vendor may switch its product from an older model to a newer one, which can change how the tool performs or what outputs it produces.

Model: A trained program that identifies patterns in data and produces an output when given new input. Different AI tools may rely on different underlying models, which can affect how they perform and what kinds of outputs they produce.

Prompt: The instructions or information given to an AI system to guide what it does. For example, a prompt might ask the system to turn body-camera audio into a draft incident report.

Sixth Amendment: Part of the Bill of Rights of the U.S. Constitution, establishing rights of criminal defendants.

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 200-officer, mid-size, ex-urban police department facing recruitment and retention pressures, overtime costs, and operational demands that are common across many similarly situated agencies. Report writing is a major source of officer dissatisfaction and a recurring administrative burden. When reports are delayed, investigations and prosecutions can also be delayed. When report quality varies across officers, the consequences can extend beyond internal workflow and affect prosecutions.

The Opportunity

The department is considering AI-generated police report writing tools because they appear to offer a way to reduce administrative burden and improve report accuracy and completeness. If the tools worked as claimed, they could allow officers to spend less time writing reports and more time on other public safety responsibilities. They might also help standardize report narratives across officers, reduce omissions, and create a clearer record for supervisors and prosecutors. However, these potential benefits should be assessed against the realities of evidentiary integrity, officer accountability, and constitutional rights.

Defining Success

In this context, the important metric is not simply how AI use compares to standard practice, especially since standard practice already includes variation in report quality, inconsistent officer writing practices, and an under-measured baseline error rate. The more useful framing is whether AI-generated report writing tools perform better than the specific alternative the agency would otherwise pursue. Those alternatives could include structured report templates, voice-to-text dictation without large language model generation, additional administrative staff, and simplified workflow design.

Success would mean more than faster production of reports or improved officer satisfaction while, at minimum, maintaining report quality and avoiding any degradation in evidentiary integrity. Such benefits have not yet been established. In fact, the strongest independent evidence available comes from a recent randomized controlled trial, and those findings contradict vendor time-savings claims.[1]

Organizational Readiness

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

Automation Bias

A significant readiness consideration is the ability to control for automation bias. Officers may treat an AI-generated draft as an authoritative record of what occurred, especially when the draft is polished and created from body-worn camera data. To mitigate that risk, the agency may need a protocol requiring officers to document independent recall before reviewing the AI draft.

Discovery Material

The agency also may need the capacity to handle prompts and drafts as potential discovery material. Many police departments do not currently treat AI artifacts as discoverable records. If system prompts, initial drafts, model versions, and final reports should be retained and produced, legal counsel, supervisors, records personnel, and technology staff will need relevant policies and training that many small and mid-size agencies do not currently have.

Outcome Tracking

Piloting an AI tool also requires more capacity than what is needed for a basic software trial. The agency would need to accumulate enough reports across incident types to detect differences in report quality, time savings, error rates, and downstream outcomes. Tracking prosecutorial acceptance, defense challenges, and case impacts requires data infrastructure that many small and mid-size departments may lack. Ongoing monitoring presents a similar challenge. The agency should define terms such as “AI-flagged content” and “override/edit rate” operationally before officially deploying the tool.

Phase 2: Classification

Assessing Specific Products

AI-generated police report writing tools vary in architecture, inputs, default scope, integration profile, and available evidence base. Some tools generate narratives exclusively from body-worn camera audio, while others rely on officer dictation, voice notes, video, or a combination of sources. Some products are offered by established public safety technology companies, while others come from newer AI-native vendors. The differences matter because the technical design of the product affects what data the tool uses, how the report is generated, what records are created, and what can later be audited or disclosed. These differences can also affect the tool's reliability: Performance may vary depending on the type and quality of the input, the underlying model, and the task the system is asked to perform, making product- and use-case-specific validation essential.

Agencies should consider key product-specific questions concerning the architecture of each system, the identity and access mode of the underlying foundation model, the role of system prompts, how training or fine-tuning data are used, what data are processed and stored, what performance claims are being made, and how the product integrates with the agency's existing body-worn camera, records management, and case management systems. Across AI-generated police report writing tools, several questions should be treated as central procurement and implementation issues rather than minor technical details:

- **System prompts and versioning:** Are prompts retained as part of the audit trail; produced in discovery; disclosed to officers, supervisors, prosecutors, or defense counsel; versioned over time; and available for retrospective comparison?
- **Underlying model and training:** What are the foundation-model identity, access mode, migration policy, and relevant fine-tuning or domain adaptation processes? Does the model continue to learn from agency usage after deployment?
- **Performance, validation, and bias:** What are the measured hallucination or error rates and the methodology used to define and measure them? What are the independent validation outcomes, under accurate operational conditions; bias audit results; and demographic differences in output quality, tone, or content?
- **Data handling:** Are data shared with third-party model providers? What are the contractual terms governing that sharing, CJIS compliance, and the handling of

personally identifiable information across the input-to-transcript-to-draft pipeline?

- **Deployment:** What other departments are currently using each product, and how have claims about adoption and performance been assessed in those contexts?

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 AI-generated police report writing tools does not produce a direct prohibited-use finding, though that result should not be read as a finding that the product category is low risk. Several concerns that do not trigger the prohibited-use screen—especially prompt opacity, potential impairment of meaningful contest, constitutional uncertainty, and disparate-impact risk in report language—remain relevant to the substantial-risk analysis. 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	Officers retain sign-off; reports are not autonomous decisions.

Screening Question	Answer	Reasoning
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	Prompt-opacity issue creates a practical impairment of meaningful contest—flagged but not as a prohibited-use trigger.
Q3: Does the system circumvent or undermine established legal or constitutional protections (e.g., due process, equal protection)?	NO at screening	Sixth Amendment confrontation and Brady obligations exist.
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	The product writes incident reports; it is not a surveillance tool.
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	Disparate-impact risk in language patterns is a Substantial-Risk factor, not a prohibited use.
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?*

AI-generated police report writing tools should be classified as higher-complexity systems under the framework's dimensions of decision transparency, predictability, validation

difficulty, and adaptability. The tool does not merely fill out a template or transcribe speech; it generates open-ended narrative text from unstructured inputs, using a large language model whose internal reasoning cannot be traced step by step due to vendor opacity.

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	LLM-generated narratives cannot be traced step-by-step; even the system prompt that shapes outputs is not currently exposed.
Predictability	Less Predictable	Same audio + same prompt on different days/model versions can produce materially different narratives; current vendor practice does not support retrospective comparison across prompt versions.
Validation difficulty	Harder to Validate	Inputs are unstructured audio; outputs are open-ended prose. "Correctness" is contestable and partially subjective.
Adaptability	Partially Adaptive	Underlying foundation models are migrated by vendors; consider whether agency usage is fed back into training.

Sector Context

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

The Starting Point

The current practice in law enforcement is that officers write reports from memory and notes after an incident or shift, sometimes using templates or other basic aids. Report quality varies by writer, supervisor review, incident type, workload, and training. These reports are not merely internal administrative documents; they carry forward into charging, plea negotiations, sentencing, impeachment, civil litigation, and public accountability.

Operational Considerations

At the operational level, AI-generated report writing changes the authorship process. The shift is from "officer writing from memory" to "officer reviewing an AI-generated account derived from body-worn camera material or other input." That shift could improve completeness, especially when the AI tool captures details an officer might omit. It could also degrade independent recall if the AI draft influences the officer's memory or encourages the officer to accept a narrative generated by the system rather than reconstructing and verifying events independently.

Legal Considerations

At the legal level, the technology introduces a new category of evidence and potential discovery material: the initial AI draft, the system prompt, the model version, and the editing history between the AI-generated draft and the final officer-approved report. Existing discovery and Brady regimes were not designed for this production process. Recently passed legislation in Utah[2] and California[3] are first-generation legislative responses that require agency policies around AI use and disclosure, but neither fully resolves the prompt issue.

A Sixth Amendment Confrontation Clause concern also exists: Officers may be unable to testify clearly about which sections of a report were generated by AI, which sections were edited, and what the officer independently remembered. Although created for a different context, the Confrontation Clause's adversarial-scrutiny purpose remains relevant. Brady concerns may also arise when AI involvement, the initial draft, or the system prompt contains or reveals material information, particularly if the final report differs from the initial AI-generated version. The precise doctrinal answer requires independent expert input, but the risk is significant enough to shape procurement, implementation, and retention requirements.

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 **substantial risk** and **substantial opportunity**.

Substantial Risk

Several factors support the substantial-risk finding. Documented hallucinations have appeared in finalized reports, including the Heber City frog[4] and King County phantom officer[5] examples. Sixth Amendment cross-examination and potential Brady disclosure obligations remain unresolved. AI-generated drafts may shape officers' memories of events, creating a confirmation-bias risk that undermines independent recall. System prompts also shape outputs but may not always be retained, versioned, or produced in discovery, creating a structural accountability gap that is not fully addressed by CA SB 524 or Utah SB 180.

Vendor refusal to retain and produce prompts in discovery should be treated as a Generally Avoid trigger, not a negotiating posture. If the prompt materially shapes the report but cannot be retained, versioned, audited, or produced when legally relevant, then the agency cannot adequately manage the evidentiary and rights-related risks of the system.

Substantial Opportunity

Several factors support the substantial-opportunity finding. Adoption is large and growing, and supervisor and prosecutor reports of quality improvements across multiple departments are credible. At the same time, the classification could shift toward Generally Avoid if vendors refuse to retain, version, and produce system prompts in discovery as standard contract terms; if independent studies continue to fail to replicate vendor time-savings claims under realistic conditions; if prosecutorial rejection[6] expands; or if independent bias audits reveal systematic disparities in report language or framing without a credible mitigation path.

Under the user decision framework, substantial risk combined with substantial opportunity requires careful implementation. That path means the agency should proceed with all Level 1 and Level 2 (enhanced) requirements, plus product-specific contract terms on prompt retention, initial-draft retention, and bias auditing.

Phase 3: Procurement

Budget and Resources

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

Procurement should, at minimum, account for the full cost of a tool's acquisition, integration, training, monitoring, and technical fixes, along with community engagement. Although an agency's ability to proactively prepare for procurement depends on vendor transparency and readily available information, key considerations for AI-generated police report writing include:

- Integration costs will depend on whether the tool can work with the department's body-worn camera and records systems, as well as any setup needed to prepare data or configure the tool.

- Meaningful training and change-management costs should be expected. Officers will need training on a review-not-adopt posture, including how to identify and correct hallucinations, omissions, cognitive biases that arise with human-machine interaction, and overconfident narrative framing. Supervisors will need training to review how much human correction a tool needs and to oversee report quality. Legal counsel will need training on prompt-and-draft retention as discovery material, including how to advise the agency on Brady, confrontation, due process, and state-law disclosure requirements.

- The agency will need infrastructure to track key success metrics, conduct periodic bias audits, and complete annual reassessment since the system is classified as substantial risk. Technical-fix costs implicate liability allocation for hallucination-driven evidentiary issues. Vendor pushback should be expected on liability and indemnification.

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 the Level 1 and Level 2 requirements in Appendix F.

Appendix F Category	Product-specific term that matters most	Notes
L1 System Requirements	Default limited scope during pilot; expand only on documented review	Agencies should want broader scope from day one
L1 Performance Standards	Define and contractually fix hallucination-rate methodology; vendor reports monthly	Vendors may resist disclosure; trade secret arguments
L1 Evidence and Validation	Independent third-party validation as a renewal condition	Frame as "validation, not audit of trade secrets"
L1 Transparency and Explainability	Prompt retention as a contract requirement (beyond CA SB 524); initial-draft retention beyond CA SB 524 minimum; underlying-model migration notification	"Prompts are proprietary" is the predictable response
L1 Data Governance and Privacy	Vendor must certify it will NOT train on agency data without written consent; data sharing with third-party model providers fully disclosed	Vendor pushback unlikely; terms vary
L1 Bias Testing and Fairness	Bias audit obligation with disaggregated reporting by subject demographics	Frame as "compliance with anticipated state laws," not internal preference
L1 Human Oversight and Override	Officer review and edit captured at the per-section level; no "approve all" UI shortcuts	Vendor pushback unlikely; this is a product-design question
L1 Training and Support	Vendor training on automation-bias module plus product use	Vendor pushback unlikely

Appendix F Category	Product-specific term that matters most	Notes
L1 System Changes and Updates	Mandatory notification for system or prompt change; right to refuse updates that change the risk profile	Vendors may prefer unilateral update authority
L1 Liability and Indemnification	Indemnification for hallucination-driven evidentiary issues	Vendor pushback likely
L1 Termination	Termination for convenience, performance failure, or rights concerns	Vendor pushback likely to mirror standard termination power pushback
L2 Enhanced Validation	Independent validation tied to renewal; pilot success criteria pre-set	Vendor pushback likely to mirror standard burden arguments
L2 Enhanced Rights Protections	Legal review of Sixth Amendment, Brady, due process; documented defendant access to prompt + initial-draft + final-report set	Vendor pushback likely to mirror standard burden arguments
L2 Community Engagement Support	Vendor participation in community meetings and public Q&A	Vendors may resist disclosure of prompts, which public groups may request
L2 Enhanced Auditability	Algorithm transparency commensurate with IP protection; audit rights for independent experts; defense-bar access in litigation	Vendor pushback likely

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

Implementation should begin with how the product presents its outputs to officers. The AI-generated report should be presented as a draft, not a recommendation, decision, or authoritative account. The user interface should not nudge officers toward approval over editing, and it should not normalize one-click acceptance of system-generated language.

Automation bias is acute in this context. Officers may treat an AI draft as the most accurate or complete account of events, especially when it is generated from body-worn camera data and presented in polished narrative form. To reduce this risk where operationally feasible, implementation should include an independent officer statement or contemporaneous notes on the officer's first-hand knowledge and memory of the event before the officer reviews the AI draft, material that should closely reflect the key elements that the generated report will include. This step is designed to preserve independent recall and reduce the chance that the AI-generated narrative reshapes the officer's memory of the incident.

Training should cover both product use and the legal and evidentiary stakes of AI-generated police report writing. Officers and supervisors should understand that the AI system is a large language model capable of hallucinations, omissions, and narrative distortions. Training should explain the role of the system prompt and prompt version, the implications of AI-generated content for Sixth Amendment and Brady obligations, and the agency's expectations for review, editing, and documentation. Supervisors should also be trained to interpret how often human correction of tool output is needed and identify patterns that may signal automation bias or system overreach.

Transparency, Accountability, and Privacy

The agency should tell the public, in plain language, where AI is used in report drafting, where it is not used, and what responsibilities remain with the officer. Public notice should make clear that the AI tool produces drafts and that officers remain accountable for the content of final reports. It should also explain the agency's retention, review, and error-correction practices.

The agency should designate a named official accountable for AI-influenced report quality

and for the prompt-and-draft retention regime. Accountability cannot rest solely with individual officers if the system design, prompt structure, model version, or vendor update affects the content officers review. A designated official should be responsible for ensuring that agency policy, vendor performance, audit trails, and legal obligations remain aligned.

The agency's discovery posture should treat the system prompt, initial AI draft, and final report as a set, all subject to discovery. This approach goes beyond CA SB 524 and responds directly to the accountability gap created when the prompt shapes the output but is not retained or disclosed. Treating prompt opacity as a structural finding rather than a vendor-implementation detail changes the procurement and implementation analysis. It makes prompt retention, prompt versioning, and prompt production in discovery core accountability features. The retention period should align with the longest applicable statute of limitations and discovery obligations, and tools should not erase initial drafts on export.[7] This approach should apply in all jurisdictions.

Pilot Program

The pilot should define scope, duration, success criteria, comparison methodology, and a go/no-go process before deployment begins. Because the AI system is substantial risk, Appendix G recommends a minimum six-month pilot. The agency could implement a longer 12-month period, given the lag between report writing and downstream prosecutorial outcomes.

When determining the scope of the pilot program, the agency should consider incident-type coverage, case-mix representativeness, and the ability to track downstream prosecutorial outcomes. Success criteria should include time savings compared to the baseline, hallucination or error rate, officer edit rate, supervisor rejection rate, prosecutor acceptance rate, defense motions targeting AI content, and demographic parity in output language. The comparison methodology should ideally use a randomized, within-officer or within-incident-type design, rather than a simple pre-/post-comparison.

The go/no-go process should be a formal one. At the end of the pilot, the agency should decide whether to terminate, modify, extend, or expand the deployment based on pre-set criteria. If the pilot reveals constitutional violations, unmitigable harms, unacceptable hallucination rates, or evidence that the system degrades report quality or legal integrity, the agency should terminate the contract rather than proceed to broader deployment.

Level 2 Enhanced Requirements For Substantial Risk Systems

Because the system is classified as substantial risk, the agency should apply additional rights protections, community engagement, evidence and validation requirements, auditing, and technical safeguards. Important additional safeguards include:

- Rights protection should begin with a formal legal analysis covering the Sixth Amendment, Brady, due process, and applicable state AI-disclosure laws, including Utah SB 180, CA SB 524, and successor statutes. The agency should also complete an assessment of human rights impacts with a mitigation plan. Defendants should have an individual challenge procedure that allows access to the prompt, initial draft, and final report, and provides them with the opportunity to present additional information.

- Community engagement should include facilitating a Community Advisory Committee, or another form of meaningful engagement with the community, with representation from directly impacted people. The agency should also provide multiple feedback channels and make particular efforts to include people directly affected by the criminal justice system. Regular public reporting should address usage, override rates, hallucination rates, and any incidents.

- Evidence and validation requirements should include independent validation by experts not affiliated with the vendor or agency. The agency should also conduct a formal alternatives analysis comparing AI-generated report writing with templates, dictation, the addition of clerical staff, and workflow redesign. That analysis should document why AI is preferable despite its risks.

- Auditing and enhanced oversight should include extended training on automation-bias resistance, continuing education requirements, whistleblower protections for officers who flag AI errors or pressure to approve drafts, real-time monitoring of edit rates and override patterns, demographic disparity tracking, and decision logs and audit trails accessible for legal review. Additional technical safeguards should include adaptive system monitoring with anomaly alerts, the ability to freeze deployment if concerns arise, and regular revalidation as underlying models are replaced.

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. Because the AI system in this case is substantial risk, scheduled reassessment should occur annually. The agency should monitor indicators that reveal both technical performance and legal-system impacts, reassessing the AI system when key product- or context-specific changes occur, including but not limited to:

- The underlying model changes
- The vendor materially changes the system prompt
- The product adds multimodal video capabilities
- Hallucination or override rates move outside agreed thresholds
- Bias findings emerge
- Defense motions target AI-generated content
- Brady disclosures arise
- New civil-rights claims and litigation outcomes surface

- New state laws change disclosure requirements

- New case law addresses the Sixth Amendment as applied to AI

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 police use of AI-generated report writing tools.* https://councilоncj.org/assessing-ai-in-action-a-case-study-on-police-use-of-ai-generated-report-writing-tools/

Endnotes

[1] Adams, I. T., Barter, M., McLean, K., Boehme, H., & Geary, I. A. (2024). No Man's Hand:

Artificial Intelligence Does Not Improve Police Report Writing Speed. *J Exp Criminol* 22, 137–154. https://doi.org/10.1007/s11292-024-09644-7

[2] Utah State Legislature. (2025). *S.B. 180, Law enforcement usage of artificial intelligence, 2025 General Session*. https://le.utah.gov/Session/2025/bills/introduced/SB0180.pdf

[3] California State Legislature. (2025). *S.B. 524, Law enforcement agencies: Artificial intelligence, 2025–2026 Regular Session*. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB524

[4] Constantino, M. (2025). Ribbit Ribbit! Artificial Intelligence Programs Used by Heber City Police Claim Officer Turned Into a Frog. FOX 13 Salt Lake City. https://www.fox13now.com/news/local-news/summit-county/how-utah-police-departments-are-using-ai-to-keep-streets-safer#google_vignette

[5] Harris, J. (2024). AI-Assisted Police Reports Not Welcome in King County Due to Potential Errors. KOMO News. https://komonews.com/amp/news/local/king-county-prosecutor-tells-police-not-to-use-ai-artificial-intelligence-for-official-reports-for-now-errors-concerns-law-enforcement-perjury-criminal-justice

[6] Prosecutor's Center for Excellence. (2025). Email from King County (WA) Prosecuting Attorney's Office Re Axon Draft One. https://pceinc.org/wp-content/uploads/2025/01/20240920-Email-to-Police-Chiefs-re-Axon-Draft-One-King-County-Prosecuting-Attorney-Dan-Clark.pdf

[7] This draft retention policy is a concrete implementation requirement responsive to the Electronic Frontier Foundation finding. See: Guariglia, M. & Maass, D. (2025). Axon's Draft One is Designed to Defy Transparency. Electronic Frontier Foundation, https://www.eff.org/deeplinks/2025/07/axons-draft-one-designed-defy-transparency