CCJ Council on Criminal Justice

Assessing AI for Criminal Justice: A User Decision Framework

March 2026

ABOUT THE COUNCIL ON CRIMINAL JUSTICE

The Council on Criminal Justice is an invitational membership organization and think tank. Independent and nonpartisan, the Council advances understanding of the criminal justice policy choices facing the nation and builds consensus for solutions that enhance safety and justice for all. The Council does not take policy positions. As part of its array of activities, the Council conducts research and convenes task forces composed of Council members who produce reports with findings and policy recommendations on matters of concern. Task force findings and conclusions are not subject to the approval of the Council's Board of Directors, its Board of Trustees, or funders. For more information about the Council, visit councilоncj.org.

ABOUT THE TASK FORCE ON ARTIFICIAL INTELLIGENCE

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. Spanning the four major sectors of the criminal justice system—law enforcement, courts, corrections, and community organizations—the group is producing credible analysis and guidance to help policymakers and practitioners navigate a complex and rapidly evolving landscape in ways that maximize benefits, minimize harms, and improve justice. Chaired by former Texas Supreme Court Chief Justice Nathan Hecht, the Task Force includes 14 other leaders representing AI technology developers and researchers, police executives and other criminal justice practitioners, civil rights advocates, community leaders, and formerly incarcerated people.

ACKNOWLEDGEMENTS

This framework from the Task Force on Artificial Intelligence is the product of its members, who graciously shared their time and expertise. Jesse Rothman produced this report, with support from Cameryn Farrow, Olivia McLarnan, Andrew Page, and others from the Council on Criminal Justice team. James Anderson and RAND serve as research partners to the Task Force. The Task Force is grateful to Kyle Moore for leading the external feedback process, as well as advisers Sorelle Friedler, Kevin Miller, Judge Scott U. Schlegel, Jonathan Wroblewski, and many others across the criminal justice field for providing invaluable guidance and insights during the production of this framework. Support for the Task Force on Artificial Intelligence comes from the Heising-Simons Foundation, The Just Trust, Microsoft, Southern Company Foundation, 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 for criminal justice: A user decision framework*. https://councilоncj.org/assessing-ai-for-criminal-justice-a-user-decision-framework/

Table of Contents

Introduction

Criminal justice agencies face urgent questions about the adoption of artificial intelligence (AI), especially concerning the usefulness and safety of existing and forthcoming tools. This framework addresses those challenges by extending AI governance principles into specific operational and ethical contexts of criminal justice practice, translating broad guidance into the detailed actions agencies and practitioners should take to navigate AI adoption responsibly.

Anchored in the Principles for the Use of AI in Criminal Justice produced by the Council on Criminal Justice Task Force on Artificial Intelligence in October 2025, this framework builds on those principles.

- Recognizing that systems should be safe and reliable, agencies should **require rigorous, independent validation** rather than relying solely on vendor claims, particularly for substantial-risk systems where errors could result in wrongful detention or public safety failures.
- Procurement serves as a critical safety net: Contracts should **establish enforceable performance standards, data rights, fairness requirements, auditability provisions, and termination rights** before any system is acquired to ensure confidentiality and security while handling sensitive criminal justice data.
- To make AI effective and helpful, multidisciplinary assessment teams—including legal, operational, technical, and community representatives—should **evaluate whether systems demonstrably outperform alternatives**, with ongoing monitoring and formal reassessments at least annually.
- Because AI should be fair and just, **regular assessment of impacts across demographic groups is essential, as is mandatory user training** that addresses automation bias and ensures operators understand system limitations.
- Upholding democratic and accountable deployment requires substantial human oversight. **Operators should retain clear authority to override AI-generated recommendations, and community input should be integrated from the outset** to ensure that those most affected by these systems help shape their adoption and governance.

While this framework offers guidance that is detailed enough to serve as an action plan, it is not intended to be rigid. Many stakeholders have needs and unique circumstances that warrant nuanced consideration of the recommendations. Users should take the liberties they need to adapt application of the framework to the capacity and limitations of their jurisdiction or organization.

Later in 2026, the Task Force plans to present a series of practical case studies that demonstrate this framework in action across different AI applications and agency contexts. These case studies will serve as implementation playbooks that agencies and communities can use to see how the framework may apply to specific tool categories.

Glossary

AI (Artificial Intelligence): Machine-based systems that operate with varying levels of autonomy, may exhibit adaptiveness after deployment, and infer from inputs how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

Algorithm: A set of rules or instructions to perform a task or solve a problem; in AI, algorithms process data to produce outputs.

Automation Bias: A tendency to over-rely on algorithmic outputs without sufficient critical evaluation.

Bias (Discriminatory): Unfair discrimination against people based on legally prohibited grounds such as race, gender, national origin, religion, or disability. Such discrimination can occur through disparate treatment or unjustified disparate impact.

Bias (Statistical): Systematic error that causes a model to consistently deviate from accuracy in a particular direction.

Black Box: An AI system whose internal workings are not visible or understandable to users or developers; decisions cannot be traced to specific rules or factors.

Classification Memo: The official document produced through Phase 2 that records an AI system's risk and opportunity assessment and recommended path forward.

Data Governance: Policies and procedures for managing data quality, security, privacy, and appropriate use throughout a system's lifecycle.

Demographic Performance: How an AI system performs across different population groups defined by characteristics such as race, gender, age, or socioeconomic status.

Disparate Impact / Discriminatory Effects: Discrimination that occurs when a facially neutral practice disproportionately harms people with a shared identity characteristic, such as race, gender, national origin, religion, or disability, without justification.

Disparate Treatment: Discrimination that occurs by intentionally treating people differently based on legally prohibited grounds such as race, gender, national origin, religion, or disability. (Contrast with disparate impact discrimination, which can be unintentional.)

Due Process: Constitutional requirement for fair legal procedures; AI must not undermine these protections.

Glossary (cont.)

Explainability: The degree to which an AI system's outputs can be explained in terms humans can understand.

Fairness Metric: A quantitative measure of whether an AI system treats different groups equitably; multiple definitions exist and may conflict.

Independent Validation: Testing of an AI system by experts not affiliated with the vendor or implementing agency.

Interpretability: The degree to which a human can understand the cause of an action taken or recommended by an AI system.

Level 1 Requirements: Baseline protections required for all AI systems (see Phase 4, Level 1 Implementation Requirements).

Level 2 Requirements: Enhanced protections required for substantial-risk systems (see Phase 4, Level 2 Enhanced Requirements).

Meaningful Human Oversight: Human review that features information access, sufficient time, training, override authority, documentation, and accountability.

Model Drift: Changes in AI performance over time due to shifts in data patterns.

Training Data: The historical data used to develop an AI system's predictive model; biases in training data can produce biased outputs.

Transparency: The availability of information about how an AI system works, what data it uses, and how decisions are made.

Vendor: A company or organization that sells or provides an AI system.

Overview

This framework walks stakeholders through sequential phases:

Phase 1: Defining the problem to be solved and assessing organizational readiness
Phase 2: Classifying the system's risk and opportunity levels
Phase 3: Establishing procurement protections
Phase 4: Implementing with appropriate safeguards
Phase 5: Conducting ongoing monitoring and reassessment

At the end of each phase, you'll reach a checkpoint, which encourages documented approval before advancement to help ensure that agencies make deliberate choices at every step.

The classification process at the framework's core first screens for prohibited uses, then categorizes remaining systems by risk level (low or substantial) and opportunity level (substantial or low). These classifications determine whether agencies should proceed with standard deployment, conduct careful implementation with enhanced safeguards, perform further evaluation, or avoid the system entirely.

Ten appendices provide the following tools to support implementation: readiness assessments, prohibited systems protocols, system complexity evaluations, sector-specific guidance, classification memo templates, procurement checklists, implementation planning guides, ongoing monitoring support, and guidance for deployed technology and general-purpose AI tools.

Taken together, these resources translate the Task Force's principles into:

- Constitutional and due process considerations tailored specifically to criminal justice applications;
- Concrete procurement and implementation steps that address the practical realities of agency operations; and
- Checklists and templates that can be adapted to jurisdictions' needs.

A Call for Critical Thinking

This framework provides a structured pathway for critical engagement with the evaluation, oversight, and use of AI in the criminal justice system. The questions, tables, and examples are meant to be illustrative guides, not inflexible decision trees. To enhance the value and validity of insights drawn from this framework, users should:

- Engage thoughtfully with difficult questions about fairness, bias, and constitutional compliance
- Challenge assumptions about what technology can and should do in justice settings
- Be prepared to say no if safeguards cannot be adequately implemented
- Leverage domain expertise, legal obligations, and contextual nuance throughout the process

User Guide

This framework is designed for use specifically with AI systems as opposed to other common forms of technology or software. The boundaries between AI and other technologies can be blurry and ambiguous. For this framework, "AI" refers to automated systems that generate predictions, recommendations, classifications, decisions, actions, or content that influence actions and decisions. It does not cover basic procedural technologies (e.g., spreadsheets, databases, standard word processing).

This framework assumes that you already have clarity on whether the tool in question should be classified as AI that could substantially influence decisions, or that you have the knowledge or external support necessary to make such a determination.

How to use this framework depends on your current engagement with potential or actual AI solutions:

- **If you are considering a specific AI system, start at Phase 1 and proceed sequentially through the framework** to evaluate that system's characteristics, risks, and opportunities before making procurement and implementation decisions.
- **If you already use an AI system,** review the full framework, then consult Appendix I, Guidance for Deployed AI Systems, to evaluate your tools and determine the proper course for ongoing management and oversight.
- **If you are conducting an open procurement process without a specific vendor in mind,** determine whether your agency should pursue this category of tool at all (using Phase 1, Phase 2, and Appendix D). If you decide to proceed, assess each finalist proposal using the risk and opportunity frameworks before making your selection. Complete a classification memo (Appendix E) for your chosen vendor before finalizing the contract.
- **If you are exploring whether AI might help with a problem but have no specific tools in mind**, begin with Phase 1 and the sector context guidance in Appendix D to understand how AI might intersect with your current practices. You should also consider whether the problems you're trying to solve might be better addressed through policy changes, training, or increased resources, rather than through technology. Proceed to exploring specific tools if your workplace has built a strong foundation for responsible AI use.

Developing Policies for General-Purpose AI Tools in Criminal Justice Settings

This framework is primarily designed for evaluating purpose-built AI systems acquired through formal procurement. However, AI increasingly enters criminal justice settings through a different pathway: staff use of general-purpose AI tools such as AI chatbots, agents, coding tools, and document analysis systems that are not purchased specifically for criminal justice work but are used in case-related contexts.

These general-purpose tools present distinct governance challenges. They are often adopted informally, without IT oversight or contractual protections. They may process sensitive case data through external servers. Their capabilities change frequently as providers update their models. And because they are intended for general-purpose use, they can be applied to an open-ended range of tasks without any single procurement decision triggering review.

Agencies should not ignore this reality. Instead, they should develop a policy governing staff use of general-purpose AI tools for case-related work. See Appendix J for more on this problem and the Task Force's recommended action.

User Profiles

The following user profiles offer illustrative, though not exhaustive, overviews of stakeholder groups that may find this framework useful, as well as tailored procedural guidance for each group:

User Group	How to Use This Framework
Agency Leaders (chiefs, directors, sheriffs, court administrators, corrections commissioners)	Focus on Phase 1 (foundation and readiness), Phase 2 (understand classification decisions), and the Introduction (principles). You approve progression past Phase 1, accept classification memos, authorize procurement for substantial-risk systems, and approve deployment after pilots.
Procurement & Legal Officials (general counsel, procurement directors, contract officers, county attorneys)	Focus on Phase 3 and Appendix F (contract protections), plus the prohibited use screening in Phase 2. You approve contract language, certify legal compliance, advise on constitutional concerns, and can recommend rejection based on legal risk.
Project Managers & IT Staff (IT directors, system administrators, data officers)	Focus on Phase 4 (implementation), Phase 5 (ongoing management), and Appendix G (implementation template). You recommend technical feasibility, approve integration plans, certify training completion, flag technical concerns during pilots, and recommend continuation or termination based on performance.
Policymakers (legislators, council members, commission staff, oversight bodies)	Review the full framework to inform legislation and oversight. Focus on Phase 2 (risk categories for regulatory frameworks) and the Introduction (specific criminal justice AI governance). You set mandatory requirements, establish reporting and oversight mechanisms, and allocate resources for AI governance.
Community Representatives & Advocates (public defenders, civil rights organizations, community advisory members, crime survivors, directly impacted individuals)	Focus on the prohibited use screening (Phase 2) and community engagement requirements (Phase 4, Level 2). You raise concerns through advisory processes, provide input on risk and opportunity assessments, advocate for specific safeguards, and escalate rights violations.
AI Developers & Vendors (technology companies, product designers, AI researchers, vendors seeking to develop or sell AI tools for criminal justice settings)	Review the sections—particularly classification (Phase 2), procurement (Phase 3), and implementation (Phase 4)—that can help you anticipate the questions, safeguards, and documentation stakeholders may expect before adopting an AI system. Understanding these expectations may help you design tools and documentation that better align with criminal justice system priorities around validation, transparency, fairness, and meaningful human oversight.

Questions for Future Work

This framework provides a structured pathway for responsible AI adoption in criminal justice, but frameworks alone are not sufficient to guarantee good outcomes. Important questions remain about what infrastructure is needed to make the full recommendations embedded in this framework accessible and considerate of the ways that AI uses may evolve. The Task Force believes the following institutional design questions highlight areas for potential future work by this body, successor entities, policymakers, or other stakeholders:

- **Agency capacity**: What minimum internal resources and expertise should agencies possess before pursuing AI adoption? How can smaller agencies access independent technical and legal expertise?
- **Bundled support**: Should professional associations, states, or regions establish centralized review bodies, shared services, pre-approved vendor lists, or pooled technical assistance to reduce the burden on individual agencies?
- **Federal role**: What guidance, standards, or grant incentives from federal agencies would support responsible local adoption of AI technology?
- **Technical assistance**: Who should provide implementation support—state agencies, academic institutions, nonprofits, or professional associations? And how should it be structured?
- **Accountability mechanisms**: What type and scope of external oversight from courts, legislatures, or civil society can reinforce these best practices?
- **Future AI capabilities:** How should criminal justice institutions and stakeholders prepare for a possible future in which AI becomes a general-purpose capability that matches or exceeds human performance across a wide range of cognitive work and professional functions?

ASSESSMENT WORKFLOW

Phase 1: Foundation and Readiness

1. Define the Problem

Before evaluating any AI tool, you should clearly define the problem you're trying to solve. Technology should not be a solution looking for a problem. Complete this exercise first:

- **Problem**: *What specific criminal justice problem are we trying to solve?*
    - Be specific and measurable
    - Who experiences this problem? How does it affect them?
    - How long has this problem existed? What solutions have been tried already?
- **Theory of Change:** *How exactly would this AI tool solve the problem better than alternatives?*
    - What is the causal mechanism by which AI improves outcomes?
    - Why wouldn't a non-AI solution work as well?
    - What assumptions must be true for the AI solution to work?
- **Success Metrics**: *How will we measure whether the problem is actually solved or improved?*
    - What data will we track?
    - What magnitude of improvement would justify the investment?
    - Over what timeframe will we evaluate success?

2. Assess Organizational Readiness

Adopting AI is a policy decision with implications for justice, safety, and public trust. Before proceeding, use the **AI Readiness Assessment Worksheet (see Appendix A)** to evaluate your organization's capacity in areas like data governance, technical expertise, legal frameworks, and community engagement. If this assessment reveals significant gaps, they should be addressed before proceeding with AI deployment.

Phase 1 Completion Checklist

Requirement	Complete?
Problem Definition	
Problem statement is documented in specific, measurable terms	☐
Theory of change explains why AI may be preferable to alternatives	☐
Success metrics are defined and measurable	☐
Relevant stakeholders have reviewed the problem scoping	☐
Organizational Readiness	
AI Readiness Assessment (Appendix A) is complete	☐
Adequate capacity is confirmed, or remediation plans are prepared for capacity gaps	☐
Leadership has reviewed and acknowledged readiness status	☐

Phase 1 Checkpoint

If you cannot clearly articulate the problem or why AI is preferable to alternatives, or if significant capacity gaps exist without clear remediation plans, **PAUSE**. Revisit the problem definition, consider non-AI alternatives, or build foundational capacity before proceeding.

Phase 2: Classification

The goal of this phase is to produce a **Classification Memo (see Appendix E)** that analyzes the AI system and recommends a path forward.

1. Assemble Your Assessment Team

Assessing the risk and opportunity associated with an AI use case requires a well-rounded set of expertise. To help increase the likelihood of accurate assessments, teams should include the experts listed below.

- **Recommended for All Systems:** An operational leader, a legal/constitutional expert, and end-users. Establish clear rules for making decisions.
- **Add for Substantial-Risk Systems:** A sector specialist, community representatives, and (for complex systems) a technical expert.

2. Screen for Prohibited Uses

Some AI applications pose an unacceptable risk to fundamental rights and should be prohibited in the criminal justice context **unless the risks can be adequately mitigated or the problematic features can be eliminated**. Answer the screening questions below.

- **If you answer YES to ANY question,** the system raises serious concerns that should be resolved before deployment. If adequate mitigation is not possible, the system is prohibited. Proceed directly to Appendix B: Protocol for Prohibited Systems.
- **If you answer NO to all questions**, proceed to the next step.

Prohibited Use Screener

Question	Yes	No
Does the system make autonomous decisions about liberty (e.g., detention, sentencing) without the possibility of substantial human review?	☐	☐
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?	☐	☐
Does the system circumvent or undermine established legal or constitutional protections (e.g., due process, equal protection)?	☐	☐
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)?	☐	☐
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?	☐	☐
Does the system systematically undermine human dignity (e.g., by publicly shaming or humiliating people or stripping them of all agency)?	☐	☐

If you are uncertain how to answer these questions, you should:

1. Request documentation from the vendor. Vendors should be able to answer all of these questions clearly.
2. Consult outside legal and technical experts for advice on these questions

Assessing Mitigation Possibilities

If you answered "yes" to any question, consider whether the concern can be addressed through:

- Design modifications that eliminate the problematic feature
- Procedural safeguards that adequately protect rights
- Technical controls that prevent the harmful application
- Alternative deployment that avoids the prohibited use

Only proceed with deployment if you can document that the risk has been fully mitigated or the problematic feature eliminated. **If mitigation is not possible or adequate, the system should be considered prohibited**.

3. Assess System Complexity and Interpretability

Next, you should evaluate the system's technical characteristics. More complex, opaque ("black box") systems require more scrutiny and more robust safeguards.

Use the questions in the **System Complexity and Interpretability Assessment (see Appendix C)** to determine if the system is transparent and predictable or opaque and ambiguous. Document your findings; this work will inform the oversight required for implementation.

Note on AI system type: *Your System Complexity and Interpretability Assessment should influence the risk assessment outlined below in step five. A system that is difficult to interpret or validate may warrant a higher risk classification, as errors may be harder to detect or mitigate.*

4. Consider the Sector Context

Before you assign a general risk or opportunity score, analyze how the AI tool will change your **current practices**. An AI system does not exist in a vacuum; its impact is relative to the baseline with which it interacts. A high-stakes context doesn't automatically mean AI increases risk. Rather, the core question is whether AI makes the existing process more or less risky, fair, and effective.

Consider the legal and operational context of your specific sector. Review the **Sector Context Guidance (see Appendix D)** to help you frame your thinking for the next steps.

5. Determine Risk and Opportunity Levels

With the sector context in mind, use the following tables to classify the system's risk and opportunity levels.

Note on classification: *The risk and opportunity levels offered in this section are not meant to imply that all AI systems and use cases fall cleanly into a binary. The "low" and "substantial" classification designations are designed to emphasize selectivity when determining risk and opportunity. For example, on a scale of 1 to 10, a substantial risk classification could correlate with levels 4 through 10 (as opposed to the traditional 6 through 10 in a true binary). This might encourage stakeholders to use the enhanced safeguards outlined in this framework even with AI systems and uses that may seem only moderately risky.*

Risk Assessment

Use this table to determine if the risk level is **low** or **substantial**.

Risk Level	Liberty Impact	Rights Impact	Error Consequences
LOW	Unlikely. No direct effect on an individual's liberty.	Unlikely. Does not affect procedural or substantive legal rights.	Errors cause minimal harm and are easily corrected.
SUBSTANTIAL	Moderate-High. Affects or influences stop, search, arrest, detention, bail, charging, plea, sentencing, parole, clemency, or similar decisions.	Affects procedural or substantive legal rights. Involves surveillance or processes sensitive personal data.	Significant harm possible. Errors could lead to wrongful detention or rights violations.

Risk Classification Questions

If the answer is **YES** to any of the following, the system is likely **SUBSTANTIAL RISK**:

- Does it influence stop, search, arrest, pretrial release or detention, charging, plea, sentencing, parole, clemency, or similar decisions?
- Does it implicate legal rights, such as freedom of speech, the right to be free from unreasonable searches or seizures, the right against self-incrimination, the right to counsel, the right to confront witnesses, the right to a fair and impartial jury, or the right to be free from discrimination?
- Does it involve the surveillance or monitoring of individuals?
- Does it process sensitive personal data?
- Could it create an unjustified disparate impact?
- Does it directly affect access to programs, services, or due process?
- Could errors result in wrongful detention or rights violations?
- Does it limit the ability for people to contest decisions?

If the answer is **NO** to all of the above, the system is likely **LOW RISK**.

Opportunity Assessment

Use this table to classify the potential for positive impact as **substantial or low.**

Opportunity Level	Performance Improvement	Evidence Quality	Stakeholder Support	Cost-Effectiveness
SUBSTANTIAL	Demonstrable improvement over existing processes.	Supported by evidence from pilots or independent research.	Community and end-users validate the value.	Favorable.
LOW	Minimal, uncertain, or no improvement.	Little or no supporting evidence; claims are speculative.	Stakeholders skeptical of necessity.	Does not favor AI.

The four factors key to evaluating the potential for positive impact will not always align. When they conflict, consider:

- **Evidence quality.** Strong performance claims mean less without credible validation. Be skeptical of promised improvements that lack independent evidence.
- **Stakeholder concerns warrant serious weight.** Opposition from affected communities or end-users can predict implementation problems, even when other factors look favorable.
- **Improvement should benefit those affected.** Efficiency gains that accrue to the organization while individuals bear the risks (errors, bias, privacy loss) represent weaker opportunity than improvements in actual outcomes.
- **Compare to alternatives, not to nothing.** The relevant question is whether AI outperforms the best alternative use of the same resources.

When factors point in different directions, use your team's collective judgment to determine the final opportunity score. In your Classification Memo, document which factors you weighted most heavily and why.

6. Finalize and Document Classification Decision

The goal of this phase is to complete and file the Classification Memo that summarizes your assessment and establishes an official record of your findings.

Use your risk and opportunity levels to find your position on this matrix.

	Substantial Opportunity	Low Opportunity
Substantial Risk	**CAREFUL IMPLEMENTATION** Potential value, but also significant risks. Recommended action: Proceed only by applying ALL Level 1 AND Level 2 requirements (details in Phase 4). Agency head or designated authority should provide written approval before deployment, documenting that all safeguards are in place.	**GENERALLY AVOID** Strong presumption against implementation. High risks are not justified by low benefits. Recommended action: Do not proceed without considering non-AI alternatives.
Low Risk	**STANDARD DEPLOY** These systems offer clear benefits with lesser risk. Recommended action: Proceed with Level 1 requirements (details in Phase 4).	**EVALUATE** The benefits are unclear and may not be worth the investment. Recommended action: Conduct a careful cost-benefit analysis. If proceeding, use Level 1 requirements (details in Phase 4).

Phase 2 Completion Checklist

Before proceeding to Phase 3, confirm the following:

Requirement	Complete?
Assessment team properly constituted	☐
Prohibited use screening complete (all NO)	☐
System complexity assessment complete (Appendix C)	☐
Sector context considered (Appendix D)	☐
Risk level determined and documented	☐
Opportunity level determined and documented	☐
Classification Memo (Appendix E) complete	☐
Classification Memo has required approval	☐

Phase 2 Checkpoint

- If classification is **GENERALLY AVOID**: Should not proceed without completing alternatives assessment and documented justification for proceeding despite low opportunity.
- If classification is **EVALUATE**: Should proceed only after completing rigorous cost-benefit analysis that justifies investment.
- If classification is **STANDARD DEPLOY** or **CAREFUL IMPLEMENTATION**: Proceed to Phase 3.

Required approval: The Classification Memo should be approved by the designated authority before procurement begins. For substantial-risk systems, this should be the agency head or a person of equivalent authority.

Phase 3: Procurement

The procurement phase establishes the contractual foundation that protects your agency, ensures accountability, and maintains compliance throughout the system's lifecycle. If your recommendation is to proceed, you should now navigate this process mindful of the appropriate requirements based on your system's classification.

The following foundational steps should be completed:

1. Budget and Resource Confirmation

- Plan for resources needed throughout the system's lifecycle for:
  - Acquisition
  - Initial implementation and integration
  - Staff training and change management
  - Ongoing monitoring and oversight
  - Technical fixes and improvements
  - Community engagement processes (for substantial-risk systems)

2. Designate Personnel

- Designate a procurement lead with appropriate authority
- Include representatives from:
  - Legal/general counsel
  - End-user departments
  - IT/technical staff
  - Finance/budget office
  - For substantial-risk systems, add: community representatives or liaison, independent technical expert (if system is complex)

3. Contract Negotiation and Essential Terms

Contract negotiation should secure protections and requirements based on your system's classification. Use **Procurement Guide (see Appendix F)** as your checklist and complete all necessary steps before moving to Phase 4.

Phase 3 Completion Checklist

Before proceeding to Phase 4, confirm the following:

Requirement	Complete?
Resources considered for full lifecycle (not just acquisition)	☐
Procurement team properly constituted	☐
Contract includes all required Level 1 terms (Appendix F)	☐
If substantial risk, contract includes all Level 2 terms	☐
Contract fully executed	☐
If substantial risk, community engagement plan developed	☐
If substantial risk, independent oversight established	☐

Phase 3 Checkpoint

Should not begin implementation until the contract is fully executed with all required protections. A contract missing essential terms creates unacceptable risk. If the vendor will not agree to the required terms, you should return to market or reconsider whether AI is appropriate for this use case.

Phase 4: Implementation

Implementation is where theoretical safeguards and contractual promises become operational reality. This phase of the framework encourages you to pay careful attention to how the system will function in your environment and how you'll ensure it performs as intended. Before any system goes live, completion of the **Implementation Memorandum Template (see Appendix G)** is recommended. This document serves as your operational blueprint, translating classification decisions and contract terms into concrete implementation steps.

Level 1 Implementation Requirements (For All Systems)

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

Human-Centered Design

The deployment of AI tools should account for the realities of human work and organizational design, and should allow human supervisors to retain ultimate authority. This requires deliberate design choices about how the system presents information and how override decisions are captured and reviewed.

- Design interfaces to present AI outputs as recommendations and show operators the relevant information behind each recommendation.
- Document every override decision along with the operator's rationale.

Training and Capacity Building

Before anyone operates the AI system, they should receive training that covers system functionality, limitations, known failure modes, and automation bias.

- Plan for regular refresher training, updates whenever the system changes, and sessions incorporating findings from ongoing monitoring.
- Keep detailed records of who has been trained, maintain current training materials, and evaluate training effectiveness.

Creating Transparency and Accountability

Members of the public should know when AI systems affect them. Develop clear public notification materials in plain language explaining what the system does, how it's used, how much it costs to operate and maintain, and who is accountable for system-related decisions.

- Internally, designate specific officials who are publicly accountable for the system.
- Document decision-making processes and override rationales.
- Publish regular reports on system performance, including how often operators override the system and why.

Protecting Privacy and Data Security

Collect and process only what is necessary for the system's legitimate purpose. Regular reviews of these requirements should be conducted to prevent unnecessary accumulation. Data collected for one use should not be repurposed without new assessment and approval.

- Implement technical safeguards per your contract, including regular security assessments, strict access controls, and monitoring for unauthorized access.
- Establish and follow clear retention policies with automated deletion where appropriate.

Pilot Program

Before it is deployed at full scale, the system should be tested in your actual operational environment. Design a pilot with a controlled scope and clearly defined success criteria before you begin. Low-risk systems may proceed with pilot duration at agency discretion, though sufficient time should be allowed to substantially evaluate success criteria. Compare the AI system's performance to your baseline practice and include diverse cases representing the full range of scenarios you'll encounter.

- During the pilot, systematically track your success metrics while remaining alert for unexpected behaviors. Gather user feedback and document issues and resolutions.
- Evaluate results against your success criteria: Did the AI improve on current practice? Are there rights concerns or disparate impacts? How well does it fit into operations? Is it cost-effective?
- Make a formal go/no-go decision. If the pilot reveals constitutional violations or unmitigable harms, you cannot proceed and should shift to contract termination procedures. If moving forward, incorporate lessons learned.

Ongoing Monitoring and Oversight

System performance should be continuously monitored; watch for degradation over time and compare results to baseline and pilot performance. Track error rates and types to identify concerning patterns.

- Pay attention to performance across demographic groups, investigating any differences that emerge.
- Establish procedures for identifying problems, addressing them, and documenting issues and resolutions.
- Share what you've learned across your organization.

Additional Considerations for High-Complexity Systems

If Phase 2 identified your system as highly complex or opaque, enhanced safeguards should be added.

- Require vendor-provided explanation mechanisms that generate case-specific rationales in plain language.
- Require vendor training, repairs, and updates for the life of the contract.
- Conduct more extensive pre-deployment testing and obtain independent technical validation.
- Continue validation during operation, checking regularly for model drift or unexpected behaviors.

Level 2 Enhanced Requirements (For Substantial-Risk Systems)

Implementation of systems classified as substantial-risk should follow everything outlined in Level 1 and include a set of enhanced protections detailed below.

Strengthening Rights Protection

Conduct a formal legal analysis that reviews due process implications, equal protection and other discrimination concerns, and surveillance or privacy impacts. Complete a thorough human rights impact assessment, identifying potential impacts and developing mitigation strategies. Ensure individual rights remain fully preserved through clear procedures that allow people to:

- Challenge AI-influenced decisions
- Access meaningful information about how AI was used in their case
- Present additional or new information

Community Engagement

- Establish a Community Advisory Committee with representatives from affected communities who have genuine authority to raise concerns and recommend changes. Consider appropriate compensation for committee members.
- Maintain ongoing communication through regular public reports and accessible forums that address questions and concerns.
- Document how community concerns are addressed, incorporate feedback into improvements, and report back to the community on actions taken.
- Acknowledge historical injustices and provide multiple accessible feedback channels. Within operational realities, seek input before decisions are finalized, and make a particular effort to include crime victims and people directly affected by the criminal justice system, wherever possible.

Demanding Evidence, Validation, and Disclosure

- Require independent validation by experts not affiliated with your vendor or agency. They should test on representative data, compare the system to alternative approaches, and publicly share results (with appropriate confidentiality protections).
- Conduct a formal alternatives analysis examining non-AI approaches and documenting why AI is preferable despite its risks.

Auditing and Enhanced Oversight

- Extend training on system limitations and automation bias resistance. Establish continuing education requirements.
- Protect staff through clear whistleblower safeguards, formal channels for reporting concerns, and investigation protocols that take reports seriously.
- Implement real-time analysis that continuously monitors performance. Establish protocols for triggering investigation and for taking corrective action.
- Maintain documentation, including detailed decision logs, complete audit trails for all system use, and records of every override that include rationale. Ensure records are accessible for legal review and appeals.

Additional Technical Safeguards for High-Complexity Systems

- When substantial risk combines with high complexity, implement adaptive system monitoring that continuously watches for unexpected behaviors and produces immediate alerts for anomalies.
- Ensure you can freeze deployment if concerns arise, and plan for regular revalidation as the system adapts.

Phase 4 Completion Checklist

Before transitioning to ongoing management, confirm the following:

Requirement	Complete?
Implementation Memorandum (Appendix G) is complete	☐
All Level 1 safeguards are operational	☐
If Substantial Risk, all Level 2 safeguards are operational	☐
All required training completed and documented	☐
Pilot program completed with documented results	☐
Formal go/no-go decision made based on pilot	☐
Monitoring systems operational and producing data	☐
Public notification completed	☐
System performing as expected	☐
No unmitigated constitutional or rights concerns	☐

Phase 4 Checkpoint

- If pilot reveals constitutional violations or unmitigable harms: **DO NOT DEPLOY**. Follow contract termination procedures.
- If pilot reveals concerns that can be mitigated: Address concerns, document mitigations, and obtain approval before deployment.
- If pilot is successful: Proceed to full deployment with all safeguards active.

Phase 5: Ongoing Management and Reassessment

Deployment is not the end of the process. All systems need ongoing monitoring and periodic reassessment to ensure they continue to function as intended without causing undue negative outcomes.

- **Scheduled Reassessment:** Substantial-risk systems should be fully reassessed annually, and low-risk systems should be reassessed before contract renewal.
- **Triggered Reassessment:** A new, full assessment should be conducted immediately if there are significant capabilities changes or major system updates, the system is applied to a new use case, performance issues arise, better alternatives become available, or new rights concerns surface.

See **Appendix H: Ongoing Monitoring and Reassessment** for detailed protocols.

ASSESSMENT TOOLS

CCJ Council on Criminal Justice

Appendix A: AI Readiness Assessment

Use this worksheet to help evaluate your organization's foundational capacity before beginning an assessment of a specific AI tool.

Area	Question	Assessment (Low/Medium/ High Readiness)	Notes / Action Items
1. Data Governance	Do we have clear, enforced policies for data quality, collection, storage, and sharing?		
	Is our data accurate for the intended AI application?		
2. Staff Capacity	How many FTE can be allocated to ongoing AI management and oversight?		
	Who will be responsible for tracking AI system performance over time?		
	Is there capacity to train staff on AI limitations and proper use?		
3. Technical Capacity	Do we have internal personnel with expertise to evaluate, implement, and monitor an AI system? If not, have we budgeted for external technical consultants?		
	Is our IT infrastructure sufficient to support the proposed AI system securely?		
4. Legal and Policy	Do we have an internal policy or legal framework governing AI use?		
	Have we consulted legal counsel on the constitutional and statutory implications of using AI?		

	Is there capacity for ongoing legal review during implementation, monitoring, and updates?		
5. Organizational Culture	Is there leadership buy-in for responsible AI adoption?		
	Have we trained relevant staff on AI literacy and critical thinking?		
	Are staff already using AI tools informally to address problems or handle tasks? Has the scope of informal AI use created risks (e.g., sensitive data entered into external tools, AI-generated content used without review) or benefits that should be addressed?		
6. Community Engagement	Do we have established channels for engaging with the community on AI-related issues? If not, how will we foster engagement and who is responsible?		
	Is there a foundation of trust with the communities we serve that would support AI adoption?		
7. Cumulative Review	**After using this worksheet, assess your overall readiness to adopt the tool under consideration.**		

Appendix B: Protocol for Prohibited Systems

DO NOT PROCURE OR IMPLEMENT

- Halt any procurement process immediately.

DOCUMENT THE PROHIBITION

- Complete the Classification Memo.
- Clearly state which prohibited indicators applied.
- Explain why the violations could not be mitigated.

EXPLORE ALTERNATIVE SOLUTIONS

- Revisit the original problem the AI system was intended to solve.
- Explore alternative approaches and non-AI solutions that respect constitutional boundaries.
- Consult stakeholders to develop preferable alternatives.

ORGANIZATIONAL LEARNING

- Share lessons learned with the larger organization to prevent future proposals for prohibited systems.
- Share information about the prohibited application with peers in your sector.

Appendix C: System Complexity and Interpretability Assessment

AI systems vary in their complexity, transparency, and the difficulty of validating their behavior and impact. The characteristics highlighted here help determine how much scrutiny, testing, and monitoring are needed.

Dimensions to Assess

1. Decision Transparency

- **More Transparent:** The system follows clear, documented rules or formulas that can be traced step-by-step. Operators can explain how and why the system produced a specific output.
- **Less Transparent:** The system uses methods that are opaque or difficult to understand. The exact reasoning path may be unknown or not intuitive to human operators.

2. Predictability and Consistency

- **More Predictable:** Given the same inputs, the system will always produce the same output. Its behavior is stable and well-understood.
- **Less Predictable:** The system may produce novel or unexpected results. It might learn and adapt over time or exhibit behaviors that surprise operators.

3. Validation Difficulty

- **Easier to Validate:** The system's performance can be tested in a straightforward fashion. Errors are easy to identify, and the system's accuracy can be verified through standard quality assurance processes.
- **Harder to Validate:** The system's behavior is difficult to test comprehensively. It may process complex, ambiguous inputs (like images or natural language) where "correct" outputs are subjective or context-dependent. Unexpected behaviors may only emerge after deployment.

4. Adaptability

- **Static Systems:** The system's logic only changes when humans manually reprogram it. Its rules are fixed.
- **Adaptive Systems:** The system may learn from new data and change its behavior over time. This adaptability creates uncertainty about future performance.

Assessment Questions

Answer these questions to help understand your system's complexity profile:

1. Can you trace and explain each step the system takes to reach its outputs?

   ☐ Yes, completely → lower complexity ☐ Partially

   ☐ No or with great difficulty → higher complexity

2. Does the system always produce the same output for the same input?

   ☐ Yes, completely → lower complexity ☐ Partially

   ☐ No or with great difficulty → higher complexity

3. Does the system use only structured, well-defined data inputs (not images, free text, or audio)?

   ☐ Yes, completely → lower complexity ☐ Partially

   ☐ No or with great difficulty → higher complexity

4. Are you confident you can fully validate the system's performance before deployment?

   ☐ Yes, completely → lower complexity ☐ Partially

   ☐ No or with great difficulty → higher complexity

5. Is the system's logic fixed and static (it won't learn, adapt, or develop new behaviors over time)?

   ☐ Yes, completely → lower complexity ☐ Partially

   ☐ No or with great difficulty → higher complexity

How to Use This Assessment

Systems with **lower complexity** (transparent, predictable, easy to validate, static) should be subject to standard oversight approaches:

- Audit the rules and logic
- Verify calculations and outputs
- Use standard quality assurance processes

Systems with **higher complexity** (opaque, ambiguous, hard to validate, processing unstructured data) should be subject to enhanced oversight:

- Demand explainability capabilities and case-specific decision rationales
- Require independent technical validation
- Plan for ongoing performance validation post-deployment

Document your findings: Note in your **Classification Memo** the system's complexity profile and what additional scrutiny or safeguards this triggers.

CCJ Council on Criminal Justice

Appendix D: Sector Context Guidance

Before classifying any AI system, you should establish your **current practice baseline** and assess how AI changes it.

Step 1: Document your current practice baseline

How are decisions currently made without AI? What are current outcomes, error rates, and fairness levels? What safeguards exist in current processes?

Step 2: Assess marginal change

Does AI add, maintain, or reduce risks compared to this baseline? Does AI create opportunities that improve on current outcomes?

Step 3: Document your reasoning

Record which sector-specific factors influenced your classification. Explain specifically how AI compares to current practice in your Classification Memo.

The key question is about marginal change: How does the AI system change risk and opportunity compared to current practices? A high-stakes decision context doesn't automatically mean AI increases risk. What matters is whether AI makes the existing process *more or less* risky, fair, and effective compared to current practice.

Sector-Specific Factors

LAW ENFORCEMENT

Risk Factors

AI May Increase Risk	AI May Maintain or Reduce Risk
Enables new surveillance scope or data aggregation	Replaces more biased decision-making with more objective criteria
Influences stops, searches, or arrests with less review	Adds accountability and transparency
Lacks safeguards present in existing processes	Operates in administrative functions not affecting constitutional practices
	Includes stronger safeguards than exist currently
	Has community support as an improvement

Opportunity Factors

High Opportunity	Low Opportunity
Independent evidence of better performance, including significant efficiency gains	Current methods already work well
Reduction in disparate impacts	Offers only marginal efficiency gains
Improved outcomes (public safety, community cooperation, constitutional compliance)	Problems could be better addressed through other changes
Community validation that it addresses real problems	

COURTS

Risk Factors

AI May Increase Risk	AI May Maintain or Reduce Risk
Reduces transparency in reasoning	Makes implicit biases explicit and measurable
Generates recommendations where judges currently exercise fuller discretion	Provides better information than currently available
Creates information asymmetries favoring one party	Improves adversarial balance
Less contestable by defense	Reduces unwarranted disparities through better information

Opportunity Factors

High Opportunity	Low Opportunity
Evidence of better decision quality	Current practice already produces clear, well-reasoned decisions
Evidence of more efficient operations	Claimed benefits aren't validated against baseline
Reduction in unwarranted disparities	Problems could be better addressed through judicial training or policy reforms
Enhanced individualized consideration	
Preserved or improved due process	
Validation showing improvement across demographic groups	

CORRECTIONS

Risk Factors

AI May Increase Risk	AI May Maintain or Reduce Risk
Reduces transparency or contestability in classification	Makes institutional classifications or decisions more auditable
Restricts programming access beyond current limitations	Identifies service needs better than informal processes
Reduces rehabilitation focus	Improves program matching
	Includes stronger protections for vulnerable populations

Opportunity Factors

High Opportunity	Low Opportunity
Evidence of better reentry outcomes	Current practice provides adequate program matching
Improved program matching	Benefits accrue to management without enhancing rehabilitation
Enhanced identification of service needs	Problems could be better addressed through policy changes
More access to information for incarcerated people and/or community members	Claims aren't validated against baseline
Demonstrated benefits for incarcerated individuals	
Significant efficiency gains	

COMMUNITY ORGANIZATIONS

Risk Factors

AI May Increase Risk	AI May Maintain or Reduce Risk
Introduces monitoring beyond trust-based relationships	Improves service matching
Creates service access barriers	Identifies needs currently missed
Shares data with justice agencies beyond current practice	Reduces bias in service decisions
Reduces human interaction	Enhances rather than replaces relationships
Makes participation less voluntary	Maintains or exceeds privacy protections
Serves vulnerable populations with fewer protections	Increases transparency
Damages trust or advocacy reputation	Increases individual agency and choice

Opportunity Factors

High Opportunity	Low Opportunity
Evidence of better individual outcomes	Current practice already provides effective service matching
Increased access to services	AI does not improve individual outcomes
Improved operational efficiency	Staff capacity increases could better address existing limitations
Confirmation of enhanced relationship-building	Current trust-based approach is already effective
Evidence of more effective advocacy	
Demonstration it responds to community-identified needs	

CCJ Council on Criminal Justice

Appendix E: Classification Memorandum Template

TO: [Recipient Name/Department]

FROM: [Assessment Team Lead Name/Title]

DATE: [Assessment Date]

RE: AI System Classification Assessment - [System Name]

SYSTEM INFORMATION

System Name: _____

Vendor/Developer: _____

Assessment Team Lead: _____

Assessment Team Members: _____

Assessment Period: _____ to _____

CLASSIFICATION RESULTS

Prohibited Use Screening: ☐ Passed (proceed to risk assessment)

☐ Failed (system is prohibited)

Risk Level: ☐ Low ☐ Substantial ☐ N/A (if Prohibited)

Opportunity Level: ☐ Low ☐ Substantial

Classification Level: ☐ Standard Deploy ☐ Careful Implementation

☐ Evaluate ☐ Generally Avoid

ASSESSMENT RATIONALE

System Complexity and Interpretability

Using the assessment questions in Appendix C, summarize the system's complexity profile across the four dimensions listed there. Note whether the system is lower or higher complexity overall and what additional scrutiny or safeguards this triggers.

Risk Assessment

Using the Sector Context Guidance (Appendix D) and Risk Assessment table and Risk Classification Questions from Phase 2, Section 5, document your assessment of how AI changes risk compared to your current practices for the following:

- Liberty Impact: Does the system affect or influence arrest, detention, bail, charging, sentencing, or release decisions?
- Rights Impact: Does the system affect procedural rights or substantive legal rights? Does it involve surveillance or process sensitive personal data?
- Error Consequences: Could errors lead to wrongful detention, rights violations, or other significant harm?
- Additional risk factors (disparate impact on protected groups, limited ability for people to contest decisions, data privacy, cybersecurity concerns, etc.)

Include specific examples or scenarios in your summary.

Opportunity Assessment

Using the Sector Specific Guidance (Appendix D) and Opportunity Assessment table from Phase 2, Section 5, document your assessment of how AI changes the opportunities across each of the following factors:

- Performance Improvement: Does the system demonstrate improvement over existing processes?
- Evidence Quality: Is there evidence from pilots or independent research, or are claims based on assumptions?
- Stakeholder Support: Do community members and end-users validate the value and helpfulness of the system?
- Cost-Effectiveness: Does a cost-benefit comparison favor AI over alternatives?

Note which factors you weighted most heavily and why. Quantify benefits where possible.

Core Findings

Summarize the most critical factors that drove the classification decision, including any concerns, limitations, or conditions that should be considered.

Stakeholder Input

Document any stakeholder consultations, community feedback, or input from subject matter experts that informed this assessment.

RECOMMENDED ACTION:

☐Approve for Implementation ☐Approve with Conditions ☐Reject

Implementation Recommendation Details

Provide specific recommendations for next steps, including any conditions, safeguards, or modifications required before implementation. If the classification is Generally Avoid and you are moving forward, explain the rationale and document alternatives considered. If the classification is Evaluate, document the cost-benefit analysis supporting the decision.

ADDITIONAL CONSIDERATIONS

Level 2 Safeguards (For Substantial-Risk Systems Only)

If the system is classified as Substantial Risk (Careful Implementation or Generally Avoid) and you are moving forward, document the additional Level 2 safeguards that will be required, including: rights protection measures, community engagement plans, independent validation requirements, enhanced audit and oversight procedures, and (if the system is also high-complexity) additional technical safeguards. See Phase 4, Level 2 Enhanced Requirements.

Compliance and Policy Alignment

Note relevant laws, regulations, policies, or ethical frameworks considered in this assessment.

Assessment Team Lead Signature: _____ **Date:** _____

Reviewer/Approver Signature: _____ **Date:** _____

Note: With appropriate confidentiality protections, this memo should be treated as public record to ensure public accountability and transparency.

CCJ Council on Criminal Justice

Appendix F: Procurement Guide

Procurement components establish the contractual foundation that protects your agency, ensures accountability, and maintains compliance throughout the system's lifecycle. Contracts should include the terms outlined below before execution.

Level 1: Core Contract Components for All Systems

System Requirements and Specifications

- Clearly define the problem the system should address
- Define requirements for integration with existing systems

Performance Standards and Metrics

- Require vendors to propose metrics for evaluating success
- Specify accuracy, reliability, and consistency requirements

Evidence and Validation Requirements

- Require documentation of system testing and validation
- Require disclosure of any known limitations or failure modes
- Request case studies or references from criminal justice implementations

Transparency and Explainability

- Require documentation of how the system works
- Demand plain-language explanations suitable for non-technical stakeholders
- For complex systems, require capability to generate case-specific explanations
- Specify access rights for auditors and oversight personnel

Data Governance and Privacy

- Define what data the system will access and process
- Establish data ownership clearly (agency retains ownership)
- Specify data retention and deletion policies
- Establish security and encryption standards
- Require compliance with relevant privacy laws and regulations
- Require vendors to disclose contractual and technical controls that prevent foreign government access to sensitive data
- Define data breach notification and redress protocols
- Vendor must certify it will NOT train models on agency/community data without explicit written consent
- If any data sharing is permitted, vendor must anonymize data per industry standards (k-anonymity, differential privacy, or equivalent)
- Agency retains right to audit vendor data practices

Bias Testing and Fairness

- Require evidence of bias testing across demographic groups where relevant
- Specify fairness metrics appropriate for your context
- Require disclosure of any known bias issues and mitigation strategies

Human Oversight and Override

- Specify that human operators must be able to override system recommendations
- Require documentation of how human oversight will be facilitated

Training and Support

- Specify vendor responsibilities for initial training
- Require ongoing no-/low-cost training resources for the life of the contract
- Require training for new staff and refresher courses for existing users
- Establish technical support requirements and response times

System Changes and Updates

- Mandate notification for any system changes
- Require re-validation after updates
- Include the right to refuse updates that change risk profile

Liability and Indemnification

- Require vendor to accept liability for system errors or failures
- Include in contract a clause requiring the vendor to indemnify the agency for civil rights violations caused by the system
- Specify that the vendor must certify compliance with all applicable nondiscrimination laws

Termination

- When additive, the contract must include a clause allowing you to terminate the contract for convenience, performance failure, or rights concerns

Level 2: Additional Contract Components for Substantial-Risk Systems

Enhanced Validation and Testing

- Require independent validation of performance claims
- Specify pilot program requirements and success criteria

Enhanced Rights Protections

- Require legal review of system impacts on due process and equal protection
- Specify how individuals' rights to challenge decisions will be preserved

Community Engagement Support

- Require vendor participation in community meetings and explanations
- Specify vendor obligations to address community concerns

Enhanced Auditability

- Require algorithm transparency, with appropriate intellectual property protections
- Establish audit rights for independent experts

CCJ Council on Criminal Justice

Appendix G: Implementation Planning and Memorandum Template

Core Planning Elements

1. Implementation Timeline

- Define clear milestones and target dates
- Build in adequate time for each phase
- Account for pilot program duration (minimum 6 months for Substantial-Risk Systems)
- Include buffer time for addressing issues discovered during pilot

2. Implementation Team and Roles

- Designate Implementation Lead with clear authority and accountability
- Assign specific responsibilities for:
    - Technical integration and system administration
    - User training and support
    - Performance monitoring and evaluation
    - Community engagement (for Substantial-Risk Systems)
    - Legal and compliance oversight
- Establish clear escalation paths for issues

3. Technical Integration

- Plan integration with existing systems and workflows
- Identify and address technical dependencies
- Establish testing protocols
- Define rollback procedures for use if issues arise
- Ensure IT infrastructure can support the system

4. Change Management

- Develop communication plan for stakeholders
- Address staff concerns and resistance
- Build organizational buy-in through engagement

CCJ Council on Criminal Justice

Implementation Memorandum

Use the following template to help ensure that system implementation aligns with the expectations established by this framework and that responsibilities are assigned for the process.

TO: [Recipient Name/Department]

FROM: [Assessment Team Lead Name/Title]

DATE: [Assessment Date]

RE: AI System Classification Assessment - [System Name]

SYSTEM OVERVIEW

System Name: _____

Vendor/Developer: _____

Classification Summary: _____

Risk Level: _____

Opportunity Level: _____

Implementation Type: _____

IMPLEMENTATION TIMELINE

Milestone		
Procurement Complete	_____	_____
Vendor Onboarding	_____	_____
Pilot Program Start	_____	_____
Pilot Program End	_____	_____
Full Deployment Decision	_____	_____
Full Deployment (if approved)	_____	_____

PILOT PROGRAM DESIGN

Duration and Scope

Pilot Duration: _____

Pilot Participants/Departments: _____

Pilot Design and Methodology

Describe the pilot program structure, including which departments or units will participate, how the system will be deployed, what processes or decisions it will support, and how it will be integrated with existing workflows. Include any limitations or constraints on pilot usage.

_____

_____

_____

Success Criteria and Metrics

Define specific, measurable criteria for evaluating pilot success. Include both quantitative metrics (e.g., accuracy rates, time savings, error reduction) and qualitative measures (e.g., user satisfaction, fairness assessments, integration effectiveness).

_____

_____

_____

Data Collection and Evaluation Plan

Describe how data will be collected, analyzed, and reported throughout the pilot. Include evaluation points, reporting frequency, and decision-making criteria for proceeding to full deployment.

_____

_____

_____

SAFEGUARDS AND RISK MITIGATION

Human Oversight Framework

Detail the human oversight model, including who will review AI-generated outputs or recommendations and at what frequency; what authority humans retain; and how human judgment will be documented. Specify whether human review is required before action or can occur after.

Bias Monitoring and Fairness Assessment

Describe procedures for monitoring and addressing potential bias, including what metrics will be tracked, how often assessments will occur, what thresholds trigger intervention, and how disparate impacts will be identified and corrected.

Security and Privacy Protections

Outline data security measures, privacy protections, access controls, and compliance with relevant regulations.

Contingency and Rollback Procedures

Describe conditions under which the system would be suspended or terminated, as well as the procedures for safe system deactivation.

CCJ Council on Criminal Justice

ACCOUNTABILITY STRUCTURE

Decision Authority and Approval Chain

Primary Decision Authority: _____

Approval Required From: _____

Escalation Path for Issues: _____

Oversight Responsibility

Day-to-Day Oversight: _____

Technical Oversight: _____

Compliance Oversight: _____

Community and Stakeholder Engagement

Describe how affected communities, employees, or other stakeholders will be informed, consulted, and involved in the implementation. Include plans for transparency, feedback mechanisms, and ongoing communication.

_____

_____

_____

RESOURCE REQUIREMENTS

Budget

Estimated Costs: $ _____

Source of Funding: _____

Personnel

Staff Time Required: _____

Training Needs: _____

Technical Infrastructure

IT Support Required: _____

Integration Requirements: _____

NEXT STEPS AND APPROVALS

Immediate Next Actions:

_____

_____

_____

Required Approvals: _____

Expected Implementation Start Date: _____

Implementation Lead Signature: _____ **Date:** _____

Approving Authority Lead Signature: _____ **Date:** _____

Appendix H: Ongoing Monitoring and Assessment

Scheduled Reassessment

- **Substantial-Risk Systems:** A reassessment should be conducted annually.

Triggered Reassessment

Vendors should immediately return to conduct a new, full assessment if **ANY** of the following events occur:

- **System Changes**: An update changes the system's core functionality, learning processes, or decision logic. Cases where an AI-powered feature has been added to existing (non-AI) products should be treated as new AI deployments for the purposes of this framework.
- **Scope Expansion:** The system is applied to new use cases or new populations.
- **Performance Issues:** There is evidence of unexpected behaviors, systematic errors, or performance degradation.
- **Rights Concerns:** Bias, disparate impacts, or constitutional compliance issues are identified.
- **Environmental Changes:** Relevant laws, policies, or the human rights environment shifts.

Risk Escalation Protocol

- **If Reclassified to a Higher Risk Level:** Should immediately begin implementing all requirements for the new, higher level.
- **If Reclassified as Prohibited:** Cease operations of the system immediately. Follow the protocol in Appendix B.

Note: *You should negotiate for ongoing monitoring and assessment responsibilities to be included in vendor contracts and service agreements where appropriate. At a minimum, you should discuss with vendors how reassessment triggers and data access requirements will be addressed throughout the system lifecycle.*

Appendix I: Guidance for Deployed AI Systems

Responsible AI governance is an ongoing obligation. Use the approach outlined below to effectively manage the system and minimize negative outcomes.

Triage Assessment

Before conducting a full retroactive assessment, perform a triage to identify systems requiring attention:

Priority 1 — Assess Immediately

- Systems that influence liberty decisions (detention, sentencing, release)
- Systems involving surveillance or monitoring of individuals
- Systems that have generated complaints, errors, or legal challenges

Priority 2 — Assess Soon

- Systems processing sensitive personal data
- Systems with known or suspected demographic performance differences
- Systems that have undergone significant updates since deployment
- Systems approaching contract renewal

Priority 3 — Assess on Regular Schedule

- Administrative systems with limited impact on individuals
- Systems with existing documentation and no identified concerns

Conduct a Retroactive Assessment

For each existing system, work through the framework phases as follows:

Phase 1 (Foundation and Readiness) — Current State: Document the problem the system was intended to solve and assess your agency's *current* capacity to manage this system responsibly. *Do you have the capacity to manage this system responsibly today to accomplish the intended solution?*

Phase 2 (Classification) — Full Assessment: Conduct the complete classification process as if you were evaluating the system for the first time. *If this system were proposed today, would you approve it?*

Phase 3 (Procurement) — Contract Review: Review your existing contract against the Procurement Guide (Appendix F). *Does your contract provide adequate protection and accountability?*

Phase 4 (Implementation) — Gap Analysis: Assess which implementation requirements are already in place versus missing. *What implementation safeguards are missing, and how will you address them?*

Decision Framework for Existing Systems

After completing the retroactive assessment, you will face one of four scenarios:

Scenario A: System passes assessment with current safeguards

- Document findings in Classification Memo
- Establish ongoing monitoring per Phase 5

Scenario B: System passes assessment but requires additional safeguards

- Document required safeguards
- Develop implementation timeline
- Seek contract amendments if needed

Scenario C: System would be classified as "Generally Avoid" if proposed today

- Conduct alternatives assessment
- Develop transition plan if alternatives are available
- If no alternatives exist, implement maximum safeguards
- Do not renew contract without leadership-level review

Scenario D: System is Prohibited

- Cease operations
- Follow Protocol for Prohibited Systems (Appendix B)
- Develop transition plan for affected operations

Documentation Requirements for Retroactive Assessments

All retroactive assessments should produce:

1. **Retroactive Classification Memo** — Use the standard template (Appendix E) with an additional section noting:
   - Date system was originally deployed
   - Whether original assessment documentation exists
   - How current assessment differs from original, if applicable
2. **Remediation Plan** (if gaps are identified) — Documenting:
   - Specific gaps to be addressed
   - Timeline for remediation
   - Responsible parties
   - Resources required
   - Interim risk mitigation measures
3. **Decision Record** — Documenting:
   - Final decision (continue, modify, discontinue)
   - Approval authority
   - Conditions or limitations
   - Next reassessment date

Appendix J: General-Purpose AI Tools in Criminal Justice Settings

While the full sequential process in this framework (Phases 1–5) is designed for formal procurement of dedicated systems, the foundational questions in Phases 1 and 2 apply equally to general-purpose tools used in case-related work:

- **What problem is the tool being used to solve?** Staff using AI informally are making implicit judgments about fitness for purpose. Making these judgments explicit helps identify where AI assistance is appropriate and where it is not.
- **What are the risks?** General-purpose tools can fabricate plausible-sounding information (sometimes called "hallucination"), produce outputs that reflect biases in their training data, and may store or learn from data entered into them. These risks exist regardless of whether the tool was formally procured.
- **Is there meaningful human oversight?** An operator who pastes AI-generated text into a court filing without substantive review has effectively delegated a professional judgment to a machine. The oversight principles in this framework apply to that decision just as they apply to a risk assessment algorithm.

Recommended Action: Develop an Acceptable-Use Policy

Your agency should develop a policy governing staff use of general-purpose AI tools for case-related work. At a minimum, such a policy should address:

- **Permitted and prohibited uses.** Which categories of work may use general-purpose AI assistance (e.g., drafting routine correspondence, researching legal questions, analyzing data), and which may not (e.g., making recommendations about individual liberty, generating evidence summaries presented as original analysis, communicating with affected individuals)?
- **Data entry restrictions.** What types of information may be entered into general-purpose AI tools? Personally identifiable information, case-specific details, confidential informant data, sealed records, and other sensitive information generally should not be entered into external AI systems unless the agency has confirmed the provider's data handling practices through a formal review.

- **Review and attribution requirements.** All AI-generated content used in official work products should be substantively reviewed by the responsible professional. Such content should not be presented as original human work product without disclosure.
- **Documentation.** How should the use of general-purpose AI tools in case-related work be recorded? At a minimum, operators should be able to identify that AI was used, what purpose it was used for, what sort of review was conducted, and who performed the review.
- **Training.** Staff should receive training on the capabilities and limitations of general-purpose AI tools. Such instruction should include the tendency of these systems to generate confident but inaccurate outputs, the potential for bias, and the professional obligations that attend the use of AI assistance.

If your agency has already deployed purpose-built AI systems through formal procurement, you should also consider whether staff are supplementing those systems with general-purpose tools in ways that fall outside existing governance structures.