

# Assessing AI for Criminal Justice

## A User Decision Framework

March 2026

### 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.

[Download PDF Version](#)

## **Overview & User Guide**

[Glossary](#)

[Overview](#)

[User Guide](#)

[User Profiles](#)

[Questions for Future Work](#)

## **Assessment Workflow**

[Phase 1: Foundation and Readiness](#)

[Phase 2: Classification](#)

[Phase 3: Procurement](#)

[Phase 4: Implementation](#)

[Phase 5: Ongoing Management & Reassessment](#)

## Assessment Tools

- [A: AI Readiness Assessment](#)
- [B: Protocol for Prohibited Systems](#)
- [C: System Complexity and Interpretability Assessment](#)
- [D: Sector Context Guidance](#)
- [E: Classification Memorandum Template](#)
- [F: Procurement Guide](#)
- [G: Implementation Planning and Memorandum Template](#)
- [H: Ongoing Monitoring and Assessment](#)
- [I: Guidance for Deployed AI Systems](#)
- [J: General-Purpose AI Tools in Criminal Justice Settings](#)

### 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.

## Table of Contents

## Overview and User Guide

[Glossary](#)

[Overview](#)

[User Guide](#)

[User Profiles](#)

[Questions for Future Work](#)

## Assessment Workflow

[Phase 1: Foundation and Readiness](#)

[Phase 2: Classification](#)

[Phase 3: Procurement](#)

[Phase 4: Implementation](#)

[Phase 5: Ongoing Management & Reassessment](#)

## Assessment Tools

[A: AI Readiness Assessment](#)

[B: Protocol for Prohibited Systems](#)

[C: System Complexity and Interpretability Assessment](#)

[D: Sector Context Guidance](#)

[E: Classification Memorandum Template](#)

[F: Procurement Guide](#)

[G: Implementation Planning and Memorandum Template](#)

[H: Ongoing Monitoring and Assessment](#)

[I: Guidance for Deployed AI Systems](#)

[J: General-Purpose AI Tools in Criminal Justice Settings](#)

## 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.

**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**

**Agency Leaders** (chiefs, directors, sheriffs, court administrators, corrections commissioners)

### **How to Use This Framework**

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.

## User Group

**Procurement & Legal Officials** (general counsel, procurement directors, contract officers, county attorneys)

**Project Managers & IT Staff** (IT directors, system administrators, data officers)

**Policymakers** (legislators, council members, commission staff, oversight bodies)

**Community Representatives & Advocates** (public defenders, civil rights organizations, community advisory members, crime survivors, directly impacted individuals)

**AI Developers & Vendors** (technology companies, product designers, AI researchers, vendors seeking to develop or sell AI tools for criminal justice settings)

## How to Use This Framework

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.

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.

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.

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.

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

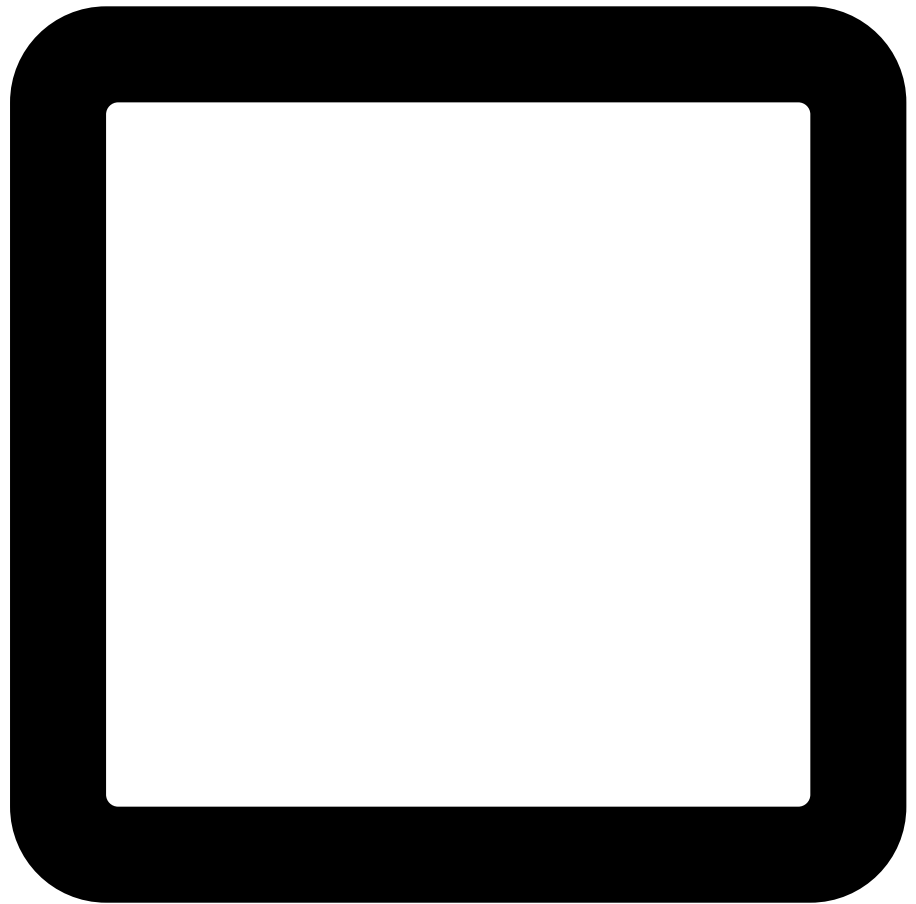
Adopting AI is a policy decision with implications for justice, safety, and public trust. Before proceeding, use the [AI Readiness Assessment Worksheet \(see Appendix A\)](#) to evaluate your organization's capacity in areas like data governance, technical expertise, legal frameworks, and community engagement.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

## Phase 1 Complete Checklist

**Requirement  
Problem Definition**

**Complete?**

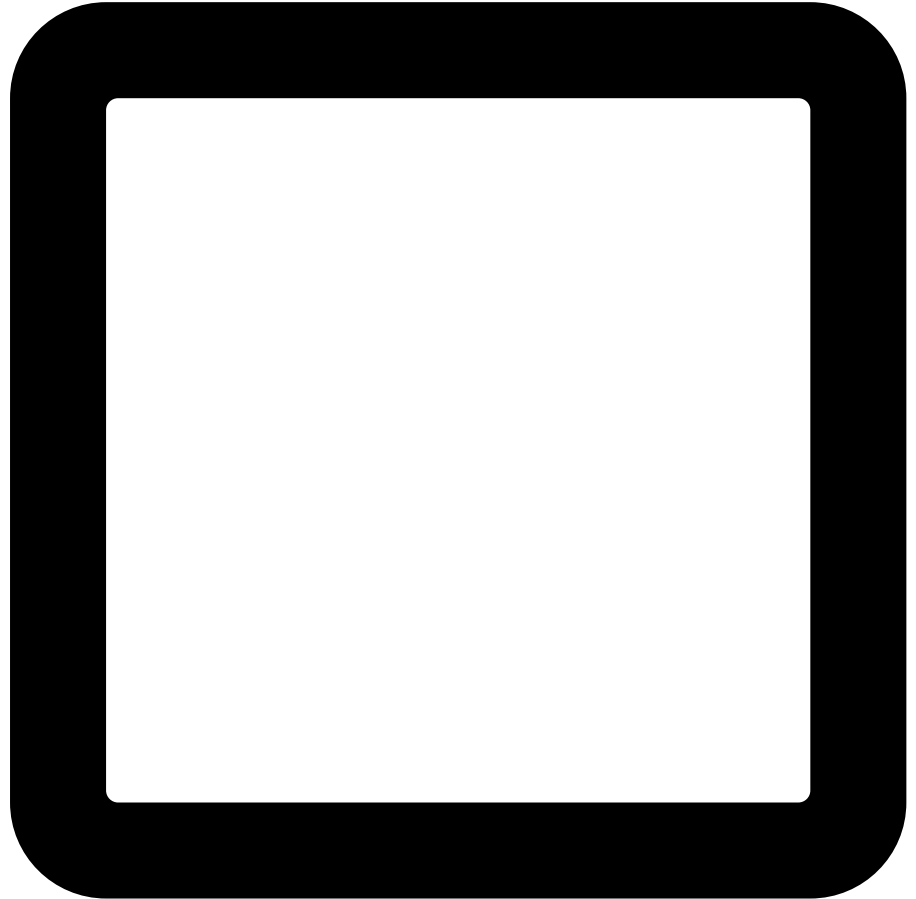
Problem statement is documented in specific, measurable terms



**Requirement**

**Complete?**

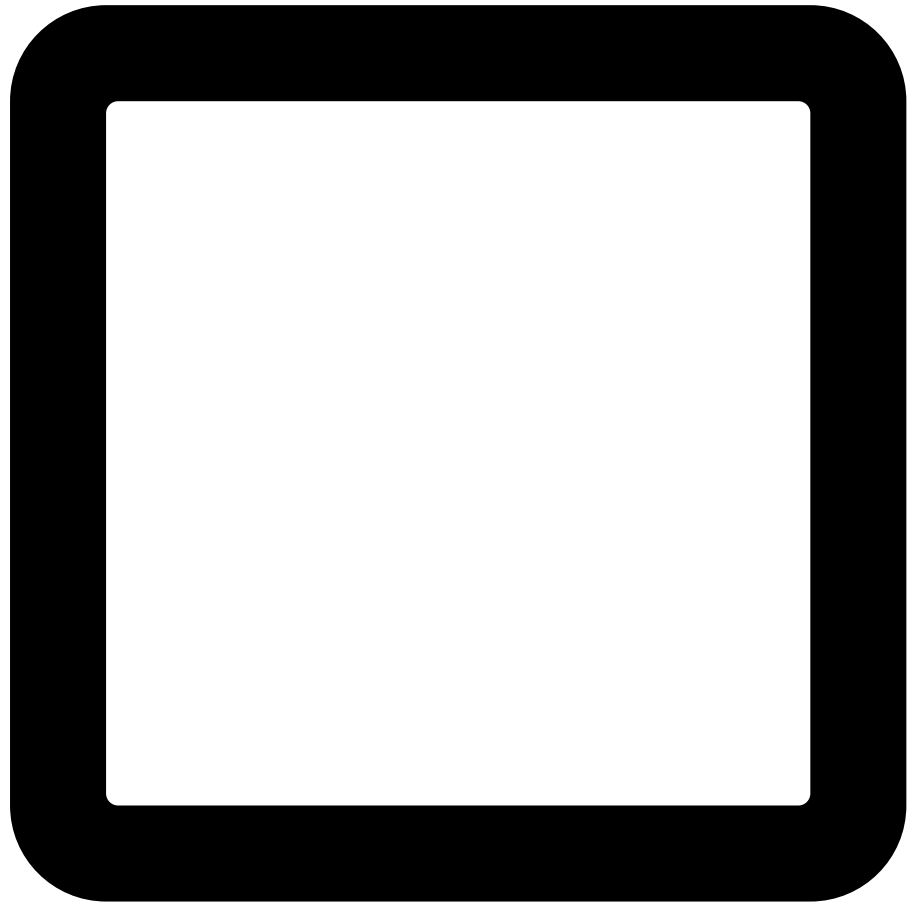
Theory of change explains  
why AI may be preferable  
to alternatives



**Requirement**

**Complete?**

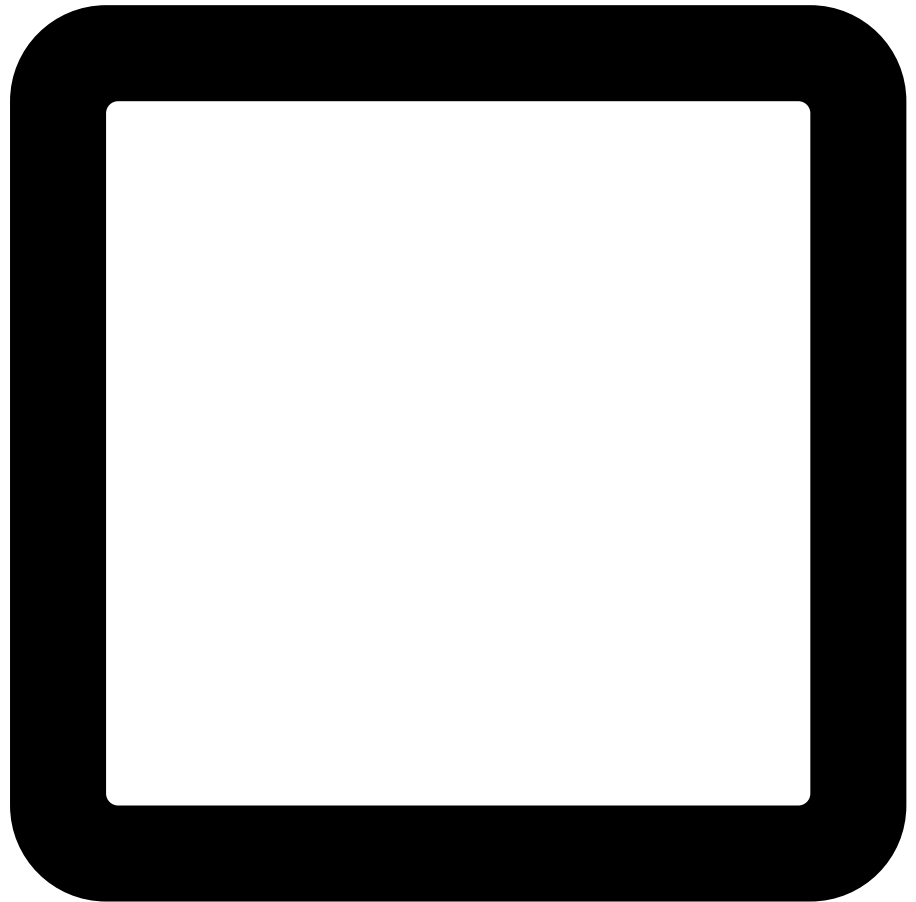
Success metrics are  
defined and measurable



**Requirement**

**Complete?**

Relevant stakeholders  
have reviewed the  
problem scoping

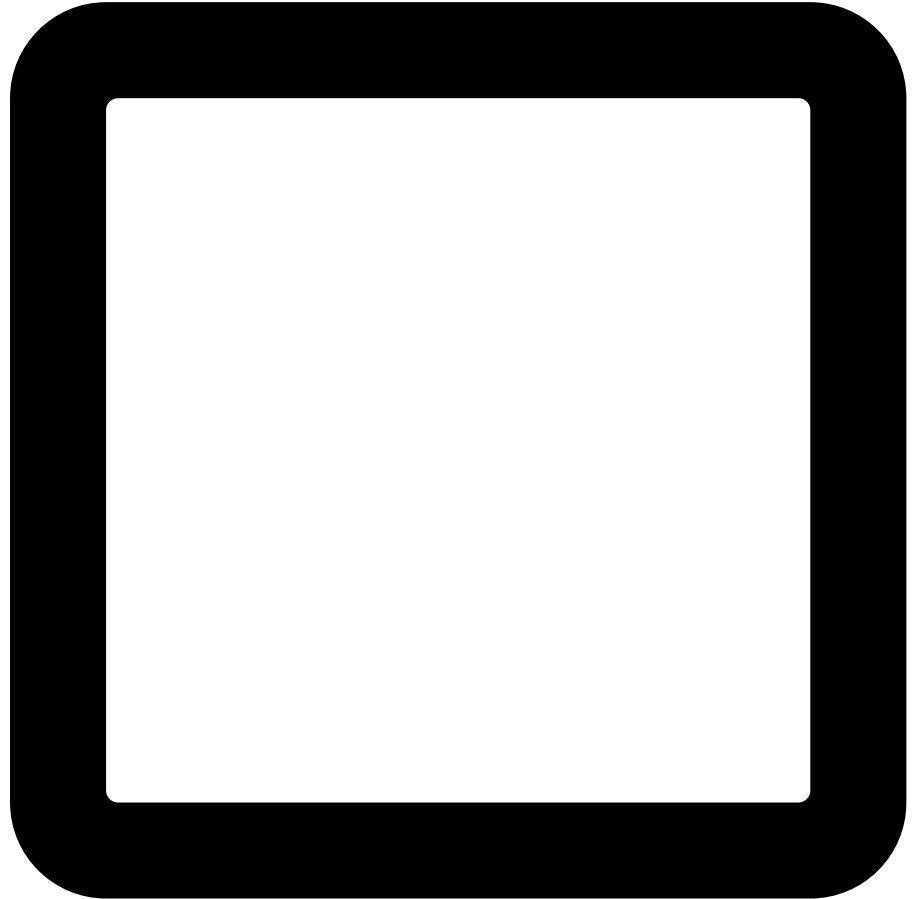


**Organizational  
Readiness**

**Requirement**

**Complete?**

AI Readiness Assessment  
[\(Appendix A\)](#) is complete



**Requirement**

**Complete?**

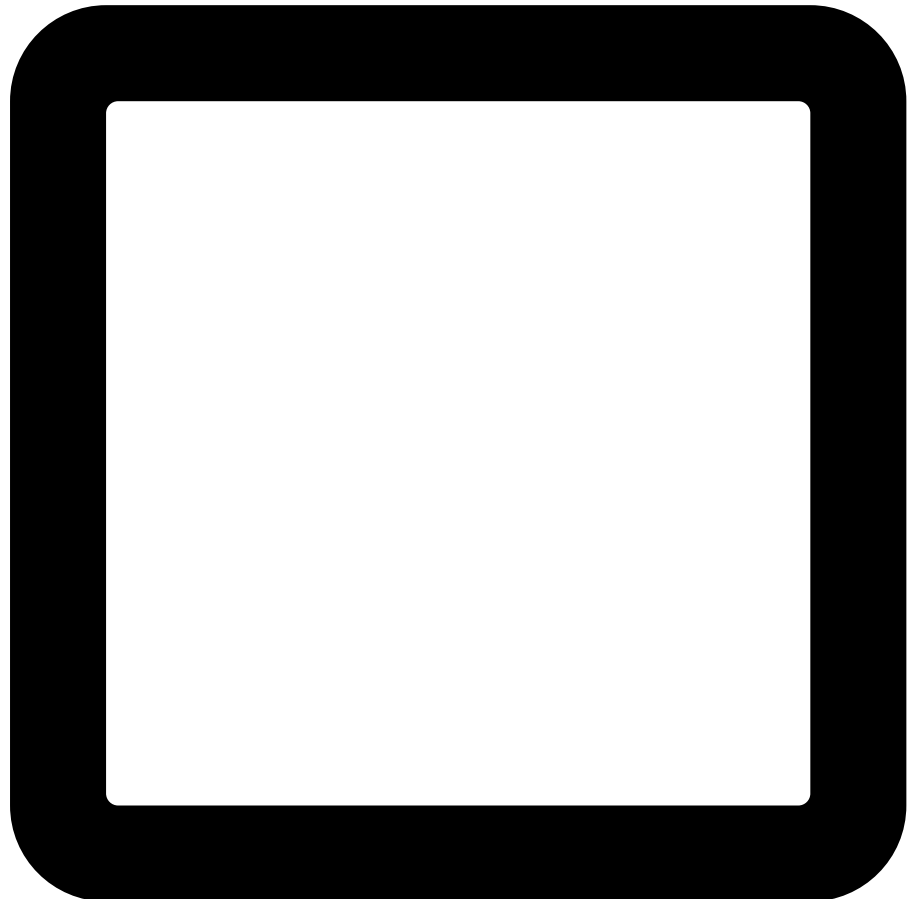
Adequate capacity is confirmed, or remediation plans are prepared for capacity gaps

A large, empty rounded square box with a thick black border, intended for a response or status.

**Requirement**

**Complete?**

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?

**Question**

**Yes**

**No**

Does the system circumvent or undermine established legal or constitutional protections (e.g., due process, equal protection)?

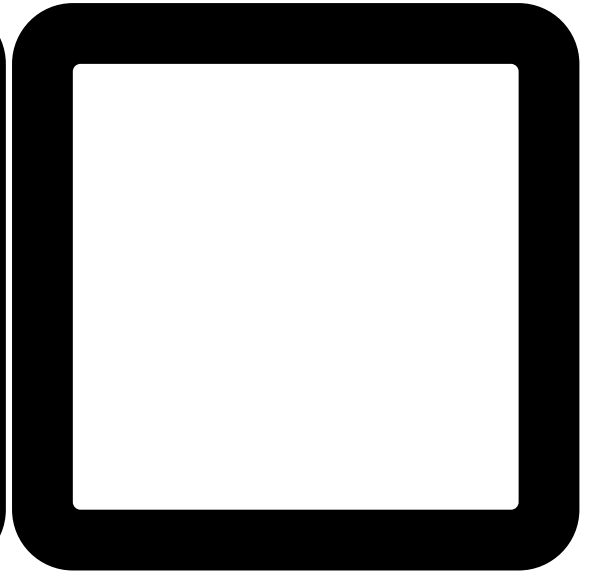
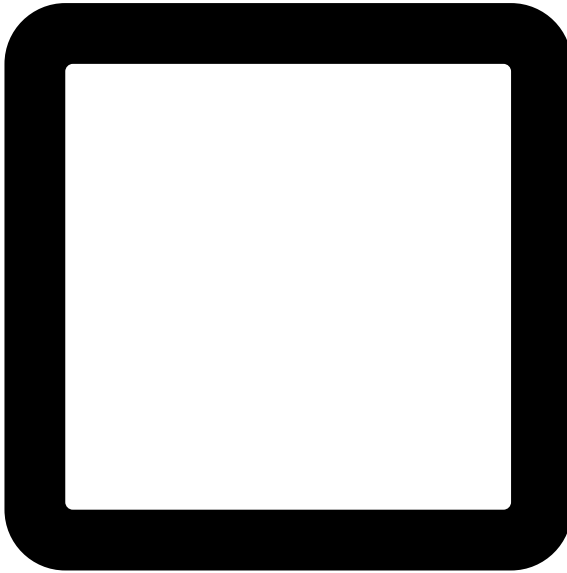
Does the system perform individualized tracking and surveillance of or otherwise have a chilling effect on a group engaging in lawful, constitutionally protected activities (e.g., First Amendment-related activities)?

**Question**

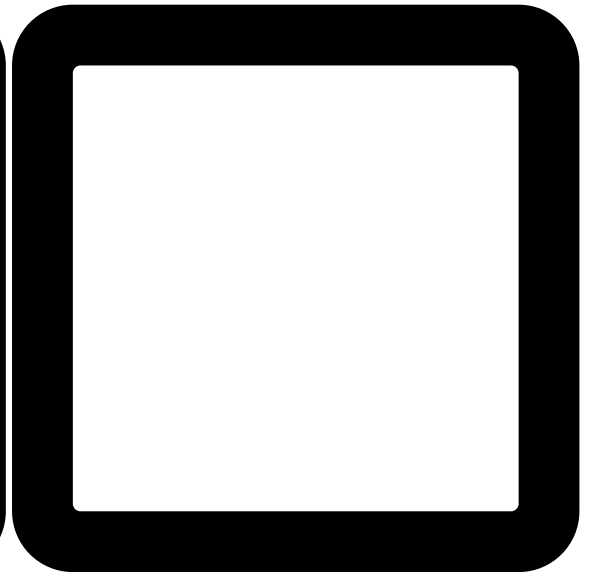
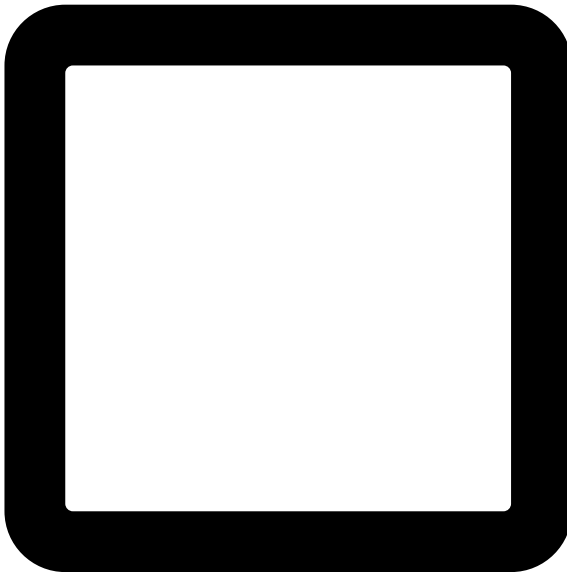
**Yes**

**No**

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
<b>LOW</b>	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.
<b>SUBSTANTIAL</b>	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**.

<b>Opportunity Level</b>	<b>Performance Improvement</b>	<b>Evidence Quality</b>	<b>Stakeholder Support</b>	<b>Cost-Effectiveness</b>
<b>LOW</b>	Minimal, uncertain, or no improvement.	Stakeholders skeptical of necessity.	Little or no supporting evidence; claims are speculative.	Does not favor AI.

<b>Opportunity Level</b>	<b>Performance Improvement</b>	<b>Evidence Quality</b>	<b>Stakeholder Support</b>	<b>Cost-Effectiveness</b>
<b>SUBSTANTIAL</b>	Supported by evidence from pilots or independent research.	Community and end-users validate the value.	Favorable.	Demonstrable improvement over existing processes.

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).

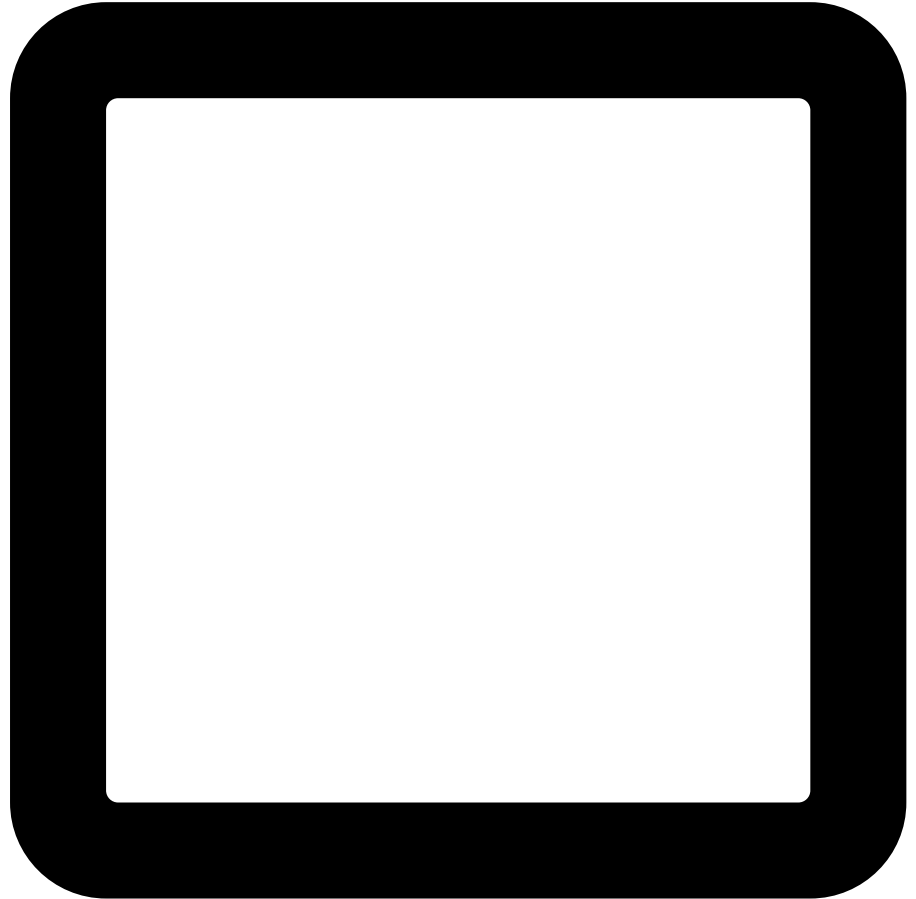
### **Phase 2 Completion Checklist**

Before proceeding to Phase 3, confirm the following:

**Requirement**

**Complete?**

Assessment team properly  
constituted



**Requirement**

**Complete?**

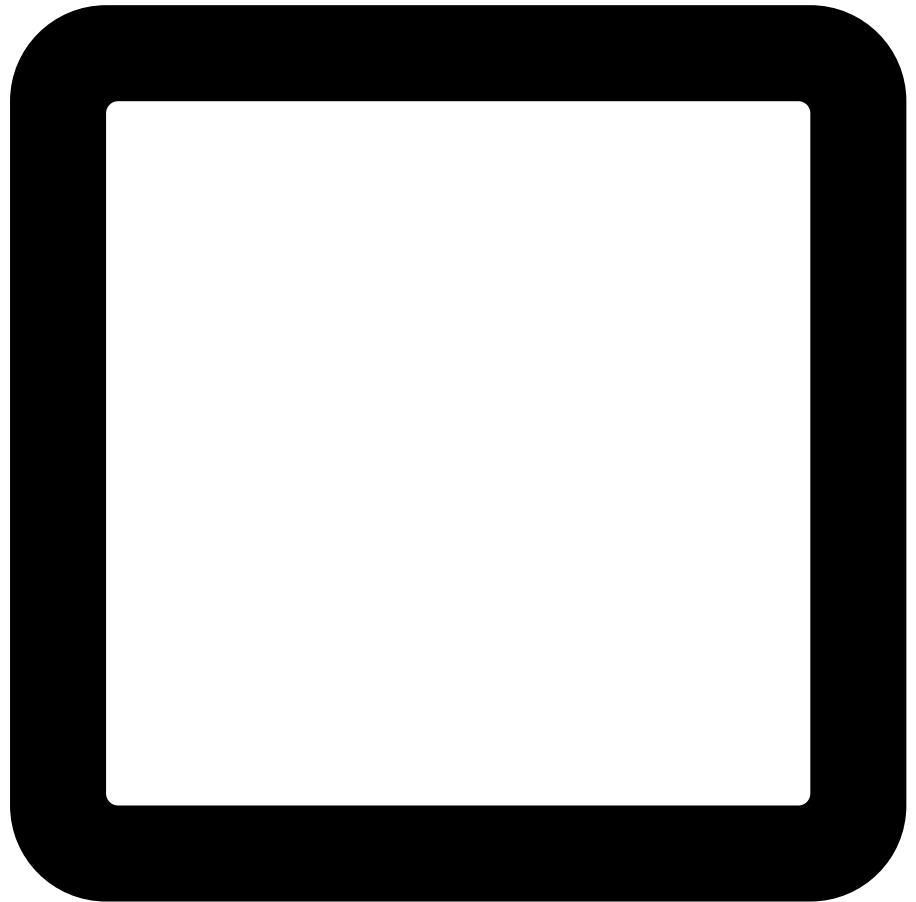
Prohibited use screening  
complete (all NO)

A large, empty rounded square box with a thick black border, intended for a response or status.

**Requirement**

**Complete?**

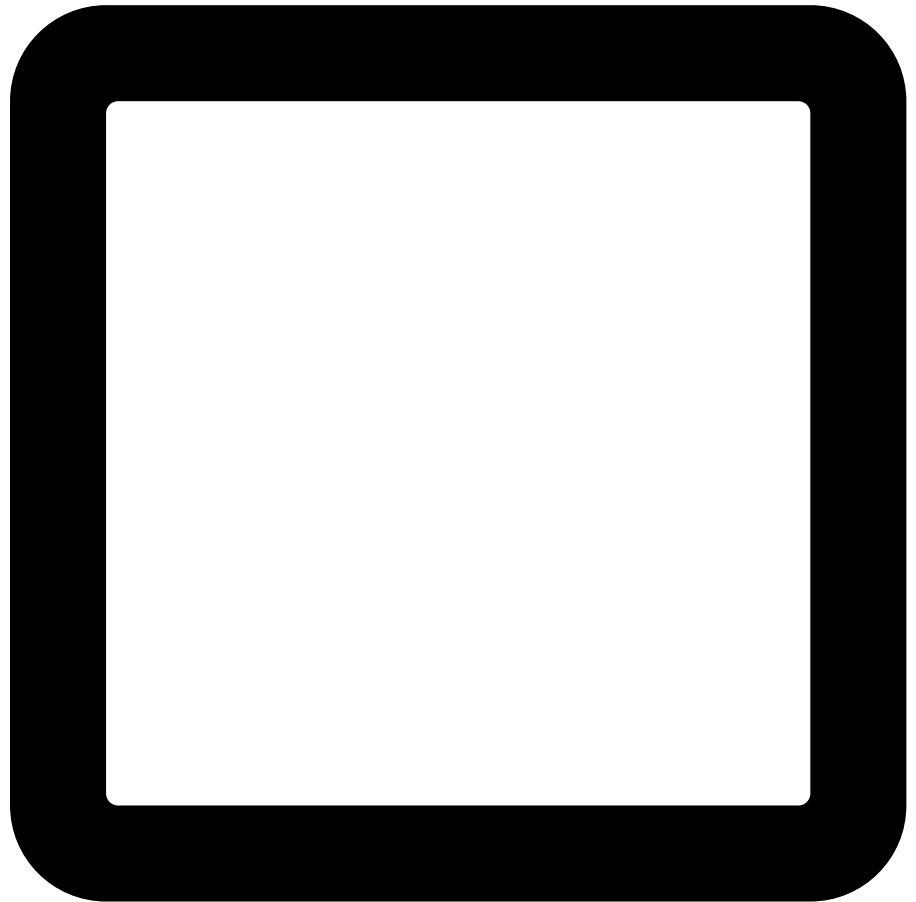
System complexity  
assessment complete  
[\(Appendix C\)](#)



**Requirement**

**Complete?**

Sector context considered  
[\(Appendix D\)](#)



**Requirement**

**Complete?**

Risk level determined and documented

A large, empty rounded square box with a thick black border, intended for a response or checkmark.

**Requirement**

**Complete?**

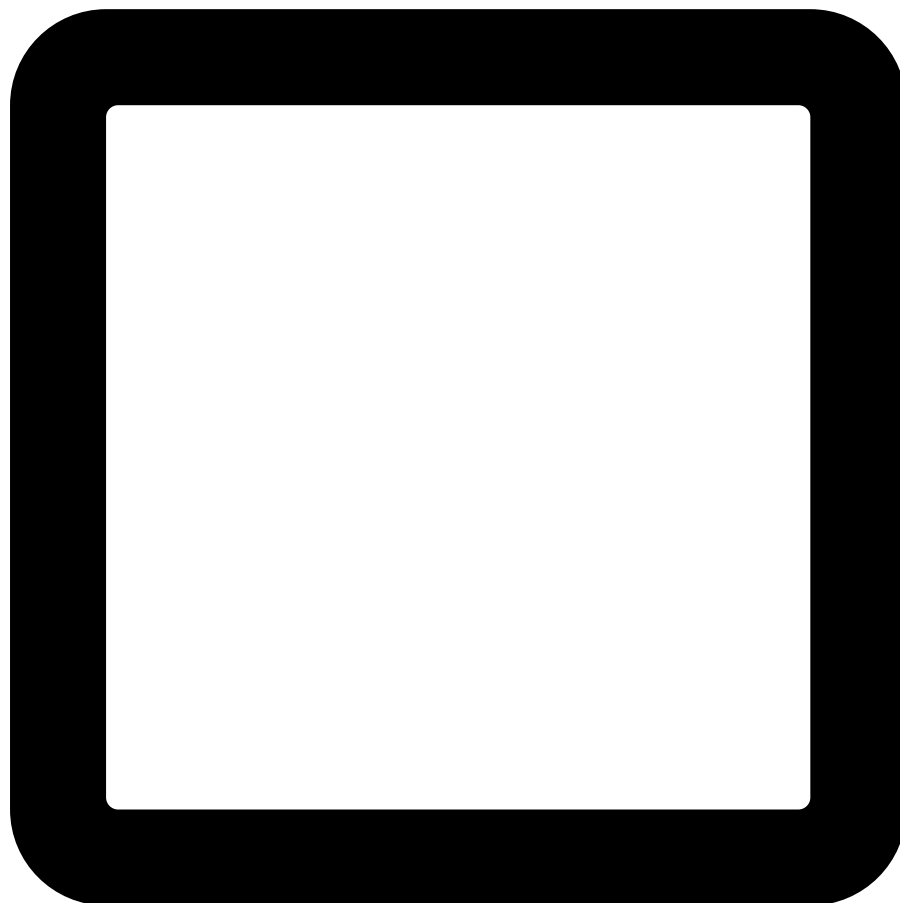
Opportunity level  
determined and  
documented

A large, empty rounded square box with a thick black border, intended for a status or completion mark.

**Requirement**

**Complete?**

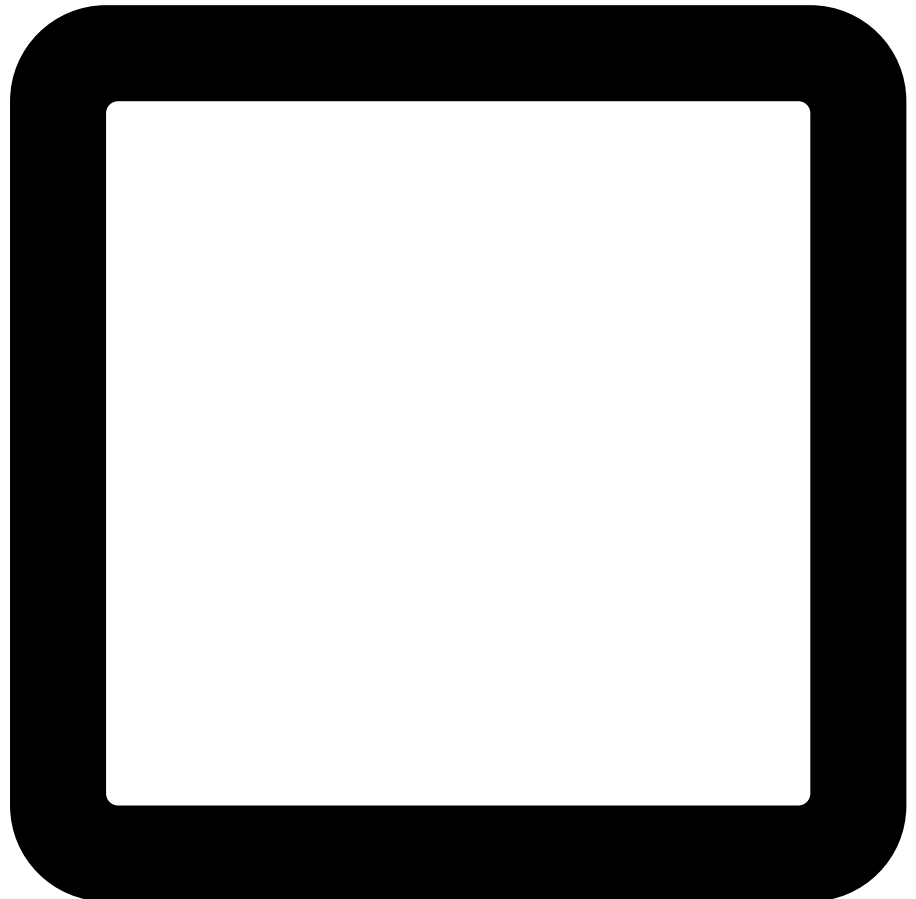
Classification Memo  
[\(Appendix E\)](#) complete



**Requirement**

**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: