ENTERPRISE
information services

Artificial Intelligence Risk Management Framework

Contents

Executive Summary

This document serves as the guide to Oregon Executive Branch agencies for assessing and managing risk of the use of artificial intelligence (AI) in information technology tools that support the work of employees and the agency. The guide helps agency leaders identify and understand potential risks from the use of AI and provides a detailed list of potential safeguards and mitigations agencies may choose to deploy in their implementations.

Why AI has both opportunity and risk

AI is a transformative technology that has the ability to support humans in numerous ways. With access to accurate and secure data, appropriate human review of AI-generated outputs, and ongoing monitoring of AI systems, as well as other key components that support transparency, accountability, safety and trust, AI can be a major support tool. At the same time, AI must be managed and its outputs verified. AI makes mistakes, and therefore the process of human review must be evaluated and specified to mitigate risks for each AI use case.

For a broad range of use cases, primarily those that use publicly available data and whose audience is internal facing, the risks are low and the state guide provides clear pathways to deploy AI responsibly. For other uses, particularly those that use data with sensitive or personally identifiable information or play a role in a process that could impact a person's rights or safety, clear pathways are still in development and will require additional documentation and approval.

What this guide applies to

The State Chief Information Officer has approved the use of Microsoft Copilot Chat by all state employees. Specific recommendations on the use of Copilot Chat, Microsoft 365 (M365) Copilot and Microsoft Teams Premium are provided in a separate policy document. Beyond these pre-approved tools, all use of generative or agentic artificial intelligence by agencies must be approved by Enterprise Information Services (EIS). This includes third party services, including Software as a Service (SaaS) subscriptions that have AI components, as well as custom-built projects. Specific definitions of generative and agentic AI are provided in the Statewide Responsible AI Usage Policy.

How this guide works

The starting point for any IT project, with or without AI, is the writing of a business case for the project. The IT investment process for obtaining approval for agency IT investments contains a list of required questions. When AI use is proposed, an additional initial list of questions on the AI Use Form must be answered to determine the risk level specific to the

proposed use of AI. If the proposed AI use is determined to be higher-risk, then additional requirements must be met.

The importance of agency governance

Consistent with their IT strategic plan and IT governance activities, agencies should define a process for intake and evaluation of AI use in agency processes. To support optimal decision-making, this process should involve a diverse group that includes leaders and frontline staff, data and IT representatives as well as business representatives. Many agencies have established IT governance and data governance groups, and AI governance should be incorporated into these provided a diverse membership is maintained. This is consistent with the principles from the Final Recommendations of the Oregon State Government AI Advisory Council[1] as well as the National Institute of Standards and Technology (NIST) AI Risk Management Framework[2].

Crawl, before you walk or run

This is the inaugural version of this guide for state of Oregon agencies, boards and commissions. It is geared toward agencies looking to use AI in low-risk ways and provides a foundation that will be added to over the coming months and years. The definition of what constitutes higher-risk use cases is also described. Agencies seeking to implement AI in use cases falling into these higher-risk categories should engage with EIS directly.

[1] Oregon State Government Artificial Intelligence Advisory Council Final Recommended Action Plan, February 4, 2025.
[2] NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0). GOVERN 3.1 is the specific governance recommendation referenced.

Initial Risk Evaluation: Completion of the AI Use Form

All uses of applications or systems that use AI, outside of enterprise-approved systems such as Copilot, should begin with the completion and submission of an AI Use Form as part of the information technology investment (ITI) submission process. The fields on the AI Use Form are used to perform an initial risk evaluation.

A key part of the AI Use Form is the identification of features that may categorize the use as higher-risk. These include:

- Public audience or users
- Use of level 3 or regulated data
- Rights-impacting or safety-impacting AI processing

In addition, for all AI uses, even those not deemed higher-risk, the AI Use Form asks you to complete a set of questions and requirements describing your understanding of the risks, mitigations, and testing and monitoring plans related to accuracy, bias, privacy, and ethics. The questions are as follows:

- What are the identified risks of Gen AI for this particular use case?
- Are there specific risks related to privacy, accuracy, bias and/or ethics? Please define.
- What specific measures will be implemented to mitigate each of these risks?
- What testing methods will be executed prior to deployment to validate accuracy, safety, and reliability?
- What processes will be used to continuously audit and validate system outputs after deployment?
- What triggers would require additional testing or risk re-evaluation?

To assist in answering these questions on the AI Use Form, below is a summary of risks and potential safeguards or mitigations.

Summary of Risks and Safeguards/Mitigations

#	Category	Example Safeguards / Mitigations
1	Data Security and Privacy	Data anonymization and redaction, encryption in transit and at rest, role-based and least-privilege access, approved secure hosting, controls on vendor data retention and reuse, protection from prompt injection
2	Vendor and Compliance Management	Vendor security and privacy due diligence, verification of certifications, contractual data-use restrictions, legal review for regulated or rights-impacting uses, breach notification and audit support
3	Acceptable Use and Policy	Defined approved and prohibited AI uses, clear human accountability, staff training on limitations and verification, prohibition on bypassing review or approval processes
4	Data Quality, Validation, and Output Review	Human review of outputs, verify that data is accurate, complete, representative and current from authoritative data sources, validation against laws and policies, source referencing where feasible, periodic quality reviews and audits
5	Bias, Fairness, Justice and Inclusion	Bias and fairness assessments, review of data representativeness, explainability for people-impacting uses, appeal or challenge mechanisms, DEIA (diversity, equity, inclusivity, accessibility) or subject-matter review, Tribal sovereignty
6	Reliability, Oversight, and Human in the Loop	Human oversight and intervention, manual override capability, defined escalation paths, fallback procedures for system failure or unreliable outputs
7	System Performance, Operations, and Resilience	Performance and availability monitoring, capacity planning, backup and recovery procedures, remediation of security vulnerabilities
8	Transparency, Accountability, and Auditability	Disclosure of AI use when material, documentation of purpose and limitations, logging of inputs and outputs, audit and public-records support, assigned accountable business owner
9	Accessibility and Usability	Accessibility standards compliance, clear and understandable outputs, multilingual support where appropriate, user testing with diverse users
10	Escalation and Continuous Improvement	Defined escalation conditions, user feedback collection, periodic reassessment of safeguards, authority to modify or suspend AI use
11	Decision Authority and Automation Limits	Human approval for rights-impacting decisions, prohibition on fully automated determinations, explainability of AI-influenced decisions
12	Records, Retention, and Public Accountability	Public-records and retention compliance, ability to reproduce AI-assisted outputs, preservation of due process and transparency

Detailed Safeguards/Mitigations

Agencies should review their particular use case and utilize this list in the production of their project design, including testing and monitoring components, as submitted in the IT investment process and in artifacts submitted through the oversight process. While these

examples and safeguards offer a solid foundation for mitigating risks associated with AI tools, state entities should tailor their approach to the specific use cases and products that they employ. Given the diverse range of AI applications and their associated risks, it is crucial to conduct a thorough risk assessment for each specific use case.

1. **RISK: Data Security and Privacy**
   Potential Mitigations:
   - Anonymize, redact, or mask sensitive, confidential, or restricted information before entering it into any AI system.
   - Limit access to AI tools and their outputs to authorized personnel only.
   - Enforce least-privilege access, with administrative access tightly restricted.
   - Encrypt all data processed by AI systems in transit and at rest.
   - Use synthetic data for testing purposes.
   - Use sensitive data only in AI systems approved for the applicable data classification. Apply strict input validation methods to filter out unnecessary sensitive data.
   - Do not use public or consumer AI tools for sensitive information.
   - Prohibit retention, reuse, or model training on agency data by vendors unless explicitly approved.
   - Host AI systems handling sensitive data in secure environments with appropriate network controls and monitoring.
   - Test and mitigate prompt injection, data exfiltration, and retrieval poisoning for any retrieval-augmented generation (RAG)/agent workflow.
   - Restrict tool permissions and connector scope to prevent “excessive access” failure modes.
   - Prevent unauthorized access and data leakage of data accessed by the AI model.

2. **RISK: Vendor and Compliance Management**
   Potential Mitigations:
   - Obtain and review vendor documentation describing data storage, protection, retention, and deletion practices.
   - Verify vendor security claims using the vendor’s own certifications, not only those of hosting providers.
   - Include contractual prohibitions on using agency data for training or secondary purposes without approval.
   - Conduct legal review for AI uses that may affect compliance, rights, or regulated activities.
   - Require vendors to support auditability, breach notification, and security incident reporting.
   - Do not use third-party AI services without contractual data protection and security safeguards in place.
   - Prohibit entering copyrighted/licensed content into AI tools unless usage rights are confirmed.

- Require vendor terms to clarify ownership of outputs and protections for agency intellectual property.
  Gartner explicitly flags third-party sharing/privacy risk and Info-Tech pushes AI compliance strategy; intellectual property is typically bundled into those compliance controls.
- Ensure vulnerabilities from third party packages, including outdated or deprecated models, are addressed.
- Establish an exit plan, including data portability and model/configuration export, to reduce vendor lock-in.

3. **RISK: Acceptable Use and Policy**
   Potential Mitigations:
   - Define approved and prohibited uses of AI tools.
   - Do not treat AI outputs as authoritative without human review.
   - Train users on appropriate use, limitations, and verification of AI outputs.
   - Assign clear responsibility for validating and approving AI outputs.
   - Do not use AI tools to bypass established review, approval, or accountability processes.
   - Inventory approved AI tools and block/monitor unapproved AI use.
   - Require training and periodic audits of AI tool usage.

4. **RISK: Data Quality, Validation, and Output Review**
   Potential Mitigations:
   - Review AI outputs by qualified staff before relying on or sharing them externally.
   - Use curated, authoritative, and up-to-date data sources where possible.
   - Validate outputs for accuracy, completeness, and alignment with applicable laws and policies.
   - Enable users to flag incorrect, misleading, or inappropriate outputs.
   - Provide references or links back to source material where feasible.
   - Periodically review outputs to identify recurring errors or degradation in quality.

5. **RISK: Bias, Fairness, Justice and Inclusion**
   Potential Mitigations:
   - Evaluate AI systems for bias that could lead to unfair or discriminatory outcomes.
   - Review training and reference data for representativeness and known bias risks.
   - Review outputs affecting people or communities with fairness and equity considerations. This includes Tribal data sovereignty.
   - Provide a mechanism to challenge or appeal outcomes influenced by AI.
   - Review culturally sensitive or demographic-specific content with appropriate subject matter experts.

6. **RISK: Reliability, Oversight, and Human in the Loop**
   Potential Mitigations:
   - Support human oversight and intervention for all AI systems.

- Do not replace human judgment in complex, sensitive, or high-impact situations.
- Define escalation paths for cases the system cannot handle reliably.
- Maintain the ability to override, correct, or disable AI outputs.
- Maintain fallback procedures when AI systems fail or produce unreliable results.
- Educate users on safe usage.

7. **RISK: System Performance, Operations, and Resilience**
   Potential Mitigations:
   - Monitor AI systems for model, performance, availability, drift, and failure conditions.
   - Plan for peak usage, system outages, and resource exhaustion.
   - Maintain backup and recovery procedures to support continuity of operations.
   - Identify and remediate security vulnerabilities in a timely manner.

8. **RISK: Transparency, Accountability, and Auditability**
   Potential Mitigations:
   - Document how AI systems are used, including purpose and limitations.
   - Disclose the use of AI when it materially affects outputs or interactions.
   - Maintain logs of inputs, outputs, and user actions.
   - Ensure records support audits, investigations, and public records requirements.
   - Assign a clear business owner accountable for AI use. Require training and periodic audits of AI tool usage.

9. **RISK: Accessibility and Usability**
   Potential Mitigations:
   - Ensure AI systems comply with applicable accessibility standards.
   - Present outputs in a form understandable to the intended audience.
   - Support multiple languages where appropriate.
   - Conduct user testing with diverse users to identify accessibility or usability barriers.

10. **RISK: Escalation and Continuous Improvement**
    Potential Mitigations:
    - Define chain of command and who has authority to pause and/or roll back AI use when risk detection conditions are met.
    - Define conditions requiring escalation to human staff.
    - Collect and review user feedback to identify recurring issues.
    - Periodically reassess AI systems to ensure safeguards remain effective.
    - Suspend or modify AI use when risks increase or harms are identified.

11. **RISK: Decision Authority and Automation Limits**
    Potential Mitigations:
    - Do not use AI to make final decisions affecting rights, benefits, eligibility, or enforcement without human approval.

- Ensure AI-influenced decisions are explainable to staff and affected individuals.
- Document how AI contributed to decisions.

12. **RISK: Records, Retention, and Public Accountability**
    Potential Mitigations:
    - Manage AI-related records in accordance with public records and retention laws.
    - Maintain the ability to reproduce AI-assisted outputs for audits, reviews, or litigation.
    - Ensure transparency in data usage by maintaining clear policies about data retention, usage and deletion.
    - Do not use GenAI in ways that undermine transparency or due process obligations.

Higher-Risk Use Cases

If your audience is public, you are planning to use level 3 or regulated data, or you have a rights-impacting or safety-impacting use of AI, your use is categorized as higher-risk and will have additional requirements, with ultimate approval requiring sign-off by the State CIO.

For higher-risk use cases, additional artifacts will be required. Specifications on these artifacts are still in development and will likely include the following:

Artifact	Brief Description
Detailed list of identified risks or potential risks, and proposed safeguards and mitigations	Following the safeguards and mitigations list in this document, a detailed list of identified potential risks and proposed safeguards and mitigations must be documented.
Detailed testing plan	Testing plan, including methods and results, for: accuracy/quality, fairness/equity/bias, privacy protections, security/red-team/adversarial testing, robustness/known limits, and pass/fail gates used prior to deployment.
Monitoring and incident management plan	Detailed monitoring and incident management plan with escalation flags; what is continuously monitored, rollback criteria, and how incidents are reported/handled. Includes triggers for re-initiating testing due to changes in model, process, etc.
Oregon AI Transparency Disclosure	A plain-language snapshot of what the system does, what data it touches, and how it was tested and is being overseen. EIS will provide a template, and all approved disclosures will be published on the EIS and agency web sites.

Risk Assessment Table and Examples

Risk/ Process	Risk Indicators	Use Case Examples
Lower-risk 1.Complete AI Use Form as part of information technology investment (ITI) process	Level 1 or 2 data only, AND Audience is internal, not the public AND Appropriate human review, including by a qualified subject matter expert if needed.	Content and communication tools used by state employees: grammar correction, summarization, reference generating tools like laws and policies, translation tools, content creation tools Internal chatbots for resource finding and process advice Network and security tools: packet inspection, system monitoring, intrusion prevention/ detection, spam filtering, malware detection, endpoint detection Certain code analysis and development tools
Higher-risk 1.Complete AI Use Form as part of information technology investment (ITI) process 2.Further consultation with EIS and submission of additional artifacts 3.Approval of State CIO required	Audience is public-facing OR Use of level 3 data OR Rights-impacting use: -Predictive analysis of crime -Evaluating a person in an employment context -Evaluating a person in a learning, training or educational context -Evaluating a person's eligibility for a service or good -Administration of justice, law enforcement or immigration OR Safety-impacting use: -Acting as a safety component in a system or products -Informing safety decisions OR AI processing that uses: -Emotion recognition or sentiment analysis -Scraping images -Biometric categorization or identification, including facial recognition	Public-facing chatbots for public interaction and information retrieval Public and direct contact services, including customer service, public relations, recommendations on legal, tax or regulatory information Internal chatbots that have level 3/ confidential data Processing of data that informs safety decisions, such as water treatment ratios, load-bearing specifications for bridges, seismic analysis Processing of confidential data that identifies or describes an individual, including in a public safety context
Not allowed	Use of level 4 data Decision-making without appropriate human review	None

Appendix: References

References:

NIST AI RMF Playbook landing page: NIST AI RMF Playbook
NIST AI Resource Center Playbook explorer (subcategories, downloadable CSV/Excel/JSON): AIRC Playbook
NIST AI RMF 1.0 (AI 100-1): Framework PDF
NIST AI RMF Core overview: AI RMF Core Functions

Additional resources:

OWASP Top 10 for LLM Applications 2025 (prevention/mitigations): LLMAll_en-US_FINAL

Revision History

Version	Date	Change Log
1.0	June 12, 2026	Initial version