ENTERPRISE information services

DAS DEPARTMENT OF ADMINISTRATIVE SERVICES

Statewide Policy

	NUMBER ###-##-###	**SUPERSEDES** Interim Guidance on Generative AI Usage and Access
	EFFECTIVE DATE	**PAGE NUMBER** Pages 1 of 4
	REVIEWED DATE	
DIVISION Enterprise Information Services (State CIO)	**REFERENCE** ORS 276A.206 107-004-130 Information Technology Investment Oversight 107-004-050 Information Asset Classification 107-004-160 Data Governance Policy Link to Procedure: ###-##-### PR	
POLICY OWNER Privacy and Artificial Intelligence		
SUBJECT Responsible Usage of Artificial Intelligence	**APPROVED SIGNATURE** Terrence Woods, State Chief Information Officer	

PURPOSE

The main purpose of this policy is the creation of a robust governance structure that instills confidence and trust in the use of AI at the state. This policy has the following objectives:

- **Establish enterprise governance:** create a governance structure for AI policy and oversight that can keep pace with rapidly evolving technology and standards.
- **Identify and address risks**: establish a risk management model to identify and address potential issues related to bias, fairness, privacy, safety, and security.
- **Ensure accountability**: mandate clear lines of responsibility for all AI systems.
- **Foster transparency**: require mechanisms for internal and public transparency regarding the state's use of AI.
- **Empower the workforce**: build AI literacy and technical competency across the state workforce.
- **Monitor progress**: establish systems for monitoring and reporting AI use.

APPLICABILITY

This policy outlines requirements for AI use in state government executive branch agencies, boards and commissions. The policy applies to generative and agentic AI systems used to conduct the business of the state, including both standalone systems or as components embedded in other software.

This policy applies to all state agencies under the purview of the State Chief Information Officer.

DEFINITIONS

Agentic artificial intelligence: AI systems designed to autonomously pursue complex goals and take actions with limited human intervention, and adapt approaches based on environmental feedback.

Artificial intelligence (AI): A machine-based system that is capable, for a given set of human-defined objectives, of making predictions, recommendations or decisions influencing real or virtual environments and uses machine- or human-based inputs.

Foundation model: an AI model trained on vast datasets, contains at least tens of billions of parameters, and is applicable across a wide range of contexts.

Generative artificial intelligence (GenAI): a software system that generates new content, data, or other output, including images, video, audio, and text, autonomously. It uses algorithms and models to create new and original information from patterns and information it learned previously from a given set of data. Large language models and foundation models are both GenAI. GenAI is distinguished from other AI systems by its primary function of creating new content rather than solely classifying, predicting, or recommending from predefined categories.

Human in the loop (HITL): human-in-the-loop refers to a system or process in which a human actively participates in the operation, supervision and decision-making of an automated system. In the context of AI, HITL means that humans are involved in the AI workflow to ensure accuracy, safety, accountability and ethical decision-making.

Oregon AI transparency disclosure: a plain language summary of what an AI system does, what data it touches, how it was tested and how it is being overseen.

Prompt: the question, instruction or other information entered by a user to generate a response from a GenAI system. It serves as the starting point for the model to generate a relevant reply.

EXCLUSIONS AND SPECIAL SITUATIONS

Non-AI powered automated decision-making systems of numerous kinds have been in use by state agencies for many years. Systems that use only data analytics, statistical modeling or human-specified machine learning, and systems that simply follow rules created solely by people to run tasks automatically, are not covered by this policy.

This policy does not apply to:

a) Basic calculations, spreadsheet operations, or mathematical computations following predetermined formulas;
b) Traditional statistical modeling using regression analysis, time series analysis, or other established statistical methods that do not incorporate machine learning;
c) Data analytics, business intelligence tools, or reporting dashboards that aggregate, visualize, or summarize data without using AI ;
d) Rule-based automation using pre-recorded conditional logic with predefined triggers that automatically initiate predetermined actions (such as "if, then" systems);
e) Database queries, sorting operations, or filtering functions; and
f) Basic AI features embedded in common commercial products used for routine productivity purposes, including predictive text in word processors or dynamic route adjustment based on real-time traffic conditions in map navigation systems.

GENERAL INFORMATION

1. Principles of Responsible AI

The following are the state of Oregon's principles related to the responsible use of artificial intelligence:

- Human Oversight and Review: Human oversight should be intentionally built into AI adoption and day-to-day use, with clear roles and responsibilities for governance and decision-making. AI should not replace subject matter expertise.

- Equity and Representation: AI systems should support equitable outcomes and be designed and used with attention to representation, accessibility, and potential disparate impacts. AI systems should be regularly monitored for fairness and accuracy.

- Privacy and Confidentiality: AI systems should protect privacy and confidentiality, with clear oversight responsibilities and heightened review where sensitive data is involved.

- Risk Management: AI risks should be identified, assessed, measured, and managed throughout the lifecycle of AI usage.

- Safety: AI systems should not reduce overall safety and should be subject to clear safety requirements, including measurable evaluation methods.

- User Experience, Disclosure and Feedback: AI should improve staff and constituent experience in their engagement with state government. Disclosure of AI usage and opportunity to provide user feedback on AI-generated results are required. Transparency and disclosure of AI use, including processes and evaluations, must increase proportionate to risk.

- Workforce Preparedness and Understanding: Workers should be involved in AI adoption and review processes and receive the training needed to use AI appropriately, while the broader workforce should build a baseline understanding of AI capabilities, uses, and implications.

2. Enterprise AI Advisory Committee

The State CIO will establish an Enterprise AI Advisory Committee comprised of representatives from EIS and executive branch agencies and chaired by the State CIO or designee. The committee will serve as advisory to EIS and the State CIO. The responsibilities of the Enterprise AI Advisory Committee include reviewing and providing recommendations to EIS on enterprise AI policy.

3. General Recommended Usage

EIS will provide guidance for generative AI use by state employees on enterprise AI tools including Copilot.

4. AI Risk Management Framework and Usage Requirements

EIS will establish an AI Risk Management Framework to provide detailed standards, requirements and guidance to state agencies for using generative or agentic AI. Any use of AI meeting the applicability standards must comply with the AI Risk Management Framework. The AI Risk Management Framework will align with the National Institute of Standards and Technology (NIST) AI Risk Management Framework and provide clear definitions for proposed use cases that are considered higher-risk.

EIS will specify standards for the following:

- Generative AI model and system card disclosure
- Use of data for model training
- User disclosure
- User feedback
- Human in the loop documentation
- Baseline testing and monitoring
- Additional requirements for higher-risk uses

5. Agency AI Adoption Plans

Agencies will create and keep updated an AI Adoption Plan for their agency that is inclusive, action-oriented and adaptable.

6. Statewide AI Registry

The State CIO or their designee will oversee the creation and maintenance of a statewide registry of all AI-using applications in use or development by state agencies, including details on their uses, as part of an enterprise application portfolio management system. This registry will serve as the single source of truth for application use inventory, monitoring, and risk.

7. Security

Secure Environments: To manage supply chain and infrastructure risk, all AI development, training, and hosting will be limited to approved, secure environments, including EIS Data Center Services facilities or other GovRAMP-certified platforms.

Integration with Cybersecurity: AI risk monitoring will be integrated into the state's Security Operations Center (SOC) and incident response plans.

8. Unauthorized Usage

Agencies are accountable for AI usage by employees for work-related purposes, and work usage by state employees of unapproved AI systems is prohibited. EIS will monitor enterprise AI usage to identify unapproved or high-risk use, evaluate associated security and operational risks, and make recommendations to the State CIO to limit or address unapproved usage.

9. Workforce Development and Training

An informed state workforce is essential to the responsible and effective adoption of AI. A statewide training program will be developed and maintained by EIS in collaboration with the Department of Administrative Services and include AI literacy and risk awareness.

10. Transparency and Public Reporting

EIS, with review by the Enterprise AI Advisory Committee, will publish an annual report on AI use in state government based on the State AI Registry. EIS will also maintain a publicly accessible website that provides these annual reports as well as Oregon AI transparency disclosure information for all higher-risk systems.

11. Sandbox for Custom Development and AI Model Installation

Agencies developing custom AI-powered software systems, including in Azure Foundry and Copilot Studio, must do so in a sandbox or development environment approved by EIS. Customers of the State Data Center are prohibited from installing, deploying, or operating AI models within EIS Data Center Services facilities without prior written approval from EIS. EIS will provide guidance and procedures for agencies to follow to support this policy.

DAS DEPARTMENT OF ADMINISTRATIVE SERVICES Statewide Procedure	**NUMBER** ###-##-###	**SUPERSEDES**
	EFFECTIVE DATE **REVIEWED DATE**	**PAGE NUMBER** Pages 1 of 2
DIVISION Enterprise Information Services (State CIO)	**REFERENCE** Responsible Usage of Artificial Intelligence Policy	
POLICY OWNER Privacy and Artificial Intelligence		
SUBJECT Responsible AI Usage Procedure	**APPROVED SIGNATURE** _____ Terrence Woods, State Chief Information Officer	

ENTERPRISE information services

PURPOSE

This procedure establishes the core activities agencies will use to implement the Statewide Responsible Usage of Artificial Intelligence policy.

APPLICABILITY

This procedure applies to Executive Branch state agencies, boards, and commissions under the authority of the State Chief Information Officer that implement, acquire, configure, test, pilot, publish, operate, or materially modify artificial intelligence systems or uses in the conduct of state business.

FORMS/EXHIBITS/INSTRUCTIONS

The following supporting materials may be required or used under this procedure:

- Attachment A: Agency AI Adoption Strategy Template and Instructions
- Attachment B: AI Risk Management Framework
- Attachment C: AI Use Form
- Attachment D: User disclosure
- Attachment E: User feedback
- Attachment F: Human in the loop documentation
- Attachment G: Higher-risk use cases requirements

PROCEDURE

RESPONSIBILITY	STEP	ACTION
Create an AI Adoption Plan for the Agency		
Agency	1	Establish an internal process for creating an AI Adoption Plan following the AI Adoption Strategy template (Attachment A) provided.
Agency	2	Create an AI Adoption Plan that is aligned with and/or included within the agency's strategic/business plan, IT strategic plan, and data governance plan.
For proposed new uses of AI, evaluate risks and submit the AI Use Form as part of the Information Technology Investment Oversight Policy and Procedure		
Agency	1	Review the proposed artificial intelligence use through the agency's internal governance, risk evaluation, and approval processes.
Agency	2	Apply the AI Risk Management Framework (Attachment B) to your potential AI use case. Refer to the questions in the AI Use Form (Attachment C), to identify risks and answer questions around safeguards, testing, monitoring, and oversight.
Agency	3	Ensure the proposal has the following requirements and/or standards identified: • Generative AI model and system card disclosure • Use of non-public data for model training prohibited • User disclosure (Attachment D) • User feedback (Attachment E) • Human in the loop documentation (Attachment F) • Additional requirements for elevated-risk or high-risk uses (Attachment G)
Agency	4	Complete and submit the AI Use Form (Attachment C) as supplemental form to IT investment intake (project or non-project.)
EIS	5	Initial review by EIS. For use cases categorized as higher-risk, EIS will prescribe additional requirements. Refer to Attachment G for additional information.
Agency	6	Complete additional requirements for higher-risk use cases, if required by EIS.
EIS	7	For higher-risk cases, State CIO (or their designee) must approve.

ENTERPRISE information services

Agency AI Adoption Strategy Template and Instructions

(Attachment A)

Contents

Overview

Why This Document Exists

Artificial Intelligence introduces new capabilities and new risks. This document aims to help agencies adopt AI deliberately, safely, consistently, and inclusively.

Enterprise Information Services' Responsible Usage of Artificial Intelligence policy directs agencies to develop their own AI Adoption Strategy. This strategy should:

1. Set intent and boundaries, not technical configurations
2. Establish who decides, who is accountable, and how risk is managed
3. Enable fast adoption of low-risk, high-value AI uses while putting structure around higher-risk uses.

The strategy should be treated as:

- A living document, reviewed annually or as appropriate;
- A decision guide for leadership and staff;
- A reference point during audits, inquiries, or incidents.

How Agencies Should Approach Completing Their Adoption Strategy

This document provides a guide to the ten major recommended components of an agency's AI Adoption Strategy, which should be integrated with the agency's overall strategic plan, IT Strategic Plan and data governance plan. These components are as follows:

1. Executive Summary: describing how the strategy relates to the agency's overall strategic plan, IT Strategic Plan and data governance plan.
2. Guiding Principles: address both agency principles and those established by the Responsible AI Usage Policy.
3. Governance and Oversight describing how AI use is overseen at the agency
4. An AI Use Case Intake and Evaluation Process: describing how AI use cases are sourced, evaluated, and prioritized.
5. Data Management and Classification: describing priorities around data management and classification, which are fundamental for the safe and responsible use of AI.
6. Tool Adoption: explaining the agency's approach to adoption and management of specific AI tools at the agency, from Copilot to custom software.

7. A Training and Employee Engagement Strategy: describing how employees will be adequately trained on AI tool use and where to go when they have ideas or questions about the use of AI.
8. Human Oversight and Accountability: describing how the agency will ensure human review of AI-generated content.
9. Transparency: describing how the agency will support AI disclosure to both employees and the public.
10. Success Metrics and Evaluation: describing how the agency will measure and evaluate AI usage, and the relative success of its AI Adoption Strategy.

In developing their AI Adoption Strategy, agencies should align with and build on existing plans and governance, including Data Governance, IT Governance, and the IT Strategic Plan, while recognizing that AI Strategy is a business-wide effort.

NOTE: References included throughout are intended as **Guidance Resources**, not mandatory citations. They are intended to help agencies make informed decisions and align with statewide direction.

Key Documents to Support Plan Development

The following documents are recommended for consultation and, where applicable, integration with your AI Adoption Strategy:

Enterprise resources:

- Statewide Responsible AI Usage Policy (107-004-###)
- Responsible AI Usage Procedure and associated attachments, including:
  - AI Risk Management Framework
  - AI Use Form
  - User disclosure
  - User feedback
  - Human in the loop documentation
  - High risk use cases requirements
- Copilot General Usage Policy
- NIST AI Risk Management Framework (AI Risk Management Framework | NIST)
- Final Recommended Action Plan of the AI Advisory Council (State Government Artificial Intelligence Advisory Council Final Recommended Action Plan)
- Agency Guide to IT Governance
- Information Asset Classification Policy (107-004-050)
- Data Governance Policy (107-004-160)
- M365 Hub and Prompt Gallery

Agency-specific resources:

- Agency Strategic Plan and IT Strategic Plan
- Agency Governance Charters
- Agency Performance and Outcome Metrics
- Agency Learning and Development Policies
- Agency Workforce Training Requirements
- Agency AI policy or Copilot use guidance

Adoption Strategy Key Components

This document outlines the ten recommended components of an agency AI Adoption Strategy and explains how each supports a practical, responsible, and organized approach to AI adoption. Agencies should use these components to build a strategy that aligns with agency goals and is integrated into the agency's broader IT Strategic Plan.

1. Executive Summary

The Executive Summary establishes the agency's strategic rationale for adopting artificial intelligence. It ensures that AI use is grounded in mission outcomes, public value, and trust, rather than driven by technology availability or external pressure.

This section should clearly communicate why the agency is pursuing AI, what problems or challenges AI is intended to help address, and what boundaries the agency is placing on AI use. It should be understandable to non-technical leadership, oversight bodies, and the public.

Helpful Resources:

- Statewide Responsible AI Usage Policy
- Agency Strategic Plan

2. Guiding Principles for Responsible AI Use

This section establishes the values-based foundation that guides all AI use within the agency. Policies and procedures define minimum requirements, but guiding principles exist to support sound judgment in situations where rules alone are insufficient. The Statewide Responsible AI Usage Policy contains guiding principles for the state that agencies may wish to supplement or define further.

AI technologies evolve rapidly, and not every potential use case can be anticipated in advance. These guiding principles are intended to help agency decision makers

consistently determine whether an AI use is appropriate, ethical, and aligned with the agency's mission, not merely whether it is technically feasible.

Helpful References

- Statewide Responsible AI Usage Policy
- Final Recommended Action Plan of the AI Advisory Council
- NIST AI Risk Management Framework (GOVERN function)

3. Governance and Oversight

This section identifies decision-making authority, accountability, and the governance structures needed to support responsible AI adoption. AI use should remain business-led, with IT, security, data, and AI specialists providing essential expertise and risk guidance. Effective governance requires both strategic oversight and operational review. In addition to executive or leadership oversight, organizations should establish an operational governance mechanism, such as a standing subcommittee or working group, to review proposed AI use cases, ensure alignment with this strategy, and escalate matters that warrant higher-level review.

Agencies with mature IT governance may already have an established process for reviewing agency IT initiatives. Some agencies may also have established a policy group and/or a data governance group. There is no one-size-fits-all approach to this as agency needs differ considerably.

<u>Two Recommended Layers of Governance</u>

1. Strategic AI Governance
    - Sets bounds and expectations for AI use across the agency.
    - Ensures alignment with:
        - Agency priorities and risk tolerance.
        - Statewide policies and guidance.
    - Typically led by the agency's executive IT governance body (e.g., IT Governance Committee).
    - Responsibilities:
        - Approve agency AI strategy and guiding principles.
        - Define escalation pathways for higher-risk AI uses.
        - Ensure compliance with statewide policy and enterprise requirements.
2. Operational AI Governance
    - Applies strategic guidance to individual AI use cases and tool requests.
    - Could be implemented through an AI Governance Subcommittee or working group under the IT Governance Committee.

- Responsibilities:
  - Establish process for sourcing AI use cases from across the agency
  - Review AI use cases for alignment with agency strategy.
  - Assess risk factors (data sensitivity, public impact, policy implications).
  - Escalate decisions when risk or impact warrants executive review.
  - Coordinate with data governance and policy groups as needed.

Key Components of Agency AI Governance

- Business-Led Decision Making: AI decisions are programmatic and policy-driven, supported by technical experts but owned by business leadership.
- Accountability: Every AI use case should have a designated business owner responsible for outcomes.
- Diverse Review: Governance bodies should include representatives from multiple divisions and disciplines to ensure balanced perspectives.
- Risk Documentation: Agencies should document risks, impacts, and mitigation strategies for each AI use case.
- External Feedback: Policies should include mechanisms to consider feedback from stakeholders and the public for high-impact AI uses.
- Third-Party Risk Management: Address risks associated with vendors and external AI systems.

Operational Workflow

- Intake and review of AI use cases before implementation.
- Streamlined review for statewide-approved tools (e.g., Microsoft Copilot), but still documented for oversight.
- Escalation of high-risk or sensitive data uses to executive governance.
- Submission to EIS for enterprise security and alignment review after agency approval.

Benefits

- Promotes responsible, transparent AI adoption.
- Maintains clear separation of strategic and operational roles.
- Builds trust internally and externally.
- Enables rapid adoption of low-risk AI while managing higher-risk uses effectively.

The National Institute of Standards and Technology (NIST) in its AI Risk Management Framework has a detailed section on the governance of AI. Below are some key

components of an AI governance framework that should be included in the strategy and applied dependent on use case:

- Govern 2.3: Executive leadership of the organization takes responsibility for decisions about risks associated with AI system development and deployment.
- Govern 3.1: Decision-making related to mapping, measuring, and managing AI risks throughout the lifecycle is informed by a diverse team (e.g., diversity of demographics, disciplines, experience, expertise, and backgrounds).
- Govern 4.2: Organizational teams document the risks and potential impacts of the AI technology they design, develop, deploy, evaluate and use, and communicate about the impacts more broadly.
- Govern 5.1: Organizational policies and practices are in place to collect, consider, prioritize, and integrate feedback from those external to the team that developed or deployed the AI system regarding the potential individual and societal impacts related to AI risks
- Govern 6.1: Policies and procedures are in place that address AI risks associated with third-party entities, including risks of infringement of a third party's intellectual property or other rights.

Helpful References

- Statewide Responsible AI Usage Policy (107-004-###)
- Agency Guide to IT Governance
- Agency Governance Charters
- NIST AI Risk Management Framework and NIST AI RMF Playbook (particularly the GOVERN function)

4. Use Case Intake and Evaluation Process

This section establishes how the agency identifies, documents, evaluates, and approves artificial intelligence use cases before they are implemented. Its purpose is to ensure that every use of AI within the agency is intentional, transparent, and appropriately reviewed.

AI use cases vary widely in complexity and impact. Some uses, such as staff productivity tools, are low risk and widely approved at the enterprise level. Others may involve sensitive data, public interaction, or decision support that affects individuals or communities. This section exists to ensure that all AI uses are visible to leadership, consistently evaluated, and aligned with agency values and statewide policy.

Review and evaluation are not intended to slow adoption or discourage innovation. Instead, they provide a shared understanding of what AI is being used for, how it works, what risks it introduces, and who is accountable. Treating intake and evaluation as a standard operating

practice, rather than an exception, reduces confusion, builds trust, and allows the agency to scale AI use responsibly over time.

Agencies are encouraged to incorporate an AI risk assessment into their internal governance process early, before a proposed use case reaches final review or approval for use. The statewide AI risk assessment is available for this purpose and can help document the proposed use case, including risks, safeguards, data considerations, human oversight, and any needed disclosures.

With the exception of AI tools explicitly approved by EIS for enterprise use, including Copilot and M365 Copilot, all AI use cases must undergo EIS review and evaluation prior to use.

Helpful References

- Copilot General Usage Policy
- AI Use Form
- Statewide Responsible AI Usage Policy (107-004-###)
- NIST AI Risk Management Framework (MAP and MEASURE functions)
- Agency data governance

5. Data Management and Classification

Managing data requires establishing clear lines of authority over that data. Agencies should identify a Data Owner, what rules apply, and how they are enforced. In addition agencies should establish and deploy strong asset classification procedures, both for policy compliance and to facilitate AI use. This may include providing agency-specific training and education on proper data classification techniques and auditing of data stores. Strong data controls prevent many AI-related risks before they occur and reduce reliance on downstream mitigation. In addition, clear data classification, quality, stewardship, and management practices help ensure AI systems are built on reliable, appropriate, and well-understood information, reducing the risk of poor outputs and avoidable harm. Ultimately, AI outputs are only as strong as the quality, relevance, and governance of the data provided.

Helpful References

- Information Asset Classification Policy
- Data Governance Policy

6. Tool Adoption

This section explains how the agency adopts and manages AI tools as organization-wide capabilities, distinct from individual use cases. Tools introduce shared functionality, data flows, and risk considerations that extend beyond any single application.

This section helps agencies avoid unmanaged tool sprawl, ensure consistent security and privacy controls, and align tool adoption with workforce readiness and statewide standards. Agencies should prioritize statewide approved tools and clearly define how new tools are evaluated and introduced.

Helpful References

- Agency Workforce Training Requirements

7. Training and Employee Engagement Strategy

This section ensures the agency's workforce is prepared to use AI tools responsibly, effectively, and in alignment with agency expectations. AI risk often stems not from the technology itself, but from misunderstanding, misuse, or unclear expectations, making clear and appropriate training essential to safe and successful adoption.

Training should not be treated as a one-time or generic requirement. It should be purposeful, role-appropriate, and aligned to the specific tools and use cases involved, with updated training provided as tools evolve or are used in new ways so staff understand capabilities, limitations, data-handling requirements, human oversight responsibilities, and escalation pathways.

Workforce enablement is a shared responsibility across business leadership, governance bodies, and staff. As part of responsible implementation, governance review of AI use cases should consider whether users are adequately trained and whether an appropriate training approach has been defined.

Helpful References

- Agency Learning and Development Policies
- Copilot General Usage Policy
- Workday Training "AI For Public Professionals"
- Microsoft Copilot Resources:
    - M365 Hub
    - Prompting Gallery

8. Human Oversight and Accountability

This section establishes human-in-the-hoop (HITL) oversight as the baseline expectation for all AI use within the agency. Artificial intelligence may assist with analysis, recommendations, or content generation, but it does not operate independently of human responsibility. Human judgment, authority, and accountability must remain central to any AI-supported process.

This section exists to ensure that AI systems are designed, implemented, and operated with intentional human involvement, particularly where AI outputs influence decisions, actions, communications, or services. Human oversight is not an afterthought or an informal practice; it is a design requirement that should be defined in advance, documented, and maintained throughout the AI system's lifecycle.

Human-in-the-loop expectations should be considered early in the design of an AI use case. Governance bodies rely on this information to understand how AI outputs are reviewed, how errors are identified and corrected, and where responsibility ultimately resides. Without clearly defined human oversight, AI use cases should not proceed.

Helpful References

- NIST AI Risk Management Framework (GOVERN and MEASURE functions)
- Statewide Responsible AI Usage Policy (107-004-###)

9. Transparency & Public Trust

This section establishes how the agency promotes transparency and public trust in its use of artificial intelligence. AI systems can influence how services are delivered, how information is provided, and how decisions are supported. As a result, the public and agency staff should have confidence that AI is being used appropriately, responsibly, and with clear accountability.

Transparency is not limited to public disclosure alone. It also includes ensuring internal stakeholders understand when AI is being used, for what purpose, and under what safeguards. This section exists to help agencies think intentionally about when disclosure is necessary, what information should be shared, and how concerns or questions can be raised and addressed.

Transparency expectations should scale with risk and impact. Public-facing or decision-support AI systems require a higher degree of openness than internal productivity tools. Governance bodies should have visibility into how transparency obligations are met as part of AI use case review and approval.

Helpful References

- Statewide Responsible AI Usage Policy

10. Success Metrics and Evaluation

This section ensures that artificial intelligence adoption is evaluated based on outcomes, value, and alignment with agency goals, rather than on activity or novelty. AI use should demonstrably support the agency's mission while remaining within established risk tolerance and governance expectations.

This section exists to reinforce that AI adoption is not static. As tools, use cases, and risks evolve, agencies should periodically reassess whether AI is delivering intended benefits, whether safeguards remain effective, and whether adjustments are needed. Regular review supports continuous improvement and prevents AI systems from becoming outdated, misaligned, or insufficiently governed.

Success measures and strategy review are also critical for executive accountability. Leadership should be able to understand where AI is adding value, where risks are emerging, and whether the agency's AI strategy remains fit for purpose.

Helpful References

- Agency Performance and Outcome Metrics
- NIST AI Risk Management Framework (MONITOR function)

ENTERPRISE information services

Artificial Intelligence Risk Management Framework (Attachment B)

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 by all state employees. Specific recommendations on the use of Copilot are provided in a separate document. Beyond Copilot, 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 added AI components, as well as large custom-built projects. Specific definitions of generative and agentic AI are provided in the Statewide Responsible AI Usage Policy, but in simple terms: any AI technology that learns, adapts, or infers beyond what humans explicitly programmed requires approval by EIS.

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 ITI 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 State of Oregon AI Advisory Board[1] as well as the 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 & 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 & 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 & 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, & Output Review	Human review of outputs, curated and authoritative data sources, validation against laws and policies, source referencing where feasible, periodic quality reviews and audits
5	Bias, Fairness, & Inclusion	Bias and fairness assessments, review of data representativeness, explainability for people-impacting uses, appeal or challenge mechanisms, DEIA or subject-matter review
6	Reliability, Oversight, & 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, & Resilience	Performance and availability monitoring, capacity planning, backup and recovery procedures, remediation of security vulnerabilities
8	Transparency, Accountability, & 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 & Usability	Accessibility standards compliance, clear and understandable outputs, multilingual support where appropriate, user testing with diverse users
10	Escalation & Continuous Improvement	Defined escalation conditions, user feedback collection, periodic reassessment of safeguards, authority to modify or suspend AI use
11	Decision Authority & Automation Limits	Human approval for rights-impacting decisions, prohibition on fully automated determinations, explainability of AI-influenced decisions
12	Records, Retention, & 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 ITI 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 & 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 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 & 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 IP. Gartner explicitly flags third-party sharing/privacy risk and Info-Tech pushes AI compliance strategy; IP 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 & 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, & 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, & 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.
   - 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, & 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, & 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, & 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 & 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 & 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 & 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, & 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 & 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 oversee. 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 resident-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	Resident-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

ENTERPRISE information services

Information Technology Investment (ITI) Form
Attachment C: **AI Use Form**

This form should accompany the ITI submitted for any investment in which the use of Artificial Intelligence (AI) is intended to be utilized. For in-flight IT project investments under oversight, a new ITI is not needed in addition to this form. This form will serve as a supplement to the IT project investment ITI submission.

*All sections must be filled out to be accepted. If left blank, the form will be returned to the agency.

1. Can the AI be turned off or disabled? ☐ Yes ☐ No
2. Audience: Is the AI functionality designed to be used outside of state employees or contracted resources? ***Note:*** If the audience will be the public, additional detailed information will be required. Please consult the Privacy & AI office at EIS for further discussion.
3. AI Disclosure mechanism: Specify how you are disclosing to users that they are interacting with an AI tool.
4. AI System outputs: Define the outputs generated by the AI System (scores, classifications, drafts, actions, etc.).
5. Rights or Safety-Impacting AI processing: Will the tool be used for any of the following purposes? Check all that apply or acknowledge that none will.
☐ Emotion recognition or sentiment analysis ☐ Scraping images from databases ☐ Predictive analysis of crime ☐ Biometric categorization or identification (including facial recognition) ☐ Acting as a safety component in a system or products ☐ Informing safety decisions (i.e. water treatment ratios, load-bearing specifications for bridges) ☐ 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 ☐ The administration of justice, law enforcement or immigration

Note: If you checked any of the boxes above, additional detailed information will be required. Please consult the Privacy & AI office at EIS for further discussion.

☐ The AI tool will not be used for any of the above purposes.

6. Data sources: Identify the data sources, type, classification (level 1-4), and use of regulated data such as HIPAA, FERPA or CJIS as an input to the AI system.

Note: If you are planning to use level 3 or regulated data, additional information will be required. Please consult the Privacy & AI office at EIS for further discussion.

7. Risk management, including privacy and ethics: Please indicate which risk management strategies are planned to address potential risks associated with AI tool use (see t*he* AI Risk Management Framework for examples).

a) What are the identified risks?
b) Are there specific risks related to privacy, accuracy, bias and/or ethics? Please describe.
c) What specific measures will be implemented to mitigate each of these risks?
d) How are you proposing to test the system before deployment?
e) How are you auditing results generated by the system?
f) What triggers would require additional testing or risk re-evaluation?
g) Describe how user feedback on each AI interaction is captured?

8. Technical details: Large Language Model used (name and provide link to publicly posted LLM System Card or Model Card)

9. Human-in-the-loop: Describe how humans are reviewing AI created content, and what the specific written policies or controls are to ensure this.

10. Defining success: Define how success of the AI will be measured.

11. AI processing categories: Check all that apply

☐ Transcription - speech-to-text from audio or video	☐ Biometric / Face Recognition – Identification or verification using facial or other biometric data
☐ Optical Character Recognition (OCR) – extracting text/structure from scanned or imaged documents	☐ Surveillance / Tracking – Monitoring individuals, locations, or behavior patterns over time (incl. computer vision on cameras)
☐ Translation/Localization – Converting content between languages or dialects	☐ Redaction / De-Identification – Removing or obfuscating personal or sensitive information
☐ Text Summarization – Condensing documents, emails, chats, or records	

☐ Content Drafting/Writing Assist – Drafting or rewriting emails, letters, reports, responses ☐ Q&A / Knowledge Assistant – Answering questions using information from approved or official sources ☐ Classification / Tagging – Assigning categories, topics, labels, or dispositions to items ☐ Routing / Triage – Directing cases, tickets, or messages to queues/teams based on AI output ☐ Recommendation / Personalization – Suggesting actions, services, content, or options to users ☐ Eligibility / Benefit Determination – Scoring or deciding eligibility, benefit level, or program fit ☐ Risk Scoring / Prioritization – Producing risk or priority scores that influence treatment, inspection, enforcement, etc. ☐ Forecasting / Trend Analysis – Predicting demand, volumes, budgets, or other future values ☐ Sentiment / Emotion Recognition – Detecting sentiment, tone, or emotional state from text, audio, or video	☐ Data Extraction / Structuring – Pulling fields, entities, or tables out of unstructured text or documents ☐ Code Assist / Dev Assist – Generating, editing, or explaining code, scripts, or queries ☐ Image / Video Generation – Creating or heavily altering images or video from prompts or source material ☐ Audio / Voice Generation – Creating synthetic speech, voices, or audio ☐ Simulation / Scenario Generation – Generating hypothetical scenarios, test cases, or synthetic data ☐ Policy / Rule Drafting – Drafting policies, contracts, rules, or legal-style text ☐ Decision Automation – Directly issuing decisions or actions (approve/deny, auto-closure, auto-enforcement) without required human review ☐ Anomaly / Fraud Detection – Flagging unusual or suspicious activity for review ☐ Other AI Processing – Free-text description when none of the above fit

PPM Tool Project ID# (for in-flight projects under oversight):_____

Definition of LLM Model Card or System Card: A structured document for a trained AI model that provides information on intended use, evaluation domains, benchmark and task performance, testing methodology, limitations, known failure modes, usage constraints, and relevant safety, security, privacy, robustness, or other risk-related findings.

Higher-risk AI Uses: If your audience is Public, you are planning to use level 3 or regulated data, or you checked the box above for a rights-impacting or safety-impacting use of AI, your use is likely categorized as higher risk. Please contact the Privacy and AI office at EIS for further discussion.

Attachment D: User disclosure mechanism

Agencies must disclose to users when they are interacting with AI-generated content.

Disclosure is "no surprises, clear expectations" and it should be user interface (UI)-proximate and role-accurate. For low-risk internal AI knowledge chatbots, disclosure should be a lightweight user interface label so staff understand they're looking at AI-assisted content. For higher-risk or public-facing contexts, disclosure becomes a user-rights feature: it tells people AI is involved, what role it plays (and does not play), and how to request explanation or human review. Done right, it prevents accidental deception and makes downstream accountability possible without heavy process.

Disclosure is about expectation-setting, not technical explanation. If someone would reasonably feel misled without disclosure, disclosure is required.

Regular, Low-Risk Use Cases:
Examples: Internal tools, drafting, summarization, research support; no rights-impacting decisions.

What to Disclose: That AI was used as assistance, not authority.

Where:

- **In the tool UI**, not buried in policy:
  - Small label, tooltip, or footer note.
- **Content that is external-facing** needs disclosure.

When:

- At the moment the user consumes the output (not later or elsewhere).

How: Use one of these plain-language patterns:

Use	Potential message
AI-generated content	"AI-assisted draft. Reviewed by staff." or "Generated with AI and edited by [team/role]." "The script you will hear throughout this training is read using an AI-generated voice. The content itself was written and reviewed by [insert agency initials here] staff."
Chatbot agent	"AI can make mistakes. Check citations to verify."
Document review or processing agent	"AI can make mistakes. Please double-check all output." "Generated with AI and edited by [team/role]."

What Not to Include:

- No model names.
- No training data claims.
- No legalistic explanations.

Governance gut check:
Ask: “Would a reasonable person feel misled if they later learned AI was involved?”
If yes → disclose.

For Higher-Risk Use Cases:
User disclosure requirements for higher-risk AI use cases, including the ability for a user to escalate to a human in certain cases, is in development. Contact EIS for more information.

NIST AI Risk Management Framework Mapping

- GOVERN: documenting and communicating AI risks and impacts more broadly (GOVERN 4.2) and setting risk management controls transparently (GOVERN 1.4).
- MANAGE**:** documenting residual risks to downstream users (MANAGE 1.4) and communicating incidents/errors to relevant AI actors, including affected communities (MANAGE 4.3).

Attachment E: User Feedback

The system should measure user satisfaction on each use and provide an opportunity for users of all types, including both state employees and the public, to provide feedback that is reviewed by a human.

Collecting and using user feedback is a critical part of a successful AI implementation. It's a good early-warning radar that can help the product owner identify issues quickly and also understand their scope.

What feedback to collect: Simple: thumbs up/down + "what was wrong?".

Where: Embedded in the tool (chatbot, document helper) or in the page footer.

When: After output is shown with a prominence that will achieve a reasonably high response rate.

How (minimum recommended fields):

- Thumbs up or down
- For Thumbs down:
  - Category (wrong, incomplete, unsafe, unclear, other)
  - Free text (please describe the issue with as much detail as possible. You can attach a file or screenshot if helpful)(optional)
  - Ability to attach one or more screenshots or files
  - Email address: You can contact me with further questions (optional)

What to do with the feedback:

- Product owner or delegate to review regularly and triage; have a process for rating issues by severity.
- Incorporate into continuous improvement workflow – data corrections, functionality updates, etc.

The gut-check: Before continuing use, the owner should be able to say: "If someone flags this as wrong or misleading, a human will actually see it, and repeated issues will change the system."

For higher-risk AI uses: Depending on the use case, for higher-risk uses of AI the user must be able to escalate issues to a human for review and response. Further requirements for user feedback on higher-risk AI uses is under development. Contact EIS for more information.

NIST AI Risk Management Framework Mapping

- GOVERN: collecting and integrating external feedback about impacts (GOVERN 5.1) and regularly incorporating adjudicated feedback into design/implementation (GOVERN 5.2).
- MANAGE: monitoring plans that include capturing/evaluating user input plus appeal/override (MANAGE 4.1) and continual improvement activities with regular engagement (MANAGE 4.2).

Attachment F: Human-in-the-loop documentation

Human review is a foundational principle for the responsible use of AI at the State of Oregon. Ensuring a human is in the loop is about accountability prior to distribution or decision. Human review is satisfied when a responsible individual confirms the output is factually correct, appropriate for use, and something they are willing to stand behind. This includes both sufficient review and approval of release of the content.

Documentation of human review for AI use cases is fundamental for the responsible use of AI. This includes the following:

- Clear requirements for human review of AI outputs from the system proposed for use. This should include what must be reviewed and who is approved to conduct the review.
- If deemed appropriate, a review log for tracking reviews, including the individual reviewer and the content reviewed.
- A further process for auditing or checking the review log, if deemed necessary for the given use case.

For many cases, human-in-the-loop control is satisfied through existing operational approvals — sending or signing off a document, giving an approval, or closing a ticket — rather than creating new logging systems or parallel audit trails.

Example: Low-Risk Use Case (internal, non-rights impacting)

What Gets Reviewed:
Most critical is to review key data points and conclusions:

- Names, dates, amounts, thresholds
- Policy references or citations
- Conclusions or recommendations (if present)

How to Review:
Reviewer must:

- Scan the output and **verify critical facts** that would cause harm if wrong.

Supporting human review:
An optimally-designed AI system will support effective human review by providing clear citations, traceable sources, and sufficient explanation of outputs to allow users to understand how conclusions or recommendations were generated.

The gut-check:
Before release, the reviewer should be able to honestly say:
“I am responsible for this. If this is wrong, it’s on me — and I’m okay with that.”

For use cases flagged as higher-risk:
More detailed identification of risks and human review requirements must be done to support any flagged use of AI, including concrete mitigations and safeguards.

NIST AI Risk Management Framework Mapping

- GOVERN: clear roles/responsibilities and oversight expectations (GOVERN 2.1), leadership accountability for risk decisions (GOVERN 2.3), and defining human-AI oversight configurations (GOVERN 3.2), with planned monitoring/review cadence (GOVERN 1.5).
- MANAGE: mechanisms to supersede/disengage/deactivate when outcomes don't match intended use (MANAGE 2.4) and post-deployment monitoring that includes appeal and override (MANAGE 4.1).

Attachment G: Additional requirements for use cases flagged as higher-risk

Higher-risk uses of AI require the approval of the State CIO or their designee. To support agency requests for this approval, the following are additional requirements that agencies will need to provide:

1. For use cases flagged due to use of level 3 data

Agencies may only deploy AI solutions for Level 3 data[1] in a manner consistent with Oregon law, information security policies, standards and principles, and consumer protection expectations, including prohibitions on unfair or deceptive practices. AI uses must comply with existing confidentiality, breach notification, and sector specific obligations (e.g., health, financial, criminal justice).

AI solutions handling Level 3 data must follow data minimization and purpose limitation; only the minimum necessary data for a defined, documented purpose may be processed, and secondary use is prohibited without explicit approval. Access to Level 3 data with AI capabilities must follow least privilege, need-to-know, and Zero Trust principles, with entitlements aligned to roles and reviewed regularly on a schedule acceptable to the agency's IT governance.

Authorization to use Level 3 data with AI has a high bar. The detailed specifications and requirements for agencies who wish to use level 3 data are still in development, but include the following:

1. **Internal Grounding**
   AI must be anchored to the agency's trusted internal data (Retrieval-Augmented Generation (RAG) architecture), where the model functions as a reasoning engine utilizing data from a secured internal search index. The model may retrieve relevant content during processing but must not retain or store that data after a session. Grounding must occur within secured internal environments or contracted private tenants, ensuring that sensitive information never leaves approved boundaries. Authoritative outputs must rely solely on curated internal sources such as approved datasets, procedures, and policies, avoiding dependencies on unvetted external content.

2. **Data Minimization**

[1] Information Asset Classification Policy 107-004-050.pdf

Only the specific data fields necessary for the AI task may be exposed to the model context window.

3. **Private Connectivity**
   All AI resources that access Level 3 data must be deployed using private networks and endpoints, ensuring that no traffic traverses the public internet. Connectivity between AI services and data stores must occur through virtual networks (VNETs) using private links or endpoints. Public ingress to AI services handling Level 3 workloads is strictly prohibited.
4. **Data Flow and Data in Transit**
   Data flows between AI components and Level 3 data sources must be explicitly documented, restricted to approved communication paths, and reviewed by the EIS CSS Security and Strategy & Design (S&D) architecture teams. Architecture diagrams reflecting these flows must be maintained as part of the system documentation record. Data in transit must use TLS 1.2+ and data at rest must be encrypted.
5. **Vendor Compliance**
   Vendors must agree to Zero Data Retention (ZDR) and provide third-party attestation/ certification (e.g. GovRAMP). Vendors must: (a) restrict data use to service delivery, (b) prohibit training on Level 3 data, (c) require breach notification within 24 hours per policy, and (d) grant audit rights.
6. **Development and Production Environments Attestation**
   Both development and production environments that handle Level 3 data must enforce the same security requirements.
7. **Logging and Auditability**
   Ensures all access, inference requests, and data interactions are logged, monitored, and available for audit by security and compliance teams.
8. **Data Loss Prevention (DLP)**
   DLP must be implemented by the agency unless a specific waiver is granted by EIS. DLP prevents unauthorized exfiltration or disclosure of sensitive data through model inputs, outputs, or training data.
9. **Emerging AI Risks**
   The Open Worldwide Application Security Project (OWASP) has defined a Top 10 list of risks for large language model applications. Agencies must provide documentation on their safeguards and mitigations for each of these risks.[2]

[2] OWASP Top 10 for Large Language Model Applications | OWASP Foundation

2. For use cases with a rights-impacting, safety-impacting, or flagged AI processing category

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 & 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 oversee. EIS will provide a template, and all approved disclosures will be published on the EIS and agency web sites. Additional detail is listed below in the next section.

3. For use cases with a public audience

The agency must provide an Oregon AI transparency disclosure for review and approval by EIS. This disclosure will give the public and project stakeholders a clear, plain-language snapshot of what the system does, what data it touches, and how it was tested and is being overseen. The disclosure is concise (2–3 pages), written for non-technical readers, and will be posted online to support transparency and trust.

Each Transparency Disclosure must include:

- Purpose and owner: program name, business goal, system owner, vendor (if any).
- Use context: where and how the AI is used, audience (internal/resident-facing), and whether it can affect rights or safety.
- Data sources and profile: Oregon data classification level (1–3) and any regulated data in scope (e.g., HIPAA/FERPA/CJIS/FTI).

- Model and hosting: model/service name and version, hosting pattern (on-prem, state cloud, GCC/GCC-H, vendor SaaS), and training-opt-out/data residency assertions.
- Human oversight and disclosures: the human-in-the-loop points, resident/user disclosures, appeals or correction paths.
- Testing snapshot (publishable): methods and summarized results for accuracy/quality, fairness/equity, privacy protections, security/red-team/adversarial testing, robustness/known limits, and pass/fail gates used prior to deployment.
- Monitoring and incidents: what is continuously monitored, rollback criteria, and how incidents are reported/handled.
- Change log: notable updates to models, data, prompts, guardrails, or controls with dates.

NIST AI Risk Management Framework Mapping

- GOVERN: documenting roles, accountability, oversight structures, and policy alignment for the AI system (GOVERN 1.1, 1.2), and supporting transparency practices through documented disclosures (GOVERN 4.1).
- MAP: describing intended purpose, use context, affected populations, and data characteristics, including potential rights, safety, or equity impacts (MAP 1.1, 2.1).
- MEASURE: summarizing testing and evaluation results for performance, bias, privacy, security, and robustness, including known limitations and validation methods (MEASURE 1.1, 2.1).
- MANAGE: documenting monitoring practices, incident response, change management, and control adjustments across the system lifecycle (MANAGE 1.1, 2.2).

Appendix: Level 3 Control Overlay

To use level 3 data with AI systems, the following specific controls must be applied to the AI Environment. Note that this list of controls is still in development.

Control Family/ OR ID	Requirement Description
AC (Access Control) -2(7)	**Just-in-Time Access:** Admin access to the AI model configuration plane must be time-bound (e.g., PIM).
	Separation of Duties: Developers who refine the "System Prompt" cannot have access to the Production Inference logs containing real Level 3 user data.
AU (Audit & Accountability)-3	**Prompt Logging:** Full capture of User Prompt + System Response.
	Exclusion: If data is CJIS/FTI, logging must be metadata only (User ID, Timestamp, Token Count) unless a compliant, encrypted logging vault is established.
SC (System & Comm) - 7	**Private Link:** The AI Model API endpoint must resolve to a private IP address within the State VNET. Public internet access to the model API is blocked.
SI (System Integrity) - 10	**Input Filtration:** Implementation of automated filters (e.g., Azure Content Safety, Bedrock Guardrails) to detect and block PII patterns or toxic content before it reaches the LLM.
Remote Access Encryption (AC-17(2))	All traffic between the Agency VNET and the AI Model must be encrypted using TLS 1.2 or higher. Public internet traversal is prohibited.
MP (Media Protection)	**Encryption:** Data in transit (TLS 1.2+) and data at rest (AES-256) within the vector database or search index used for grounding. Keys must be managed via State-controlled Key Vault.
Developer Security Testing / SA-11	Agency led "Red Teaming" results must be documented prior to production to test for jailbreak vulnerability.

Appendix: Definitions

Grounding: The process of connecting an AI model to verifiable, authoritative sources of information (e.g., agency policy documents) to reduce hallucinations and improve accuracy.

Hallucination: The production of erroneous or false content by a large language model that is presented with confidence.

Internally Grounded: A specific architectural requirement where AI solutions must rely on curated internal sources (policies, procedures, approved datasets) as the primary knowledge base and avoid reliance on unvetted external content.

Private Endpoints: A network interface that uses a private IP address from a virtual network to connect to an AI service, ensuring that traffic remains on the private network backbone and does not traverse the public internet.

Private Networks (VNETs): Logically isolated network environments within a cloud tenant that prevent public ingress, ensuring resources are accessible only through controlled, secure pathways.

Retrieval Augmented Generation (RAG): An architecture where the AI model does not rely solely on its training data but retrieves relevant information from a specific, trusted knowledge base (internal search index) before generating an answer.

Zero Data Retention: A contractual and technical configuration where the AI provider processes the prompt and generates the response in memory, but does not store, log, or use the data for model training after the session ends.