## Higher-risk use cases requirements

To support agency requests for approval of higher-risk use cases, 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 artificial intelligence (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 where the model functions as a reasoning engine utilizing data from a secured internal search index, and avoid dependencies on unvetted external content. 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.

2. **Data Minimization**
   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

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

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 Cyber Security Services security and Strategy and Design (S&D) architecture teams. Architecture diagrams reflecting these flows must be maintained as part of the system documentation record. Data at rest and in transit must use EIS-approved security protocols.
5. **Vendor Compliance**
   Vendors must agree to Zero Data Retention (ZDR) on AI processing 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. 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 |
|---|---|

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

| | |
|---|---|
| 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.<br><br>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. Additional details are 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: Additional 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.

**Revision History**

| Version | Date | Change Log |
|---|---|---|
| 1.0 | June 12, 2026 | Initial version |