

State Policy on Responsible AI Use

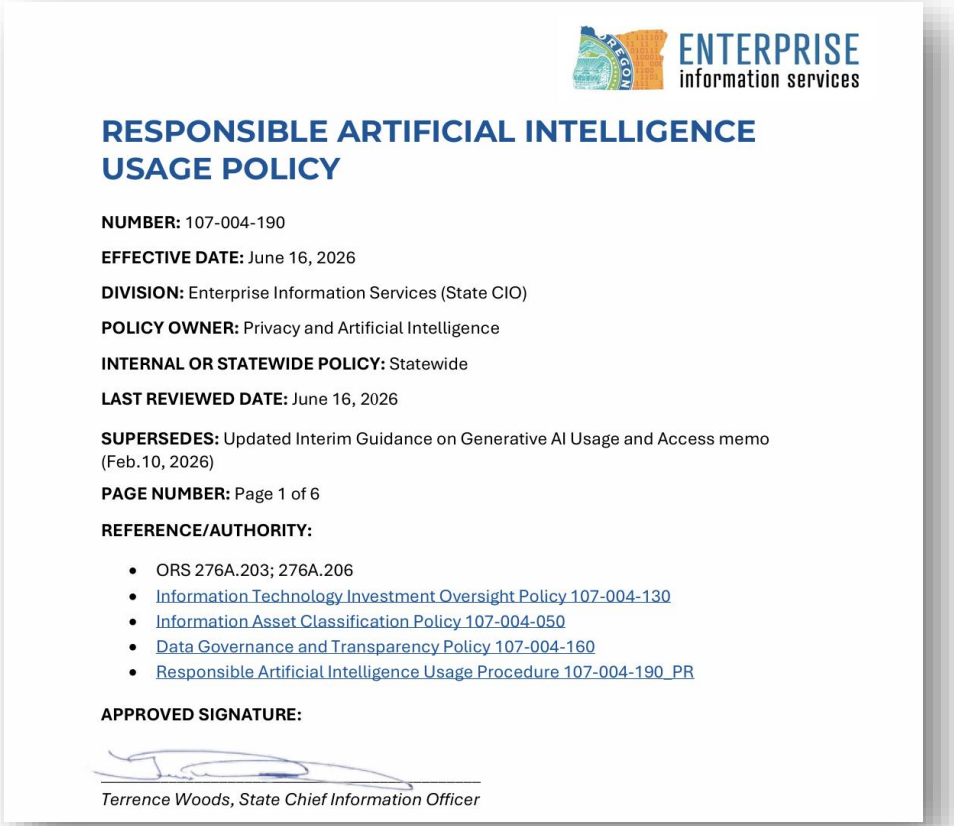
Data Equity Summit

Amy Paul, PhD
AI Policy and Safety Advisor

October 2026

Responsible AI Usage Policy – State of Oregon

- Statewide Policy Released in June, 2026
- Primary Goal: Governance that instills confidence and trust in the state's use of AI
- Multiple sections and supporting documents



ENTERPRISE information services

RESPONSIBLE ARTIFICIAL INTELLIGENCE USAGE POLICY

NUMBER: 107-004-190

EFFECTIVE DATE: June 16, 2026

DIVISION: Enterprise Information Services (State CIO)

POLICY OWNER: Privacy and Artificial Intelligence

INTERNAL OR STATEWIDE POLICY: Statewide

LAST REVIEWED DATE: June 16, 2026

SUPERSEDES: Updated Interim Guidance on Generative AI Usage and Access memo (Feb.10, 2026)

PAGE NUMBER: Page 1 of 6

REFERENCE/AUTHORITY:

- ORS 276A.203; 276A.206
- Information Technology Investment Oversight Policy 107-004-130
- Information Asset Classification Policy 107-004-050
- Data Governance and Transparency Policy 107-004-160
- Responsible Artificial Intelligence Usage Procedure 107-004-190_PR

APPROVED SIGNATURE:

Terrence Woods, State Chief Information Officer

Link to Policy and Supporting Docs: Enterprise Information Services : Artificial Intelligence : Privacy and Artificial Intelligence : State of Oregon

ENTERPRISE information services

Principles of Responsible AI

- Human Oversight and Review
- **Equity and Representation**
- Privacy and Confidentiality
- **Risk Management**
- Safety
- **User Experience, Disclosure, and Feedback**
- Workforce Preparedness and Understanding

Based on the Principles from the State Government Artificial Intelligence Advisory Council Final Recommended Action Plan February 2025.

AI Risk Management Framework

- **Goal**: Help agencies understand the risk involved with potential AI uses, and provide recommendations for mitigations and safeguards
- **Key Policy Documents**:
  - *AI Risk Management Framework*
  - *AI Use Form*
  - *Standards for: User Feedback, Disclosure, Human Review*
- **Details**:
  - AI Use Form serves as initial intake and helps determine what constitutes higher-risk use cases

Risk Mitigation For Every Proposed AI Use

- The AI Use form asks about risks and mitigations for every proposed use case as an initial assessment of risk.

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

Transparency: AI Model or AI System Cards

2. Core Principles and Structural Taxonomies

Canonical model cards embody five foundational principles: transparency, accountability, risk management, regulatory compliance, and FAIR (Findable, Accessible, Interoperable, Reusable) data practices (Mitchell et al., 2018, Everitt et al., 2024). Structurally, recommended section sequences include:

- **Model Details**: identity, version, architecture, developer, license, persistent ID, feedback channel.
- **Intended Use**: applications, users, non-applications (contra-indications), taxonomy, comparison to alternatives.
- **Factors**: axes of performance variability (demographics, hardware, input distribution).
- **Metrics**: statistical definitions, uncertainty quantification (e.g., 95% CI, bootstrap).
- **Evaluation Data**: datasets, preprocessing, representativeness, edge cases.
- **Training Data**: composition summaries, biases.
- **Quantitative Analyses**: unitary and intersectional performance disaggregation.
- **Ethical Considerations**: data sensitivity, potential harms, mitigations.
- **Caveats/Recommendations**: known failure modes, suggested deployment constraints.

Modern extensions may add sections on **Trustworthiness** (with NIST/EU-inspired subcategories: accountability, explainability, privacy, fairness, reliability, safety, security, transparency) and **Risk Environment and Management** (Kennedy-Mayo et al., 2024).

Model Cards for Reporting AI Models, *Emergent Minds,* updated March 2026

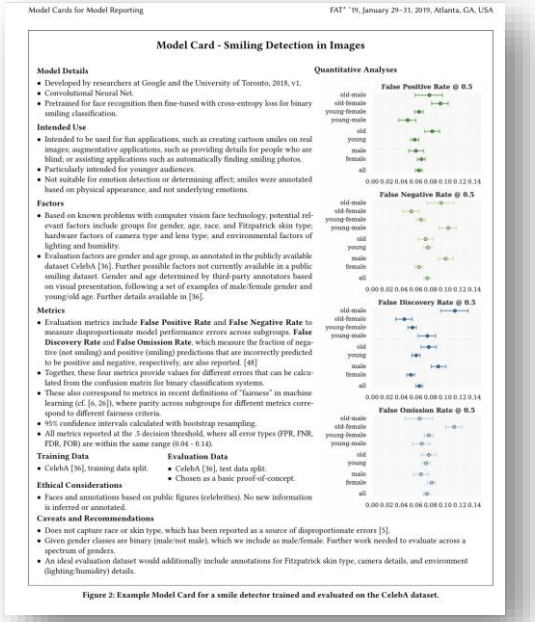
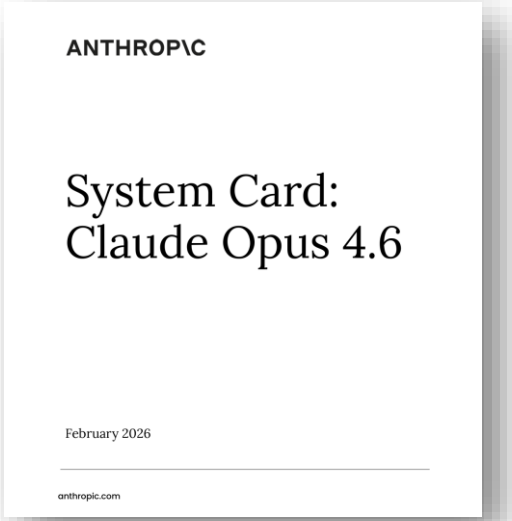


Figure 2: Example Model Card for a smile detector trained and evaluated on the CelebA dataset.

Model Cards for Model Reporting, 2018



Model system cards \ Anthropic

Higher-Risk Use Cases

Public Facing AI Systems	Use of Level 3 or Regulated Data	Rights or Safety Impacting AI Use Cases
• Chatbots for information retrieval on public services	• Internal AI-tools for analysis or retrieval of sensitive agency data	• Evaluation of a person in an employment or education setting • Benefit eligibility evaluation • Informing safety decisions on critical infrastructure • AI processing of emotion or biometrics

Example Risk Mitigation Approaches

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.

AI Risk Management Framework

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

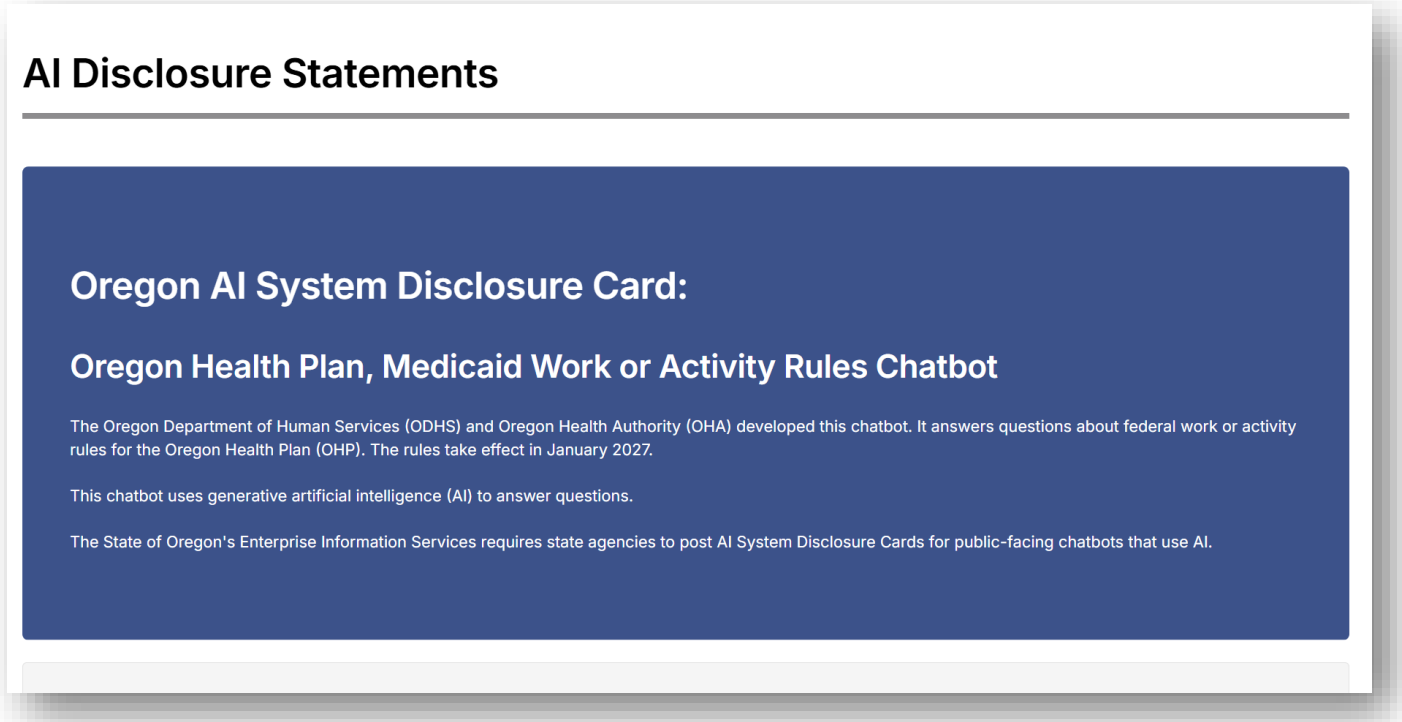
Additional Requirements for Rights and Safety Impacting Use Cases

- Detailed list of risks and mitigations
- Detailed testing plan describing how models have been evaluated for accuracy, fairness, equity, bias and safety, and security metrics
- Monitoring and incident preparedness
- Oregon AI Transparency Disclosure
- Involvement of primary users of and communities affected by AI systems in design and testing
- Considering independent, third-party review for higher-risk use cases

ENTERPRISE information services

Oregon AI Transparency Disclosure

- Greater transparency for public-facing systems
- Engagement of users and community organizations in testing and in disclosure design
- All led to improved product and more accessible disclosure



Oregon Department of Human Services : AI Disclosure Statements : Agency Information : State of Oregon

Agency-level AI Adoption Plans

- Guiding Principles
- Governance and Oversight
- Use Case Intake and Evaluation
- Data Management and Classification
- Tool Adoption
- Training and Employee Engagement
- Transparency and Public Trust
- Success Metrics and Evaluation


ENTERPRISE
information services
Thank you!
Team Email: AI@eis.Oregon.gov
Amy Paul, AI Policy and Safety Advisor – amy.paul@eis.oregon.gov