TLP Level: W – G – A - R 

INTERIM GUIDELINES FOR RESPONSIBLE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE

Introduction.

Generative Artificial Intelligence (GenAI) holds great potential to transform the Administrative, Education, Health, Natural Resources, Public Safety and Transportation sectors within the state government of Oregon. However, ensuring that GenAI solutions are used in a secure manner is essential to protect sensitive data, maintain public trust, and uphold regulatory compliance. This interim guide provides tailored recommendations for state government agencies to enhance the security of their AI initiatives.

The aim of this document is to responsibly integrate GenAI into the work of state agencies in a secure fashion, while ensuring it aligns with our values of service, equity, and inclusivity.

GenAI is a technology that excels in producing, editing, summarizing, and reshaping content such as text, images, audio, and video. It has recently become widely available for almost anyone to use, often at no direct cost, and its adoption by both individuals and organizations has been growing exponentially. While there is immense potential for improving the agencies' capacity and capabilities, using GenAI appropriately and effectively requires care and planning to address the risks that the use of AI will introduce for the state.

The guidelines in this document are based on the core principles laid out in the NIST AI Framework (Link provided in the Resources section). These are general-purpose guidelines, and it is important to recognize that more detailed guidance will be needed in specific areas.

This document serves as an initial framework for the responsible and ethical use of GenAI technologies for the Oregon state government. Recognizing the rapidly evolving nature of AI, these interim guidelines will be periodically reviewed and updated to align with emerging technologies, challenges, and use cases.

Definition.

"Generative artificial intelligence (GenAI) is the class of AI models that emulate the structure and characteristics of input data to generate derived synthetic content. This can include images, videos, audio, text, and other digital content". (Definition from NIST AI 600-1)

Trends of AI Use.

GenAI has started a new gold rush of tools being developed as companies explore all the different ways that this technology can be used. In this section we will list several of the most well-known tools and their specialties. Information from this section was provided by turing.com. (Link in references) It is worth noting that all free use AI have in their terms of service that they take users prompts/files to train and improve the GenAI. See Sensitive Data Section for more information.

- ChatGPT

  - The AI that started it all ChatGPT was the first GenAI to be released to the public. Its features are as follows:
    - Generates text that closely resembles human language.
    - Engages in conversations that feel natural and authentic.
    - Provides detailed and insightful answers to a wide range of queries.
    - Offers valuable support and suggestions for creative writing endeavors.
    - Demonstrates the ability to understand and respond in contextually relevant ways.
- Copilot
  - Microsoft's response to the creation of ChatGPT. This GenAI is fully integrated with the Microsoft environment which allows high level of data integration than can normally be achieved. Its features are as follows:
    - Fully integrated with all Microsoft products.
    - Security settings and access to data can be controlled through Azure Security Dashboard.
    - Contextually relevant communications.
    - Able to assist with a wide range of tasks and can adjust output based on the nature of the input provided.
    - Able to summarize documents and meetings into other formats (Can provide a text summary of a PowerPoint presentation or a Teams meeting.)
- Gemini (Formerly Bard)
  - Google's response to ChatGPT, this GenAI is fully integrated with the Google Account environment. The list of features is as follows:
    - Harnesses the power of LaMDA, a robust transformer-based model.
    - Limited-access waitlist catering to select customers in the US and UK.
    - Incorporates a user response rating mechanism.
    - Accessible via individual Google accounts.
    - Empowers users with assistance in software development and programming-related tasks.
- Scribe
  - An AI writing assistant designed to assist in document creation and academic writing its features are as follows:
    - Dedicated AI writing assistant.
    - Generates content in diverse styles and formats.
    - Summarizes articles, creates reports, and aids academic writing.
    - Assists in training and documentation.
    - Simplify documenting complex tasks.
- AlphaCode

  - An AI developed as a coding assistant. This system was designed to specialize in code creation, fixing bugs and aiding in efficient programming. It's stated features are as follows:
    - Utilizes advanced generative AI for coding assistance.
    - Supports various programming languages and paradigms.
    - Offers real-time code suggestions and bug fixes.
    - Provides code optimization solutions.
    - Assists in collaborative coding with shared suggestions.
- GitHub Copilot
  - A Microsoft product merging the services of GenAI and the GitHub platform to assist with the creation and development of code. Its key features are as follows:
    - Seamlessly integrates with popular code editors such as Visual Studio Code.
    - Generates not only code snippets but also explanations and contextual information to aid developers.
    - Provides instant and relevant suggestions for code completion, improving coding efficiency.
    - Offers a diverse range of programming language support, accommodating various projects.
    - Learns from usage patterns, adapting to individual developer preferences.
- Dall-E2
  - Crafted by OpenAI, this AI specializes in image creation from text prompts. Its features are as follows:
    - Generative AI model specializing in image synthesis.
    - Transforms textual prompts into intricate visual content.
    - Accommodates a wide range of image styles and genres.
    - Offers control over image attributes like composition and lighting.
    - Supports both high-level concept visualization and detailed imagery.
- Synthesia
  - This GenAI specializes in video creation via text inputs, its features are as follows:
    - The AI-powered platform converts text to videos efficiently.
    - Creates dynamic visuals, avatars, and scenes.
    - Automated voice synthesis for seamless audio.
    - Users can tweak visuals, text, and voice style.
    - Rapidly produce multiple videos simultaneously.
- Claude
  - AI assistant program
    - Handles extensive text data processing.
    - Engages in natural and fluent conversations.

        - Has multilingual proficiency, including common languages and programming languages.
        - Can handle complex workflows.
        - Adapts to user feedback for continuous improvement.
- Duet AI
    - GenAI that specializes in integrating with Google Applications. Currently in Beta
        - Code assistance for cloud users.
        - Image generation based on prompts.
        - Assists with create intelligent business applications.
        - Build workflows within Google Workspace using natural language commands.
- Cohere Generate
    - This GenAI specializes in crafting dynamic dialogue systems to enhance user engagement.
        - Delivers human like responses.
        - Can craft personalized email content.
        - Has been successfully deployed in various sectors, such as e-commerce and customer support.
        - Makes it easy to create conversational agents.
- Perplexity.ai
    - GenAI with a focus on citation of sources used to craft responses.
        - General Purpose AI
        - Less vulnerable to hallucinations
        - Able to summarize multiple sources together for its responses while clearly citing its sources.

This list is intended to showcase the variety of GenAI services that are available on the public market today and is not a complete list or an endorsement of any product or service.

Use Cases and Risk.

The introduction of GenAI introduces several specific types of risk for the governments computer systems and networks. In McKinsey & Companies “Implementing Generative AI with Speed and Safety” they identify four specifically:

- Increasingly sophisticated security threats resulting from using GenAI to augment the creation of Malware.
- The risk of unknown data exposures resulting from third-party vendors using GenAI to support their services.
- Malicious use such as the creation of deepfakes of agency personal or fabricating official documentation.
- The theft of Intellectual property (images, music, and text)

These risks have different potential consequences with the different use cases for GenAI. In general, the uses of GenAI can be consolidated into four use cases.

- Customer Automation Assistance (Chatbot etc.)
- Document generation (summarizing meetings etc.)
- Code Development
- Creative content (Video and Image Generation)

Each of these use cases generally has different primary risks, though some risks are common to all uses cases. The table below will layout the risks for each use case.

Gen AI Use Cases	Third Party Exposure	IP Infringement	Malicious Use	Security Threats
Customer Automation	Primary Risk	Tertiary Risk	Primary Risk	Primary Risk
Code Development	Secondary Risk	Primary Risk	Primary Risk	Primary Risk
Documentation Generation	Secondary Risk	Primary Risk	Primary Risk	Tertiary Risk
Creative Content	Tertiary Risk	Primary Risk	Primary Risk	Tertiary Risk

Primary Risk: This risk is of primary importance for this use case and should be covered in user training/education and considered in all risk management discussions.

Secondary Risk: This risk has a moderate chance of being exploited in this use case though there are others that are more likely. It should still be of some consideration.

Tertiary Risk: This is risk has a low chance of being exploited in this use case.

As agencies develop their plans for implementing GenAI into their processes it will be vital that they consider the new risks that GenAI will introduce to their environments. They will need to develop and work with the upcoming state framework on GenAI on the following points:

- Education and Training
- Safe Procurement
- Data Protection
- Data Governance
- Technology and Solutions

Principles.

The intention of the state of Oregon is to follow the principles in the NIST AI Risk Framework, which serve as the basis for the guidelines in this document. A foundational part of the NIST AI Risk

Framework is to ensure the trustworthiness of systems that use AI. The guiding principles from that document are as follows:

- **Valid and reliable**: Agencies should ensure GenAI use produces accurate and valid outputs and demonstrates the reliability of system performance. It should be assessed by ongoing testing or monitoring to confirm the systems are performing as intended.
- **Safe:** AI systems should "not under defined conditions, lead to a state in which human life, health, property, or the environment is endangered" (Source: ISO/IEC TS 5723:2022). Safe operation of AI systems is improved through:
    - responsible design, development, and deployment practices.
    - clear information to deployers on responsible use of the system.
    - responsible decision-making by deployers and end users; and
    - explanations and documentation of risks based on empirical evidence of incidents.
- **Fairness, and managed Bias**: Fairness in AI includes concerns for equality and equity by addressing issues such as harmful bias and discrimination. Standards of fairness can be complex and difficult to define because perceptions of fairness differ among cultures and may shift depending on application. Organizations' risk management efforts will be enhanced by recognizing and considering these differences.
- **Privacy and data protection**: AI should be used to respect user privacy, ensure data protection, and comply with relevant privacy regulations and standards. Privacy values such as anonymity, confidentiality, and control generally should guide choices for AI system design, development, and deployment. Privacy-enhancing AI should safeguard human autonomy and identity where appropriate.
- **Accountability and responsibility**: As public stewards, agencies should use generative AI responsibly and be held accountable for the performance, impact, and consequences of its use in agency work.
- **Transparency and auditability**: Acting transparently and creating a record of AI processes can build trust and foster collective learning. Transparency reflects the extent to which information about an AI system and its outputs is available to the individuals interacting with the system. Transparency answers "what happened" in the system.
- **Explainable and interpretable**: Agencies should ensure AI use in the system can be explained, meaning "how" the decision was made by the system can be understood. Interpretability of a system means an agency can answer the "why" for a decision made by the system, and its meaning or context to the user.
- **Public purpose and social benefit**: The use of AI should support the state's work in delivering higher quality services and outcomes to its residents.
- **Ownership and Vendor usage**: Unexpected security and privacy concerns can arise in the case of derived data, or in the mishandling of source data in generative AI systems. Clear ownership of the data, as well as full transparency over how it is being collected and used,

regardless of whether it is the source data, the derived data, or generative AI, should always be maintained.

Recommended Guidelines.

Fact-checking, Bias Reduction, and Review.

All content generated by AI should be reviewed and fact-checked, especially if used in public communication or decision-making. State personnel generating content with AI systems should verify that the content does not contain inaccurate or outdated information, or potentially harmful or offensive material. Given that AI systems may reflect biases in their training data or processing algorithms, state personnel should also review and edit AI-generated content to reduce potential biases. GenAI can also potentially suffer from hallucinations where it creates fictitious events to answer the users prompt. When consuming AI-generated content, be mindful of the potential biases and inaccuracies that may be present. (See NIST Special Publication 1270-Towards a Standard for Identifying and Managing Bias in Artificial Intelligence)

Disclosure and Attribution.

AI-generated content used in official state capacity should be clearly labeled as such, and details of its review and editing process (how the material was reviewed, edited, and by whom) should be provided. This allows for transparent authorship and responsible content evaluation. See NIST AI Risk Framework.

> **Sample disclosure line for general document creation**: This memo was summarized by Google Bard using the following prompt: "Summarize the following memo: (memo content)". The summary was reviewed and edited by [insert name(s)].
>
> **Sample disclosure line for code development**: (In the file header comments section) "This code was written with the assistance of ChatGPT3.5". The initial code was created using the following prompt: "Write HTML code for an Index.HTML page that says, 'Hello World'". The code was then modified, reviewed, and tested by the web development team at (Insert Agency Name).

Additionally, state personnel will need to conduct due diligence to ensure no copyrighted material, including images, audio, and video, is published without appropriate attribution or the acquisition of necessary rights. This includes content generated by AI systems, which could inadvertently infringe upon existing copyrights.

Sensitive or Confidential Data.

Agencies should not integrate, enter, or otherwise incorporate any non-public data (non-level 1 data) or information into publicly accessible generative AI systems (e.g., ChatGPT). The use of such data will likely lead to unauthorized disclosures, legal liabilities, and other consequences (see "Compliance with Policies and Regulations" section below).

If your agency has a usage scenario that requires non-public data to be used with generative AI technology, contact your agency privacy/security team.

Similarly, where non-public data is involved, agencies should not acquire generative AI services, enter into service agreements with generative AI vendors, or use open-source AI generative technology unless they have undergone a Security Design Review and received prior written authorization from the relevant authority, which may include a data sharing contract. Contact your agency privacy/security team for further guidance.

Compliance with Policies and Regulations.

Oregon State Law allows for a generally open information transparency for state information. To protect PII or other sensitive information the state is implementing sensitivity labels to mark information that is not suitable for disclosure. Only AI systems with which the Agency has entered into a data sharing/ enterprise level agreement with should be allowed access to data above sensitivity label Level 1 - Published via user entered AI prompts or any other means.

Proof of Concept or Pilot Creation.

Agencies should notify the Office of the State Technology Officer prior to deploying proof of concepts or AI pilot projects. The proof of concepts or pilots should not include sensitive data.

Generative AI Usage Scenarios and Dos and Don'ts.

Below are several usage scenarios alongside some do's (best practices) and don'ts (things to avoid):

When using GenAI, the four main uses are generally as follows:

- Text Generation/Editing
- Code Development
- Chatbot Development
- Image/Video Generation

When using GenAI in these areas here are some Do's and Don'ts to keep in mind.

- **Text Generation/Editing.**
  - **Do**
    - Review results to ensure the text is easily understandable and matches the intended reading level and check the rewritten documents for biases and inaccuracies.
    - Edit and review the document, label the content appropriately (see "disclosure and attribution" above), and remember that you and the state of Oregon are responsible and accountable for the impact and consequences of the generated content.
  - **Don't**
    - Include sensitive or confidential information in the prompt.

        - Use generative AI to draft communication materials on sensitive topics that require a human touch.
- **Code Development**
    - **Do**
        - Understand what the code is doing before deploying it in a production environment, understand the use of libraries and dependencies, and verify the code is complaint with state regulation with vulnerabilities and other security considerations.
    - **Don't**
        - Include sensitive or confidential information (including passwords, keys, proprietary information, etc.) in the prompt and code. (See ORS 192.345 and 355)
- **Image/Video Generation**
    - **Do**
        - Review generated content for biases and inaccuracies and engage with your communication department before using AI-generated audiovisual content for public consumption.
    - **Don't**
        - Include sensitive or confidential information in the prompt.
- **Chatbot Development**
    - **Do**
        - During development, conduct thorough user testing; after deployment, ensure users are informed they are interacting with an AI chatbot and provide a clear path to human assistance if needed.
    - **Don't**
        - Use generative AI as a substitute for human interaction or assume it will perfectly understand residents' queries. Provide mechanisms for residents to easily escalate their concerns or seek human assistance if the AI system cannot address their needs effectively.

Please contact the CSS Info email, eso.info@das.oregon.gov, if you have further questions.

Resources.

- Oregon Governor Kotek's executive order for a State Government Artificial Intelligence Advisory Council
    - Governor's Executive Order 23-26
- Council home page and meeting information
    - State Government Artificial Intelligence Advisory Council
- Implementing GenAI with Speed and Safety
    - Implementing generative AI with speed and safety | McKinsey

- Turing.com (Best AI Tools of 2024)
  - 11 Best Generative AI Tools and Platforms in 2024 (turing.com)

Federal Government.

- President Biden's executive order on Artificial Intelligence
  - Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence
- Blueprint for an AI Bill of Rights
  - Blueprint for an AI Bill of Rights
- OMB draft policy on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence
  - Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence
- Federal AI hub
  - AI.gov
  - NIST.AI.600-1.GenAI-Profile.ipd.pdf
  - Artificial Intelligence Risk Management Framework (AI RMF 1.0) (nist.gov)
  - Towards a Standard for Identifying and Managing Bias in Artificial Intelligence (nist.gov)
- U.S. General Services Administration IT Modernization centers of Excellence
  - AI Guidelines for Government