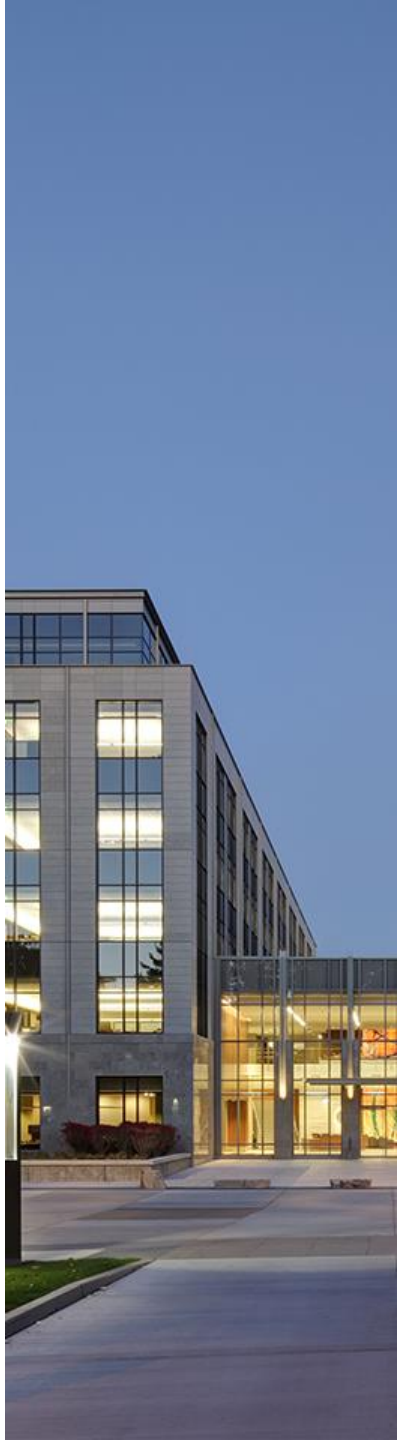
WaTech
Washington Technology Solutions

Introduction to Navigating AI Risks

May 2025



WaTech
Washington Technology Solutions

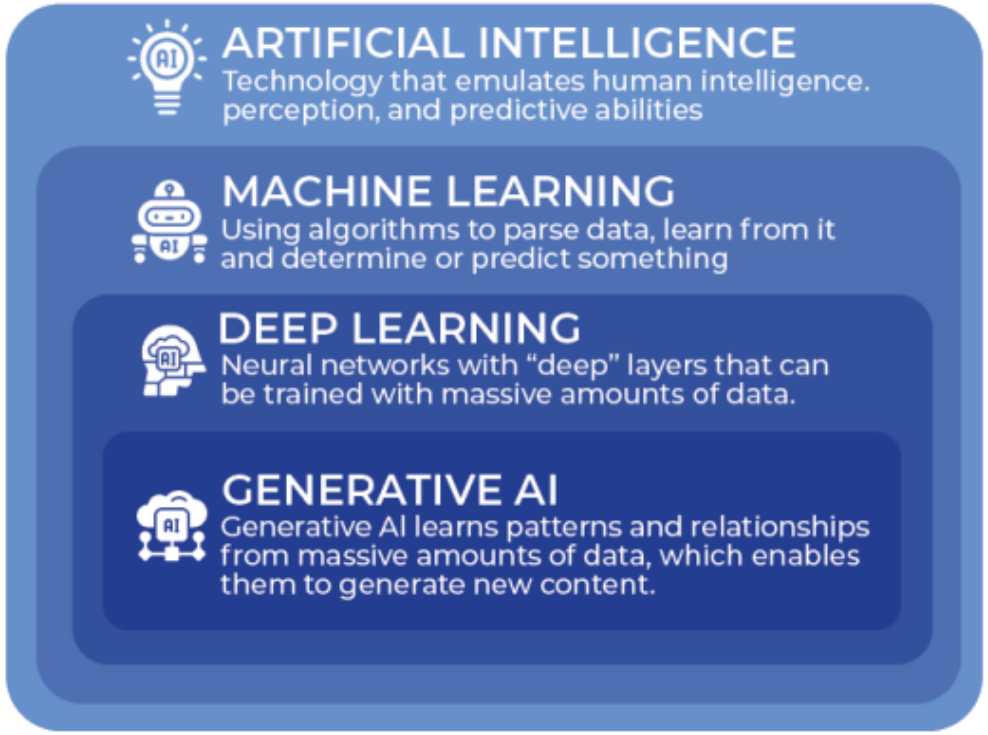
- Intro to Public Sector AI Use
- Washington's Responsible AI Principles
- Risk assessment guidance
- Identifying high-risk use cases
- Risk determination requirements

WaTech
Washington Technology Solutions

Intro to Public Sector AI Use

WaTech
Washington Technology Solutions

What is AI?



- **Large Language Model:** A specialized type of AI that has been trained on vast amounts of text to understand existing content and generate new, original content.
- **Neural Networks:** A model that, taking inspiration from the brain, is composed of layers consisting of simple connected units or neurons followed by nonlinearities.

WaTech
Washington Technology Solutions

Common Ways AI Shows Up

Generative AI

Creates content when prompted by the user. Learns from data to generate more targeted content over time (e.g. Co-Pilot, chatGPT)

Agentic AI

AI agents that are designed to perform tasks in the service of human goals, without direct human intervention (e.g. chat bots, virtual assistants)

Computer Vision

Processes, interprets, and provides insights into visual information (e.g. eel grass scanning, wildfire tracking)

Natural Language Processing

Understands, interprets, and generates human language in a meaningful way. (e.g. language translation)

Sources: WA State Generative AI Report ; Responsible AI in the Public Sector

WaTech
Washington Technology Solutions

AI in Our Technology

Intentional AI

- Solutions that are acquired and used specifically for their AI capabilities

Incidental AI

- New or existing solutions that have generative AI embedded, but it's not their primary purpose

3rd Party AI

- Solutions that are used by entities that interact with, but are not a part of WA state government

WaTech
Washington Technology Solutions

Why use AI?

- **Enhanced Customer Service.** AI helps facilitate access to information in near-real time, regardless of language preference.
- **Increase Industry Compliance.** AI assists regulatory agencies with summarizing state laws and regulations and creating guidance that's easy to understand and comply with.
- **Enables the Workforce through Automation.** AI can augment common tasks like transcribing meetings, summarizing lengthy documents, and searching through large data sets, thereby freeing up staff for more complex work.
- **Modernize the State's Technology.** IT staff can use AI to translate legacy coding languages into modern versions, quickly review and test code, and generate reference guides.

Common Public Sector Adoption Challenges

Content Accuracy

- Align responsible AI training with the use of AI technology
- Conduct thorough testing of outputs created by a newly adopted AI tool
- Conduct QA on all generatively created content

Staff Adoption

- Develop a strategy and training plan for your workforce
- Invest in training and upskilling of staff that are impacted by AI
- Ensure adoption of AI is in alignment with statewide principles and policies

Customer Adoption

- Conduct organizational change management and public engagement activities for internal staff and external customers
- Ensure all public-facing AI technology is clearly aligned with the WA State Agency Privacy Principles

Data Maturity

- Align AI-related initiatives to modernization and/or data management activities
- Leverage existing technology evaluation and adoption processes for all AI solutions

Sources: WA State Generative AI Report ; Responsible AI in the Public Sector

WaTech
Washington Technology Solutions

Responsible AI Principles

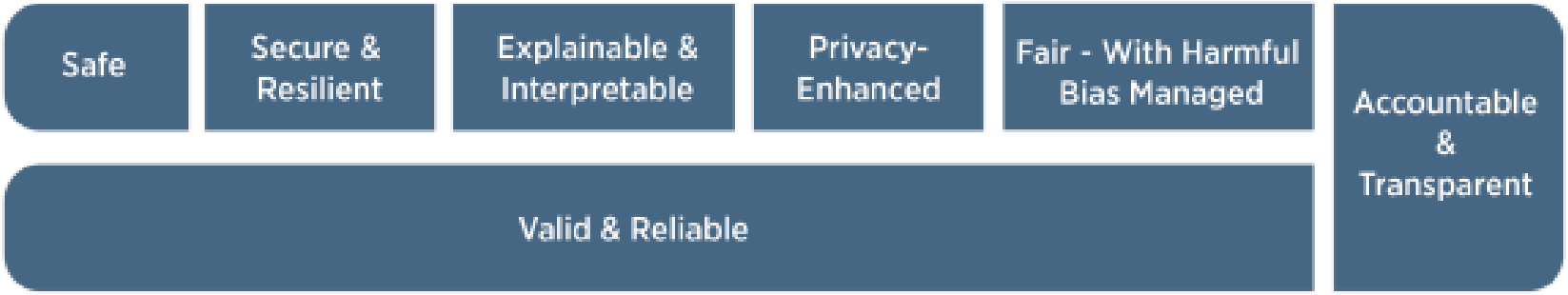
WaTech
Washington Technology Solutions

Washington AI Principles

- Safe, secure, and resilient
- Valid and reliable
- Fairness, inclusion, and non-discrimination
- Privacy and data protection
- Accountability and responsibility
- Transparency and auditability
- Explainable and interpretable
- Public purpose and social benefit

Washington AI Task Force

- Proposed adoption of NIST AI RMF Artificial Intelligence Trustworthiness Principles



WaTech
Washington Technology Solutions

Risk assessment guidance

WaTech
Washington Technology Solutions

By December 2024, WaTech will produce guidance on the risk assessments for the deployment of High-Risk Generative AI Systems. Assessments must leverage existing security and privacy assessment processes, and must include the following…

- Executive Order 24-01, sec. 8

WaTech
Washington Technology Solutions

a. Information about the High-Risk Generative AI System, including whether the High-Risk Generative AI System is provided by a third party, the name and address of the third party, and relevant state agency;
b. The intended uses of the High-Risk Generative AI System;
c. Assessment of the fitness of the High-Risk AI System for the intended purpose;
d. Assessment of impacted communities, benefits, harms, and mitigations of the High-Risk AI System;
e. An evaluation of the potential harms of the High-Risk Generative AI System which may include harms to individuals and groups, discriminatory or unfair outcomes, deceptive practices, societal risks, privacy and cybersecurity considerations, and national security concerns;
f. An assessment of mitigations including but not limited to consideration of restricted uses and limitations on use, policies, deidentified data, and commercial terms; and
g. Information about the agency approach to generative AI governance that is consistent with the AI Risk Management framework published by the National Institute of Science and Technology.

WaTech
Washington Technology Solutions

Artificial Intelligence Resources

WaTech
Washington Technology Solutions

Identifying high-risk use cases

Identifying AI Uses

	Intentional AI	Incidental AI
Original functionality	Existing review processes are well-equipped to identify these use cases.	Existing review processes can be used to identify these use cases.
		Processes and templates can be modified to capture appropriate information.

Identifying AI Uses

	Intentional AI	Incidental AI
Changes to capability or use	Existing review processes can be used to identify these use cases.	Existing review processes can be used to identify these use cases.
	Additional training and periodic review will help ensure these uses are identified.	Additional training and periodic review will help ensure these uses are identified.
		Contract terms can help identify these uses.

WaTech
Washington Technology Solutions

- Compared to existing technology, AI can:
  - **Pose similar risks,**
  - **Magnify existing risks, or**
  - **Introduce new risks**
- Risks may or may not decrease as technology advances
- More creativity = less predictability

WaTech
Washington Technology Solutions

NIST GenAI Profile

- **CBRN Information**: Eased access to or synthesis of materially nefarious information or design capabilities related to chemical, biological, radiological, or nuclear (CBRN) weapons, or other dangerous materials or agents.
- **Confabulation**: The production of confidently stated but erroneous or false content (known colloquially as "hallucinations" or "fabrications") by which users may be misled or deceived.
- **Dangerous or violent recommendations**: Eased production of and access to violent, inciting, radicalizing, or threatening content as well as recommendations to carry out self-harm or conduct illegal activities. Includes difficulty controlling public exposure to hateful and disparaging or stereotyping content.

WaTech
Washington Technology Solutions

NIST GenAI Profile

- **Data privacy**: Impacts due to leakage and unauthorized use, disclosure, or de-anonymization of biometric, health, location, or other personally identifiable information or sensitive data.
- **Environmental**: Impacts due to high compute resource utilization in training or operating GAI models, and related outcomes that may adversely impact ecosystems.
- **Harmful bias or homogenization**: Amplification and exacerbation of historical, societal, and systemic biases; performance disparities between sub-groups or languages, possibly due to non-representative training data, that result in discrimination, amplification of biases, or incorrect presumptions about performance; undesired homogeneity that skews system or model outputs, which may be erroneous, lead to ill-founded decision-making, or amplify harmful biases.

WaTech
Washington Technology Solutions

NIST GenAI Profile

- **Human-AI configuration**: Arrangement or interactions between a human and an AI system which can result in the human inappropriately anthropomorphizing GAI systems or experiencing algorithmic aversion, automation bias, over-reliance, or emotional entanglement with GAI systems.
- **Information integrity**: Lowered barrier to entry to generate and support the exchange and consumption of content which may not distinguish fact from opinion or fiction or acknowledge uncertainties, or could be leveraged for large-scale dis- and mis-information campaigns.
- **Information security**: Lowered barriers for offensive cyber capabilities, including via automated discovery and exploitation of vulnerabilities to ease hacking, malware, phishing, offensive cyber operations, or other cyberattacks; increased attack surface for targeted cyber attacks, which may compromise the confidentiality or integrity of training data, code, or model weights.

WaTech
Washington Technology Solutions

NIST GenAI Profile

- **Intellectual property**: Eased production or replication of alleged copyrighted, trademarked, or licensed content without authorization (possibly in situations which do not fall under fair use); eased exposure of trade secrets; or plagiarism or illegal replication.
- **Obscene, degrading, and/or abusive content**: Eased production of and access to obscene, degrading, and/or abusive imagery, including synthetic child sexual abuse material (CSAM), and nonconsensual intimate images (NCII) of adults.
- **Value chain and component integration**: Non-transparent or untraceable integration of upstream third-party components, including data that has been improperly obtained or not cleaned due to increased automation from GAI; improper supplier vetting across the AI lifecycle; or other issues that diminish transparency or accountability for downstream users.

WaTech
Washington Technology Solutions

Other challenges

- **Transparency**

- **Interpretability and explainability**
  - Interpretability – Transparency of the model itself
  - Explainability – After the fact explanations for outputs

- **Data readiness**:
  - Data availability
  - Inconsistent data
  - Outdated data
  - Mislabeled data

WaTech
Washington Technology Solutions

Risk determination requirements

High-Risk Generative AI System

- "High-Risk Generative AI System" means systems using generative AI technology that creates a high risk to natural persons' health and safety or fundamental rights. Examples include biometric identification, critical infrastructure, employment, health care, law enforcement, and administration of democratic processes.

- Executive Order 24-01

- Whether a system creates a high risk is dependent on:
  (1) the magnitude of an impact to natural persons' health and safety or fundamental rights, and
  (2) the likelihood of that impact occurring.

- Risks include both:
  - Direct impacts, such as when an AI system is used to determine eligibility for benefits, and
  - Indirect impacts, such as when an AI system is used to allocate resources in a way that impacts the people in a particular community.

Magnitude	Likelihood
1 – Negligible. No foreseeable direct or indirect impact to natural persons.	**1 – Remote or improbable**. Very low chance of occurring.
2 – Low. Any impact to natural persons is very unlikely to impact health, safety or fundamental rights.	**2 – Unlikely**. Low chance of occurring.
3 – Moderate. Some impact to natural persons that may include indirect impact to health, safety or fundamental rights.	**3 – Possible**. Moderate chance of occurring.
4 – Significant. Major effect causing substantial harm or disruption to health, safety or fundamental rights. May include direct impact in individual circumstances, or indirect, systemic impacts.	**4 – Likely**. High chance of occurring.
5 – Severe or catastrophic. Extreme impact resulting in serious harm, injury or violation of fundamental rights.	**5 – Probable**. Very high chance of occurring.

WaTech
Washington Technology Solutions

The intended use and operating context

- What is the intended use of system outputs and who will benefit?
- Who might be adversely impacted and how? Consider at least the workforce, customers/residents, and vulnerable or underserved populations.
- What is the role of humans in reviewing system outputs? For example, do they have little to no review, do humans review in defined circumstances, or do humans verify, modify, or review all outputs?
- Who are the intended users? Is it intended for internal or external use? Is it intended for broad use or a small group of people?
- What laws apply to agency use of the system?
- What alternatives were considered and why is this the preferred option?

WaTech
Washington Technology Solutions

Data characteristics

- What is the highest data classification level for included information? Consider both information used for training and information processed during operation, including intended prompts, inputs, fine-tuning, augmentation or any other system customization.
- What type of information is involved? Will the system process personal information?

WaTech
Washington Technology Solutions

System characteristics and safeguards

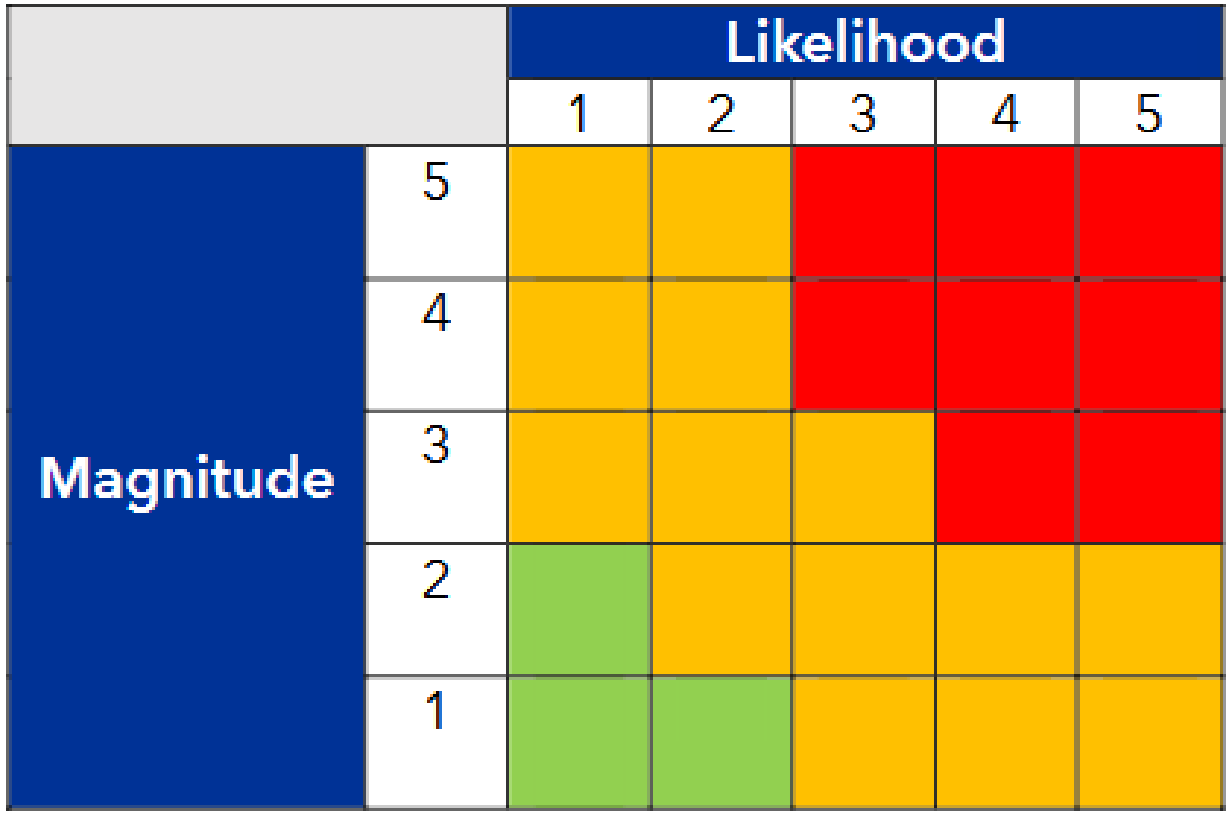
- Is there evidence of biased, discriminatory, inaccurate or otherwise unreliable outputs? What measures are in place to address unreliable outputs?
- Are there built-in filters or monitoring to detect harmful outputs, manipulation, or misuse?
- How explainable is the system? Do users know the system uses AI? Does the system provide sources or otherwise help identify to users why it is producing a particular output?
- How transparent is system implementation? Are users trained or notified of appropriate use and system limitations

[Magnitude] x [Likelihood] = Risk level

1-2 = Low

3-11 = Moderate

12-25 = High



11. Does your deployment of this solution use AI technologies? *

Definition of AI:
https://watech.wa.gov/policies/definition-terms-used-policies-and-reports#:~:text=the%20change%20request.-,Artificial%20Intelligence,MGMT%2D01%2D01%2DS%20Technology%20Portfolio%20Foundation%20%2D%20Applications,-Asset

- (●) Yes
- ( ) No

12. What AI risk level has your agency assigned to this solution? *

See Artificial Intelligence Risk Level Guidance: https://watech.wa.gov/sites/default/files/2025-02/AI%20risk%20guidance%20final.pdf

- ( ) Low
- (●) Moderate
- ( ) High
- ( ) Undetermined

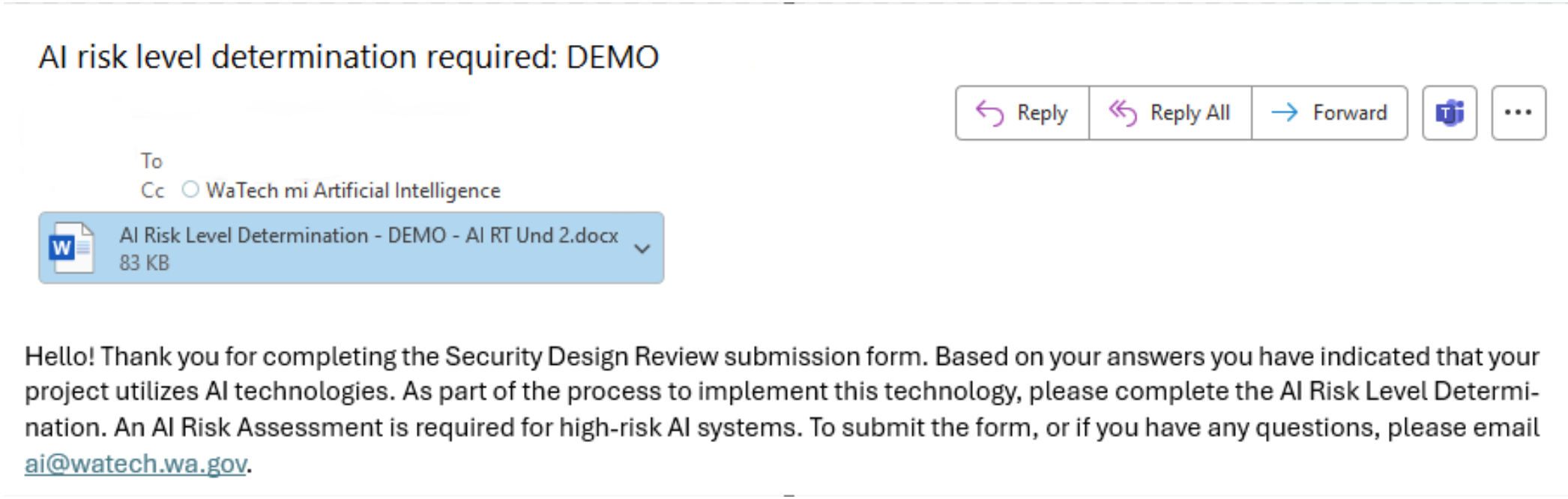
13. Please enter the email address for the individual who made the AI risk level determination: *

Enter your answer

OR

13. You will be required to submit the AI risk level determination (Low/Moderate/High) for this solution to the ai@watech.wa.gov mailbox before this SDR can be completed. *

- ◯ I understand


AI risk level determination required: DEMO
Reply
Reply All
Forward
To
Cc
WaTech mi Artificial Intelligence
AI Risk Level Determination - DEMO - AI RT Und 2.docx
83 KB
Hello! Thank you for completing the Security Design Review submission form. Based on your answers you have indicated that your project utilizes AI technologies. As part of the process to implement this technology, please complete the AI Risk Level Determination. An AI Risk Assessment is required for high-risk AI systems. To submit the form, or if you have any questions, please email ai@watech.wa.gov.

Artificial Intelligence Risk Level Determination

How to use this document.

This guidance describes when an AI system is "high-risk," and includes considerations to help make that determination. The information in this document specifically addresses the minimum requirements to identify AI systems that pose a high risk to health, safety or fundamental rights. Agencies may prioritize other types of risks posed by AI systems. For example, an AI system could pose a high reputational or operational risk even if it does not impact health, safety or fundamental rights. Similarly, agencies may have different risk tolerances based on their operating context. Agencies are encouraged to consider additional types of risk.

Using the considerations below, identify the AI risk level for this project and send the completed document to ai@watech.wa.gov.

Agency	
Project name	
AI risk level	☐ Low ☐ Moderate ☐ High
Email of person who made risk determination	

WaTech
Washington Technology Solutions

Thank you!

privacy@watech.wa.gov