WaTech
Washington Technology Solutions

Navigating AI in the Public Sector: Challenges and Best Practices

October 2024

WaTech
Washington Technology Solutions

Today's Goals

- Artificial intelligence, what is that?
- What are the different types of AI? What is generative AI?
- What are the basic components/ how does it work?
- What can I do to enable responsible innovation?

WaTech
Washington Technology Solutions

Today's Goals

Understand how to determine when the benefits of identified use cases outweigh risks, innovate effectively, and avoid unnecessary risks.

WaTech
Washington Technology Solutions

What is AI?

- "AI system" means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments"

  - EU AI Act, Article 3

WaTech
Washington Technology Solutions

What is AI?

- Components of common definitions:
  - Technology
  - Some type of human involvement
  - Some type of autonomy
  - Some type of output
  - Bonus – performs a task otherwise expected to require a human

WaTech
Washington Technology Solutions

What is AI?

- Robotic vision
- Natural language processing
- Expert systems
- Edge AI
- Biometrics
- Generative AI (GenAI)
- Deep learning
- Machine learning

*not exhaustive, not mutually exclusive

WaTech
Washington Technology Solutions

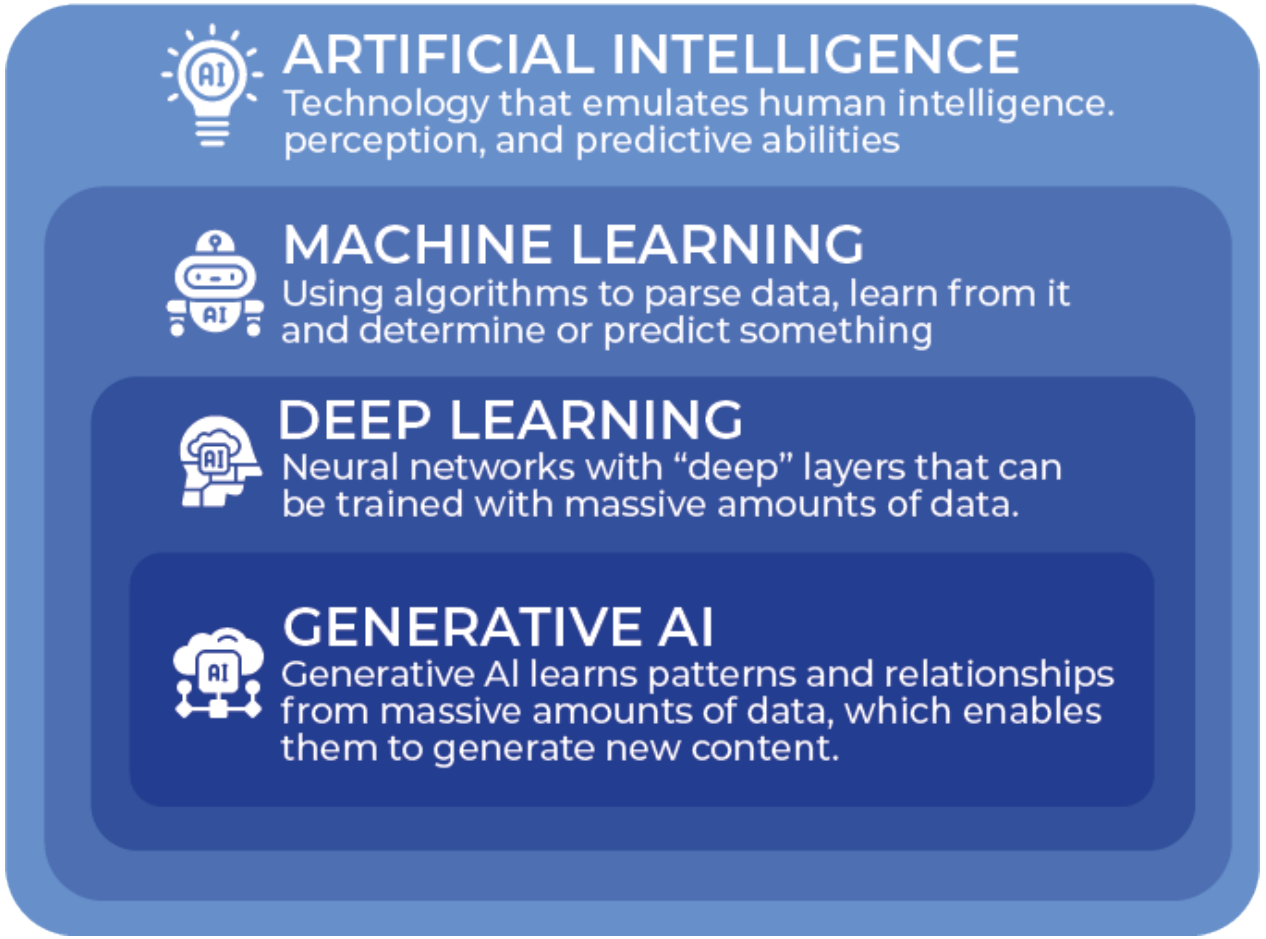
What is GenAI?

- ". . . a technology that can create content, including text, images, audio, or video, when prompted by a user. GenAI systems learn patterns and relationships from large amounts of data, which enables systems to generate new content that may be similar, but not identical, to the underlying training data."

  -Executive Order 24-01, 2024

WaTech
Washington Technology Solutions

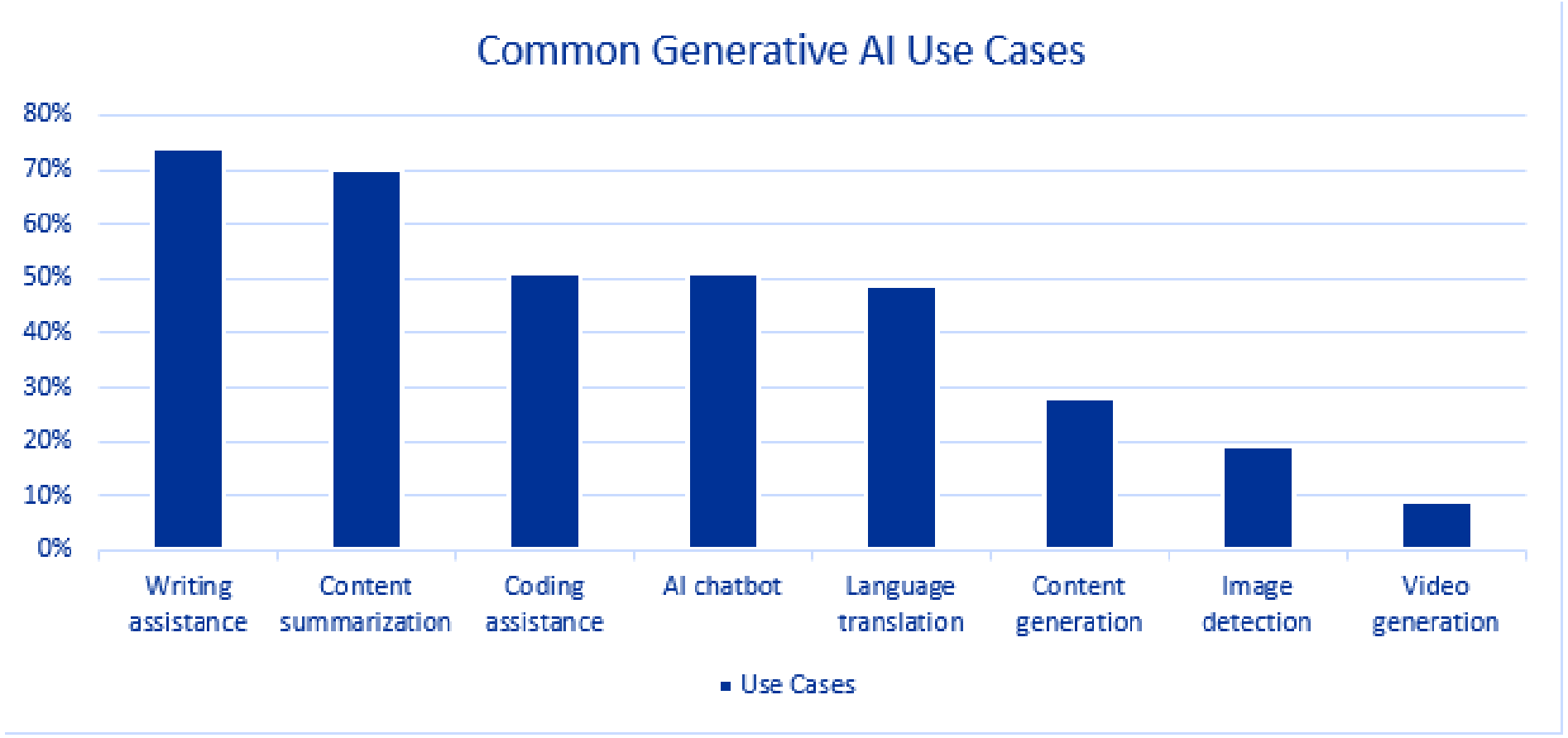
What is GenAI?



WaTech
Washington Technology Solutions

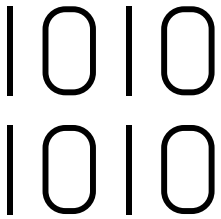
Common GenAI Use Cases	
Writing assistance	Chatbots
Content summarization	Language translation
Data analysis	Audio generation
Coding assistance	Image or video generation

Based on May 2024 Agency Survey Results



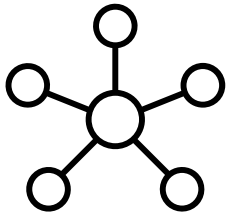
WaTech
Washington Technology Solutions

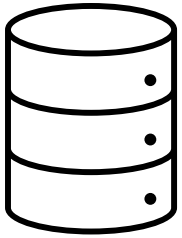
Infrastructure



Compute

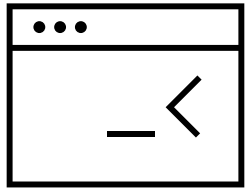
Network





Storage

Software



WaTech
Washington Technology Solutions

Data, data, data

- Training the model
  - Supervised, unsupervised, reinforcement
- Using the model
  - Prompts
  - Context window
  - Inputs
  - Retrieval augmented generation

WaTech
Washington Technology Solutions

What can I do to enable responsible innovation?

WaTech
Washington Technology Solutions

Try using it

WaTech
Washington Technology Solutions

Understand risks

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: Lowered barriers to entry or eased access to materially nefarious information related to chemical, biological, radiological, or nuclear (CBRN) weapons, or other dangerous biological materials.
- Confabulation: The production of confidently stated but erroneous or false content (known colloquially as "hallucinations" or "fabrications").
- 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 criminal or otherwise illegal activities.

WaTech
Washington Technology Solutions

NIST GenAI Profile

- **Data privacy**: Leakage and unauthorized disclosure or de-anonymization of biometric, health, location, personally identifiable, or other sensitive data.
- **Environmental**: Impacts due to high resource utilization in training GAI models, and related outcomes that may result in damage to ecosystems.
- **Human-AI configuration**: Arrangement or interaction of humans and AI systems which can result in algorithmic aversion, automation bias or over-reliance, misalignment or mis-specification of goals and/or desired outcomes, deceptive or obfuscating behaviors by AI systems based on programming or anticipated human validation, anthropomorphization, or emotional entanglement between humans and GAI systems; or abuse, misuse, and unsafe repurposing by humans.

WaTech
Washington Technology Solutions

NIST GenAI Profile

- **Information integrity**: Lowered barrier to entry to generate and support the exchange and consumption of content which may not be vetted, may not distinguish fact from opinion or acknowledge uncertainties, or could be leveraged for large-scale dis- and mis-information campaigns.
- **Information security**: Lowered barriers for offensive cyber capabilities, including ease of security attacks, hacking, malware, phishing, and offensive cyber operations through accelerated automated discovery and exploitation of vulnerabilities; increased available attack surface for targeted cyber attacks, which may compromise the confidentiality and integrity of model weights, code, training data, and outputs.
- **Intellectual property**: Eased production of alleged copyrighted, trademarked, or licensed content used without authorization and/or in an infringing manner; eased exposure to trade secrets; or plagiarism or replication with related economic or ethical impacts.

WaTech
Washington Technology Solutions

NIST GenAI Profile

- **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.
- **Toxicity, bias, and homogenization**: Difficulty controlling public exposure to toxic or hate speech, disparaging or stereotyping content; reduced performance for certain sub-groups or languages other than English due to non-representative inputs; undesired homogeneity in data inputs and outputs resulting in degraded quality of outputs.
- **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:
  - Inconsistent data
  - Outdated data
  - Mislabeled data

WaTech
Washington Technology Solutions

Identify risk levels

EU AI Act Risk Levels

Unacceptable risk – Prohibited uses.

High-risk – Requirements include:

- Risk management system
- Data governance
- Technical documentation
- Record-keeping
- Instructions for use
- Ability to implement human oversight
- Appropriate accuracy, robustness and cybersecurity

Limited risk – Requires transparency.

General Purpose AI (GPAI) – Separate requirements for GPAI depending on whether it presents systemic risk.

WaTech
Washington Technology Solutions

EU AI Act – Unacceptable Risk

- Subliminal, manipulative, or deceptive techniques
- Exploiting vulnerabilities
- Biometric categorization
- Social scoring
- Crime risk prediction
- Facial recognition databases
- Emotion detecting on school or workplace
- Real-time remote biometric identification

WaTech
Washington Technology Solutions

EU AI Act – High-risk

- Non-banned biometric categorization
- Non-banned emotion detection
- Non-banned remote biometric identification
- Education and vocational training
- Employment, workers management and access to self-employment
- Critical infrastructure
- Access to and enjoyment of essential public and private services
- Law enforcement
- Migration, asylum and border control management
- Administration of justice and democratic processes

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.

- Consider at least:
  - The intended use and operating context
  - Data characteristics
  - System characteristics and safeguards

WaTech
Washington Technology Solutions

Less risk	More risk
Category 1	Category 2, 3 or 4 information
Recommendations or suggestions	Automated decisionmaking
No impacts to safety or fundamental rights	Potential impacts to safety or fundamental rights
Internal facing	External facing
Agency governance and workforce education	Implemented without oversight or workforce education
Appropriate agreements	Free version
Controlled data inputs, high data quality	Poor or unknown data quality

WaTech
Washington Technology Solutions

Adopt principles

WaTech
Washington Technology Solutions

State principles for responsible GenAI use

- **Safe, secure, and resilient**: AI should be used with safety and security in mind, minimizing potential harm and ensuring that systems are reliable, resilient, and controllable by humans.
- **Valid and reliable**: Agencies should ensure AI use produces accurate and valid outputs and demonstrates the reliability of system performance.
- **Fairness, inclusion, and non-discrimination**: AI applications must be developed and utilized to support and uplift communities, particularly those historically marginalized. Fairness in AI includes concerns for equality and equity by addressing issues such as harmful bias and discrimination.
- **Privacy and data protection**: AI use should 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.

WaTech
Washington Technology Solutions

State principles for responsible GenAI use

- **Accountability and responsibility**: As public stewards, agencies should use 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 that "how" a 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 better and more equitable services and outcomes to its residents.

See Interim Guidelines for Purposeful and Responsible Use of Generative Artificial Intelligence.

WaTech
Washington Technology Solutions

Adopt policies or guidelines

WaTech
Washington Technology Solutions

- Establish minimum requirements now, iterate and improve later

- Consider at least:

  - Requirements for confidential information
  - Role of human review and accountability
  - Transparency of consequential uses

WaTech
Washington Technology Solutions

Train staff

WaTech
Washington Technology Solutions

Purposeful adoption

WaTech
Washington Technology Solutions

- What is the business problem?
- Why is AI the best solution?
- What does success look like?
- How will success be monitored?

WaTech
Washington Technology Solutions

Get the right people in the room

WaTech
Washington Technology Solutions

Prioritize use cases

WaTech
Washington Technology Solutions

Criteria example

Potential impact

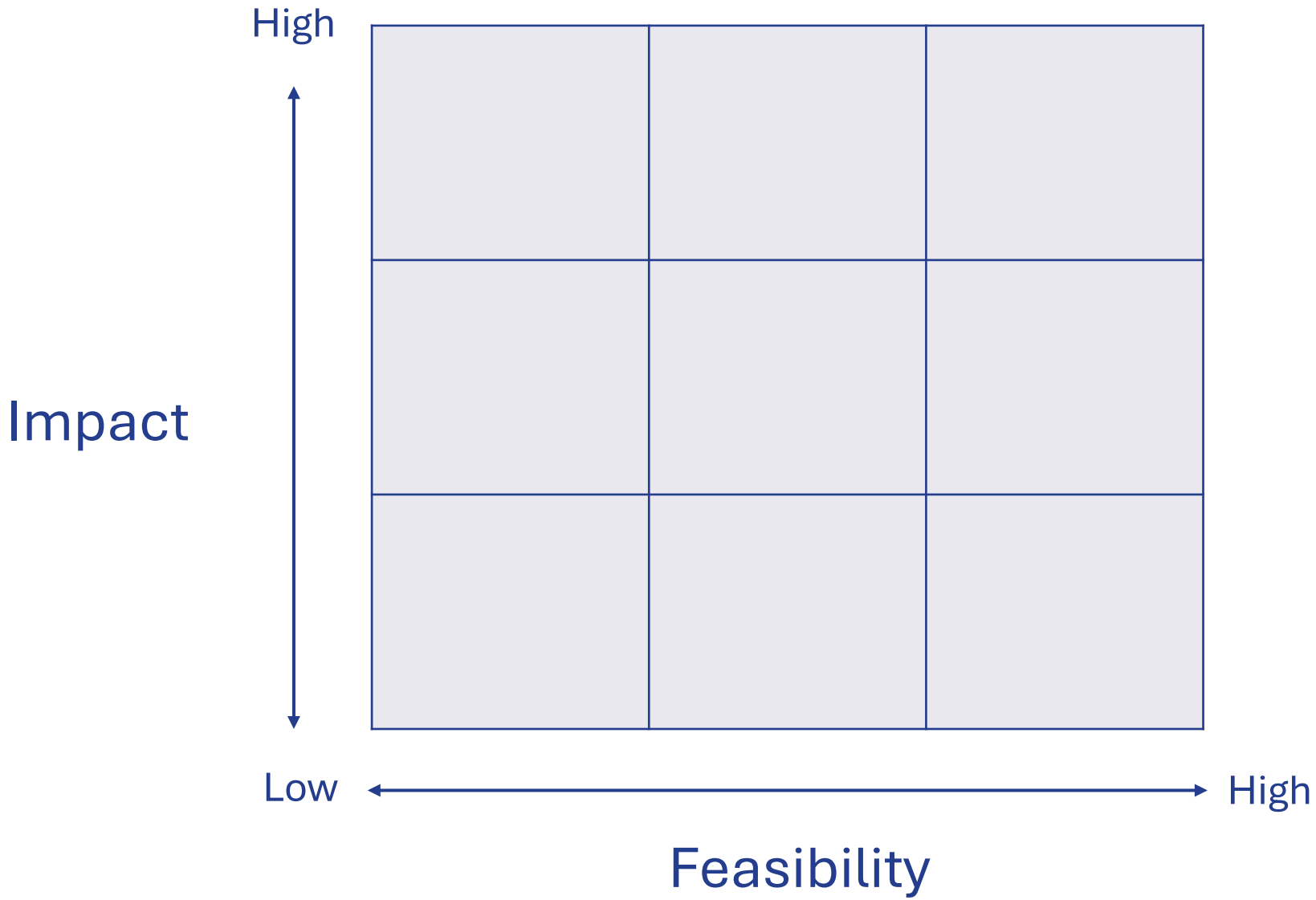
- Expected outcomes
- Strategic alignment
- Scalability

Feasibility

- Risk
- Technical requirements
- Data readiness

WaTech
Washington Technology Solutions

Example matrix



WaTech
Washington Technology Solutions

Use existing resources, get involved, stay up-to-date

WaTech
Washington Technology Solutions

AI Initiatives in State of Washington

AI CoP

- Steering Committee
- State Agencies
- Local Gov't
- Subcommittees
  - Risk
  - Policy
  - Use cases
  - Local Gov't

EO 24-01

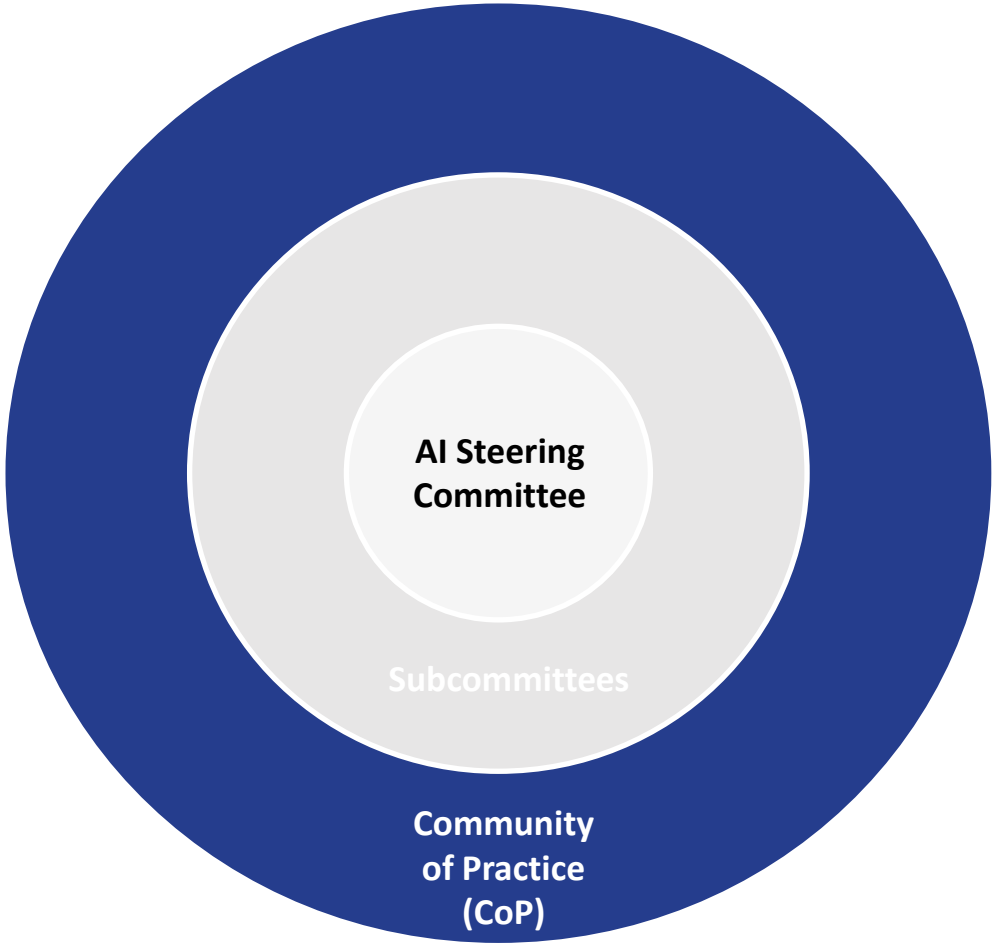
- Deliverables
  - WaTech
  - DES
  - OFM
  - OOE
  - WTB

AGO AI Taskforce

- Executive Committee
  - (19 members)
  - Industry, advocacy, government, legislative members
- Subcommittees

AI CoP

- **Governance Structure**
  - **Representation from WaTech, State Agency, and Local Government**
- **Steering Committee Objectives**
  - **Develop a set of guidelines and policies**
  - **Identify and document best practices**
  - **Establish a governance structure and develop mechanisms for accountability and oversight**
  - **Document use cases and examine potential societal impact**
  - **Facilitate collaboration and knowledge sharing**
  - **Promote alignment of new AI technologies to business and IT strategies**



WaTech
Washington Technology Solutions

EO SEC.	LEAD AGENCY	NAMED COLLABORATORS	DELIVERABLE	DEADLINE
2	WaTech	Cabinet Agencies	Report of Gen AI initiatives for agencies	September 2024
3	WaTech	DES	Initial Guidelines for Procurement	September 2024
4	DES	Office of Equity WaTech	Training plan for state workers	January 2025
5	WaTech	Office of Equity Community members Tribal governments SMEs State agencies	Guidelines on impact of adopting Gen AI on vulnerable communities	December 2024
6	DES	WaTech	Contract terms templates	January 2025
7	Office of Equity	WaTech DES WTB	Accountability Framework	September 2024
8	WaTech		Risk assessments	December 2024
9	OFM	Labor organizations WTB WaTech	Report of impact of Gen AI on state workforce	December 2024
10	WTB		Identify and create research opportunities	January 2025

WaTech
Washington Technology Solutions

Thank you!

privacy@watech.wa.gov