GUIDELINE

Generative Artificial Intelligence Policy

POL-209

Purpose

The purpose of this policy is to set forth requirements City departments will observe when acquiring and using software that meets the definition of "generative artificial intelligence."

Scope

All City departments. Vendors, contractors, and volunteers who operate on behalf of the City are also subject to this policy.

Definitions

Generative Artificial Intelligence (Generative AI) is a class of computer software and systems, or functionality within systems, that use large language models, algorithms, deep-learning, and machine learning models, and are capable of generating new content, including but not limited to text, images, video, and audio, based on patterns and structures of input data. These also include systems capable of ingesting input and translating that input into another form, such as text-to-code systems.

While this policy document includes principles that apply to AI technologies generally, the policy statements apply only to generative AI systems.

Artificial Intelligence (AI) Principles

Principles describe general codes of conduct that represent the City's values and are aligned with our responsibilities to the residents we serve. These principles serve to guide City employees in their use of both generative and traditional AI technology. City employees shall adhere to the principles and requirements outlined in this policy, and will be held accountable for compliance with these commitments.

1. **Innovation and Sustainability:** The City values public service innovation to meet our residents' needs. We commit to responsibly explore and evaluate AI technologies, which will improve our services and advance beneficial outcomes for both people and the environment.
2. **Transparency and Accountability:** The City values transparency and accountability and understands the importance of these values in our use of AI systems. The City will ensure that the development, use, and deployment of AI systems are evaluated for and compliant with all laws and regulations applicable to the City prior to use, and will make documentation related to the use of AI systems available publicly.
3. **Validity and Reliability:** The City will work to ensure that AI systems perform reliably and consistently under the conditions of expected use, and that ongoing evaluation of system accuracy throughout the development and/or deployment lifecycle is managed, governed, and auditable, to the greatest extent possible.
4. **Bias and Harm Reduction and Fairness:** We acknowledge that AI systems have the potential to perpetuate inequity and bias resulting in unintended harms on Seattle residents. The City will evaluate AI systems through an equity lens, in alignment with our Race and Social Justice

commitments, for potential impacts such as discrimination and unintended harms arising from data, human, or algorithmic bias to the extent possible.

5. **Privacy Enhancing:** The City values data privacy and understands the importance of protecting personal data. We work to ensure that policies and standard operating procedures that reduce privacy risk are in place, and are applied to the AI system throughout development, testing, deployment, and use to the greatest extent possible.
6. **Explainability and Interpretability:** The City understands the importance of leveraging AI systems, models, and outputs that are easily interpreted and explained. We work to ensure all AI systems and their models are explainable to the extent possible, and that system outputs are interpretable and communicated in clear language, representative of the context for use and deployment.
7. **Security and Resiliency:** Securing our data, systems, and infrastructure is important to the City. We will ensure AI systems are evaluated for resilience and can maintain confidentiality, integrity, and availability of data and critical City systems, through protection mechanisms to minimize security risks to the greatest extent possible, in alignment with governing policy and identified best practices.

Policy

1. Acquisition of Generative AI Technology

1.1. Consistent with the City's standards for Acquisition of Technology Resources, City employees may be authorized to use pre-approved generative AI software tools or they may request a non-standard acquisition of generative AI software through Seattle IT's current request process.

1.2. Seattle IT shall review exception requests according to its current risk and impact methodology, which shall include specific review criteria for generative AI technology. Seattle IT shall either approve or deny a request according to its criteria.

1.3. The City's standard for technology acquisition applies to all technology, including free-to-use software or software-as-a-service tools.

1.4. If a technology that has already been approved for use in the City adds or incorporates generative AI capabilities, no additional approval is required to use those capabilities, however all other aspects in this policy apply to said use.

1.5. Seattle IT may revoke authorization for a technology that adds AI capabilities, or may restrict the use of those AI capabilities, if, in its judgment, those AI capabilities present risks that cannot be effectively mitigated to comply with this policy or other City policies.

2. Use of Generative AI Outputs

2.1. Outputs of Generative AI systems must be reviewed by humans prior to each use in an official City capacity ("Human in the Loop" or HITL). HITL review processes shall be documented by owning departments and shall demonstrate how the HITL review was conducted to adhere to the principles outlined in this document.

2.2. Documentation of HITL reviews shall be retained according to the appropriate records retention schedule.

3. Attribution, Accountability, and Transparency of Authorship

3.1. All **images and videos** created by Generative AI systems must be attributed to the appropriate Generative AI system. Wherever possible, attributions and citations to the City of Seattle should be embedded in the image or video (e.g., via digital watermark).

3.2. If **text** generated by an AI system is used substantively in a final product, attribution to the relevant AI system is required.

3.3. If a significant amount of **source code** generated by an AI system is used in a final software product, or if any amount is used for an important or critical function, attribution to the appropriate AI system is required via comments in the source code and in product documentation.

3.4. All attributions should include the name of the AI system used plus an HITL assertion (which should include the department or group who reviewed/edited the content).

Example: Some material in this brochure was generated using ChatGPT 4.0 and was reviewed for accuracy by a member of the Department of Human Services before publication.

3.5. Departments shall interpret "substantive use" thresholds to be consistent with the principles outlined in this document as well as relevant intellectual property laws.

4. Reducing Bias and Harm

4.1. Generative AI systems may produce outputs based on stereotypes or use data that is historically biased against protected classes. City employees must leverage RSJI resources (e.g., the Racial Equity Toolkit) and/or work with their departmental RSJI Change Team to conduct and apply a Racial Equity Toolkit (RET) prior to the use of a Generative AI tool, especially uses that will analyze datasets or be used to inform decisions or policy. As per the objectives of the RSJ program, the RET should document the steps the department will take to evaluate AI-generated content to ensure that its output is accurate and free of discrimination and bias against protected classes.

5. Data Privacy

5.1. Use of generative AI tools shall be consistent with the principles and standards described in the City's Data Privacy Policy and Information Security Policy.

5.2. Unless suitable enterprise controls and data protection mitigations are in place, employees shall not submit data that is classified by the City's data classification guidelines as Confidential or Confidential with Special Handling, or that otherwise not considered to be acceptable to disclose to the public, shall not be submitted to Generative AI systems.

5.3. No City data or records, including inputs or prompts, are to be used for training or parameter-tuning for Generative AI models outside the City's control. AI technologies that cannot prevent City data or records from contributing to their language models may not be used by City employees.

6. Public Records & City Records Management

6.1. All records generated, used, or stored by Generative AI vendors or solutions may be considered public records and must be disclosed upon request.

6.2. All Generative AI solutions and/or vendors approved for City use shall be required to support retrieval and export of all prompts and outputs (either via exposed functionality or through vendor contract assurances).

6.3. City employees who use generative AI tools are required to maintain, or be able to retrieve upon request, records of inputs, prompts, and outputs in a manner consistent with the City's records management and public disclosure policies and practices.

Exceptions

Exceptions must be approved in advance through submission of a Seattle IT Exception Review Approval request in Service Hub. This can be submitted directly or with the assistance of Client Engagement personnel. Note: this section refers to exceptions to *this policy* as it relates to generative AI tools that are in use by the City. It does not refer to requests for acquisition of non-standard applications or technologies.

Non-compliance

The Chief Technology Officer (CTO) is responsible for compliance with this policy. Enforcement may be imposed in coordination with individual division directors and department leaders. Non-compliance may result in department leaders imposing disciplinary action, restriction of access, or more severe penalties up to and including termination of employment or vendor contract.

Related Standards and Policies

- City Privacy Policy [POL-202]
- Acquisition of Technology Resources [STA-209]
- Information Security Policy [POL-201]
- Data Classification Guideline [GUI-110]

Responsibilities

The policy will be maintained through the Data Privacy, Accountability and Compliance (DPAC) division, owned by the Director of DPAC and City of Seattle Chief Privacy Officer. Their responsibilities include creating and maintaining the generative AI risk and impact criteria and the documents and forms to support the exception review process for this technology.

Document Control

This policy shall be effective on 11/1/2023 and shall be reviewed annually.

Version	Content	Contributors	Approval Date
v 1.0	Initial Draft	Reviewer: Greg Smith – Chief Information Security Officer (CISO)	**10/23/2023**
	Final	**Approver:** **Jim Loter – Interim Chief Technology Officer (CTO)**	**10/23/2023**