Commonwealth of Massachusetts
Executive Office of Technology Services and Security (EOTSS)
Enterprise Privacy Office

Guidelines for the Use and Development of Generative Artificial Intelligence

Document Name: GenAI Guidelines	Effective Date: 4/12/2024
Document ID: AI.001	Last Revised Date: 4/12/2024

Preamble

The Executive Office of Technology Services and Security (EOTSS) recognizes the potential that generative artificial intelligence offers in assisting Commonwealth Agencies and Offices to improve the delivery and efficacy of services for Massachusetts constituents. At the same time, use of this technology raises risks related to societal harms such as fraud, data integrity, discrimination, bias, and disinformation. Harnessing AI for its potential will require mitigating these risks. Similar to the evolution of the internet, the responsible use of generative AI has the potential to enhance our work by helping employees to learn new skills, improve their performance, and adapt to changing environments. Generative AI is a tool that has the potential to augment employee's capabilities and empower them to achieve more. The Commonwealth will continue to leverage the talent pipelines that exist within Massachusetts schools, colleges and universities, and the emerging technology sector to safely, equitably, and responsibly implement generative AI so that Massachusetts remains an economic leader and a competitive place to live and work.

These guidelines are intended to establish minimum requirements for the development and use of generative AI by Commonwealth Agencies and Offices. The goal of these guidelines is to foster public trust, support business outcomes, and ensure the ethical, transparent, and accountable development and implementation of generative AI technology by Commonwealth Agencies and Offices (as defined below). Recognizing the rapidly evolving nature of generative AI, EOTSS will periodically update and assess these guidelines to align with emerging technologies, challenges, use cases, and input from the Commonwealth's Artificial Intelligence Strategic Task Force.

Scope

These guidelines apply to Executive Department Agencies including all executive offices, boards, commissions, agencies, departments, divisions, councils, bureaus, and offices within an executive office. Other Commonwealth entities that voluntarily use or participate in services provided by the Executive Office of Technology Services and Security, such as mass.gov, must agree to comply with this document, with respect to those services, as a condition of use (hereinafter referred to as "Commonwealth Agencies and Offices"). These guidelines apply to all state employees, contractors, consultants, vendors, and interns, including full-time, part-time, or voluntary (hereinafter referred to as "Personnel"). Offices, agencies or entities other than Executive Department Agencies may adopt additional standards, guidance, or policies with respect to the use of AI in their own agency/office/entity, provided those standards, guidelines, or policies are at least as protective as those set forth herein.

Definitions

"Artificial Intelligence," as used in these guidelines, is a machine-based system that can, for a given set of human objectives, make predictions, recommendations, or decisions.

"AI Systems" means any data system, software, hardware, application, tool, or utility that operates in whole or in part using AI.

"Generative Artificial Intelligence" ("Generative AI") is a type of artificial intelligence technology that can generate many forms of content including but not limited to texts, images, and multimedia.

Principles

Similar to risks associated with other types of technology, Generative AI risks can emerge in a variety of ways. To reduce risk, minimize negative impacts, and promote trustworthiness in the use of Generative AI, EOTSS adopts the following technical and social principles. These principles will serve as the foundation for the responsible use and development of Generative AI. The following principles largely track the National Institute of Standards and Technology (NIST) AI Risk Management Framework. The following principles outline the characteristics of trustworthy AI systems to include validation and reliability; safety, security, and resiliency; accountability; transparency; explainability and interpretability; privacy by design; fairness with mitigation of harmful biases; knowing your content; and public purpose and social benefit, as further outlined below.

Validation and Reliability: Generative AI systems can produce inaccurate, unreliable, or poorly generalized data which increases risks and reduces trustworthiness. Validation processes are key to ensuring accuracy and reliability of their Generative AI systems.

Safety, Security, and Resiliency: Generative AI systems should be designed to operate safely, securely, and exhibit reliability, and resilience under various conditions. Generative AI must not endanger human life, health, property or the environment and must be controllable by humans.

Accountability: Clear accountability measures for Generative AI systems are vital to public trust and users should be held accountable for the performance, impact, and consequences of the use of Generative AI in their work.

System and Procedural Transparency: Documentation of decision-making processes around testing, datasets, and modeling is essential for building transparency in the use of Generative AI systems. Users should provide a record of Generative AI processes and information about AI's outputs to those interacting with or relying on the system.

Explainability and Interpretability: Explainability and interpretability assist those operating or overseeing a Generative AI system, as well as users of a Generative AI system, to gain deeper insights into the functionality and trustworthiness of the system, including its outputs. There should be documented and understandable explanations for the decisions made by and with Generative AI systems.

Privacy by Design: Establishing robust data privacy designs and measures at the early stage of development and use of Generative AI fosters trust between the public and the Commonwealth. Generative AI systems should be utilized in a way that respects the rights and obligations of individuals and organizations with respect to the collection, use, retention, disclosure, and destruction of personally identifiable information.

Fairness with Mitigation of Harmful Bias: Generative AI should be developed and used in such a way that it respects, serves, and protects humans' personal and cultural sense of identity, physical and mental integrity. Generative AI systems should be evaluated for biases and users should create mitigating strategies to ensure that individuals and groups are free from unfair bias, discrimination, and stigmatization.

Know your Content: Users need to "know the content" that Generative AI creates using a label or other mark informing individuals when any output is produced by a Generative AI model rather than a human being. This labeling obligation must protect the public from the alteration of original content and the creation of "deep fakes."

Public purpose and social benefit: Generative AI should be used in a way that improves the delivery and efficacy of equitable services for Commonwealth residents. Generative AI should be used to support Personnel in their work and must not be used to replace existing jobs.

Guidelines

Fact Check, Bias Reduction, Review

Commonwealth Agencies and Offices must implement processes for fact-checking and human verification of Generative AI-generated content and ensure that all Generative AI practices comply with applicable laws, regulations, policies, and guidelines including those concerning discrimination, privacy, and data protection. This should include a multi-level internal review and approval process. Generative AI systems have the potential to mirror biases from the source materials used for training, encompassing harmful and discriminatory biases that can perpetuate and amplify negative impacts on individuals, organizations, and society. These biases can be related to but not limited to culture, gender, ethnicity, finance, health, and other societal factors. Moreover, the algorithms responsible for parsing and processing contents may introduce such biases as well. Bias identification, management and mitigation must be considered at all stages of the Generative AI development lifecycle, designing, developing, deploying, evaluating, using, and auditing. Personnel must reasonably evaluate any content generated by Generative AI systems or applications to identify any inaccurate information and mitigate unintended or undesirable instances of biases. This includes addressing contents that could potentially be offensive or harmful.

Disclosure and Attribution

Commonwealth Agencies and Offices must be transparent in their use of Generative AI and properly identify and credit AI-generated content. Generative AI content used by Commonwealth Agencies and Offices must be clearly labeled in a conspicuous manner, including details of its review, and editing process. This allows for transparent authorship and responsible content evaluation.

Examples of disclosures lines that can be used:

- This memo was summarized by [generative AI Tool] using the following prompt: "Summarize the following memo: (memo content)".
- The summary was reviewed and edited by [insert name(s)].
- This code was written with the assistance of [generative AI tool].

Commonwealth Agencies and Offices must also perform reasonable due diligence, including consultation with agency legal counsels, to reasonably ensure that no copyrighted material is released without proper attribution or the acquisition of necessary rights. Generative AI systems might unintentionally violate existing copyrights across various media, including audio, video, books, publications, and other forms of

intellectual property. Personnel should consult with their legal offices if they have questions about using AI-generated content.

Privacy and Confidentiality

Prior to utilizing PII or other regulated data within a Generative AI system, Personnel must consult with their own legal and security teams. Personnel may consult with the Commonwealth Chief Information Security Officer (CISO) and Commonwealth Chief Privacy Officer (CPO), as needed.

Commonwealth Agencies and Offices should use the minimum amount of personally identifiable information (PII) or other regulated data necessary to support the objective of the Generative AI system. When handling PII or other regulated data, Commonwealth Agencies and Offices must take appropriate steps to protect such information based on its level of sensitivity and confidentiality. This includes, but is not limited to, complying with the Commonwealth Enterprise Security Policies and Standards, applicable state and federal laws and regulations, and requirements related to access controls, encryption, data retention requirements, and data sharing agreements.

AI-generated recording tools that record or transcribe a meeting may jeopardize privacy and confidentiality. Personnel must not use any such recording tool unless approved by agency legal counsel or in accordance with agency recording policies.

Compliance with Policies and Regulations

Commonwealth Agencies and Offices utilizing Generative AI systems must comply with the Commonwealth Enterprise Security Policies and Standards. Commonwealth Agencies and Offices that use personally identifiable information or other regulated data within a Generative AI system must comply with applicable state and federal laws, regulations, contractual requirements, policies, and guidelines pertaining to such information or data. Personnel should be aware that Generative AI generated content may be subject to the Commonwealth's public records law.

All Personnel must use their Commonwealth-issued email address when creating accounts or using Generative AI tools. The use of personal email addresses is strictly prohibited.

Procurement of Generative AI Software or Services

Commonwealth Agencies and Offices must notify and seek approval from the Commonwealth Chief Technology Officer (CTO) if they intend to procure Generative AI specific software or services from a third-party vendor. Notification must be made in advance of procuring the product or services. Requests should be submitted in accordance with the existing process established in the applicable statewide contract for procurements that require EOTSS' approval. Pursuant to the Commonwealth IT Terms and Conditions, third party-vendors must comply with EOTSS Enterprise Security Polices and Standards. EOTSS reserves the right to review and approve IT contracts for software or services where Generative AI functionality is included.

AI Inventory

Commonwealth Agencies and Offices must create a Generative AI Inventory documenting all Generative AI Systems currently in use or under consideration and provide said inventory to EOTSS. Commonwealth Agencies and Offices must provide an inventory of all applications currently in use, under development, or in initial discovery phases within their agency that include, either in whole or in part, a Generative AI solution. Reporting will be in a manner prescribed by EOTSS.

Ongoing Evaluation

EOTSS will periodically assess and update these guidelines as appropriate and with input from the Commonwealth's Artificial Intelligence Strategic Task Force. EOTSS will continue strengthening collaborations with academia, other governmental entities, industry and policy experts, organizations representing employees, and community-based organizations to help inform these guidelines.

Establishment of an AI Sandbox

EOTSS has created an AI Sandbox designed to provide a safe and secure testing environment for Commonwealth Agencies and Offices, and other authorized entities, to develop AI use cases with the goal of improving the operation of Commonwealth entities and the provision of government services. The AI Sandbox Program is designed to foster innovation and collaboration while providing the necessary guardrails for the development of Generative AI Systems and tools. Additional details about the AI Sandbox Program are included in the AI Sandbox Program Terms and Conditions, to be released separately.

Contact Information

Questions regarding these guidelines shall be directed to EOTSS Legal/Enterprise Privacy Office at:

Thomas Myers, General Counsel and Chief Privacy Officer, Thomas.B.Myers2@mass.gov

Document Control

Version No.	Revised By	Effective Date	Description of Changes
1.0	EOTSS Legal/Privacy Office	4/12/2024	Initial release of the Guidelines for GenAI