	Georgia Technology Authority	
Title:	**Artificial Intelligence Responsible Use**	
PSG Number:	**SS-23-002**	
Effective Date:	12/12/23	**Review Date:**
Synopsis:	Establishes requirements for the use of AI tools within agency operations	

PURPOSE

As artificial intelligence (AI) technology continues to evolve and further integrates itself into the core of daily business operations, it becomes imperative for the State to engage with AI in an ethical, trustworthy and responsible manner. This Standard seeks to provide -support of the appropriate use of AI tools, while mitigating the risks and undesired outcomes, including AI misapplication, unethical consequences, inherent biases, inaccuracies, and breaches of information security. It also delineates the protocols for the deployment of AI, Generative AI, Deep Learning, and Machine Learning tools within government services and operations, with special emphasis on their utilization in conjunction with sensitive data and state information.

SCOPE and AUTHORITY

O.C.G.A 50-25-4(a)(8) – *State Government, Georgia Technology, General Powers*
O.C.G.A 50-25-4(a)(20) - *State Government, Georgia Technology, General Powers*
PM-04-001 – *Information Technology Policies, Standards and Guidelines*
PS-08-005 – *Enterprise Information Security Charter*

TERMS AND DEFINITIONS

Agency - every state department, agency, board, bureau, commission, and authority but shall not include any agency within the judicial or legislative branch of state government, the Georgia Department of Defense, departments headed by elected constitutional officers of the state, or the University System of Georgia and shall also not include any authority statutorily required to effectuate the provisions of Part 4 of Article 9 of Title 11.

Artificial Intelligence - a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments.

Generative AI - tools or systems used to create models that can generate new and original content, such as images, music, or text, based on patterns and examples from existing data.

Deep learning – model, tool or system used to recognize complex patterns in pictures, text, sounds, and other data to produce insights and predictions.

Machine Learning - computer systems or tools that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data.

STANDARD

1. Responsible Use of AI Tools within Agencies

State agencies are at the forefront of harnessing the potential of these technologies to enhance services, improve efficiency, and drive innovation. One facet of AI that has garnered particular attention is generative AI, a class of systems that have demonstrated remarkable capabilities in tasks such as language generation, content creation, and data analysis.

The responsible use of AI tools including generative AI, is rooted in the principles of ethics, accountability, transparency, and user protection. State agencies have a duty to ensure that AI technologies are deployed in a manner that respects individuals' rights, avoids harm, and adheres to the highest standards of ethical conduct.

Agencies shall:

A. Use AI tools only for their intended purpose and in accordance with ~~any~~ applicable laws and regulations.
B. Ensure that AI tools are used in a manner consistent with the mission, values, and business objectives of the agency by regularly consulting with supervisors, colleagues, and relevant stakeholders to align AI-generated outputs with organizational goals.
C. Exercise due diligence and critical thinking when using AI-generated outputs and guard against intentional and unintentional misuse of AI.
D. Report all tools using AI components to GTA via email at gta.psg@gta.ga.gov.

All entities other than an Agency shall:

A. Disclose to the respective agency in a report, how their products or developed solutions undergo testing to ensure the safety, accuracy, and reliability of data and algorithms.
B. Agree for their solutions to be reviewed for potential ethical, legal, and societal implications.
C. Submit an AI development plan for ongoing monitoring and maintenance to ensure that the system continues to operate safely and effectively and that any new issues or concerns are promptly addressed.

2. Procuring AI Solutions

Agencies shall submit a *Business Solutions Review* (BSR) Request form to GTA (gta.psg@gta.ga.gov) prior to procurement if they believe a new AI tool could be advantageous to business operations. GTA will assess the tool for safety, privacy, and compliance. The State CIO will provide a response to the senior executive of the requesting agency.

Service and technology contracts shall provide standards of third-party provider, and their partners or subcontractors, including:

1. Adhering to GTA policies, standards and guidelines ("PSGs") can be found at Enterprise Policies, Standards, and Guidelines (georgia.gov) https://gta-psg.georgia.gov
2. Adhering to NIST guidance, risk management frameworks and standards for responsible artificial intelligence practices.
3. Providing copies of and, any updates thereof, the Contractor's and Subcontractors' framework and policies concerning responsible AI (RAI), of which shall be provided to GTA upon the Effective Date of the Agreement and within thirty (30) days of any updates. Such policies and framework shall provide in sufficient detail to allow for understanding of Contractor's and any Subcontractor's operationalization of RAI, including the following:
    a. Understanding of Subcontractor's use or proposed use of RAI:
        i. operation of RAI models and stated purpose for each model;

ii. data preparation, enhancement, detection and mitigation of bias;

iii. nature and purpose of tools used to detect bias and other risks;

iv. consistently utilize software engineering best test practices to ensure the RAI is working as intended and be trusted; including:

1. conducting rigorous unit tests to test each component of the system in isolation
2. conducting integration tests to understand how individual components interact with other parts of the GTA environment
3. proactively detect input drift by testing the statistics of the inputs to the RAI system to make sure they are not changing in unexpected ways
4. using a gold standard data set to test the Work Product and ensure it continues to behave as expected. Update this test set regularly in line with changing users and use cases, and to reduce the likelihood of training on the test set
5. conducting iterative user testing to incorporate a diverse set of users' needs in the development cycles
6. applying the quality engineering principle of poka-yoke: building quality checks into the environment so that unintended failures either cannot happen or trigger an immediate response.

v. model(s) monitoring including use of multiple metrics to assess training and monitoring

vi. for models that cannot be explained, include a detailed description of the output ability to enable any output of the RAI or machine learning (ML) model.

b. Bias detection and mitigation:

i. Description of data, and source of such data, used to train RAI models;

ii. Articulate type of data and classification of data being used

iii. Description of how bias is detected and mitigated (or adjusted for bias
iv. Description of elements in creating the model, including parameters and algorithms used with an indication of parameters with the most influence.

c. Agency's right, at its option, to authenticate and verify Contractor, and any partner or subcontractor working under the contract, application and adherence to such standards.

3. Identifying and Mitigating Bias in AI-generated Outputs

To identify and mitigate potential biases in AI-generated outputs, agencies shall:

A. Be aware of common biases that may be present in AI systems, such as data bias, algorithmic bias and confirmation bias.
B. Regularly review and evaluate AI-generated outputs for potential biases and inaccuracies, seeking input from diverse perspectives and stakeholder groups.
C. Use AI tools with transparent methodologies and documentation to better understand their decision-making processes.
D. Collaborate with AI vendors and developers to improve AI systems and address identified biases, reporting any issues and working together to develop solutions.

4. Ensuring Accuracy and Appropriateness of AI-generated Outputs

To ensure that AI-generated outputs are accurate and appropriate, agencies shall:

A. Verify the accuracy of AI-generated outputs by cross-checking with reliable sources, human judgment, and other relevant methods.
B. Review AI-generated outputs for appropriateness, taking into account the context, audience, and potential impact of the content.
C. Establish a system of checks and balances involving multiple reviewers to minimize the risk of errors or inappropriate content.
D. Ensure that AI-generated content is properly reviewed and approved by the business owner or designated personnel before it is published or used for decision-making.
E. Develop and implement guidelines for the responsible use of AI-generated outputs in different contexts and situations, tailored to the specific needs and requirements of the agency.

F. Be aware of risks present when using AI tools such as hallucinations, prompt injections and vulnerabilities, copyright and intellectual property infringements.
G. Maintaining a record of AI tool usage, including the purpose, inputs, outputs, and any actions taken based on the AI-generated results.

5. Data Protection, Safety and Privacy

Agencies shall adhere to the Enterprise's Data Protection Policies, and Guidelines Standards (Data Security Electronic Records SS-08-003; Data Location and Access SS-15-002) when using AI systems. This includes:

A. Ensuring that all confidential information or sensitive and personal data must be authorized by appropriate agency authority for use in the manner proposed and, when used, is anonymized, encrypted, or otherwise protected when used with AI tools.
B. Implementing sufficient security measures to protect data from unauthorized access, modification, or deletion.
C. Obtaining appropriate consent from data subjects when necessary.
D. Reporting any data breaches or incidents involving AI systems to the designated Data Protection Officer (DPO) or other appropriate person(s).

6. Training and Awareness

All users of AI systems shall undergo appropriate training to ensure the responsible and ethical use of these tools and to prevent intentional and unintentional use and abuse. This includes:

A. Being familiarized with this AI Usage Policy and any other relevant policies, guidelines, and best practices.
B. Participating in regular training sessions and workshops to stay updated on AI-related developments, risks, and mitigation strategies.

7. Reporting Misuse

Each agency shall designate a staff member who shall be responsible for ensuring full compliance with this Standard and the requirements set forth herein.

Agencies shall report the following suspected misuse of AI systems, whether intentional or unintentional, to the agency CIO or senior executive and GTA OIS. Reports shall be handled as confidentially as permitted under the Georgia Open Records Act, O.C.G.A. 50-18-72 (25) and (44).

If GTA OIS determines that further investigation is warranted, the agency shall comply and report additional findings. GTA OIS may recommend corrective action and the agency must submit documentation of compliance or request deviation subject to GTA OIS's approval.

Misuse of AI

A. *AI-based fraud*: AI systems used to manipulate or cheat unsuspecting individuals or organizations through phishing scams, identity theft, or fraudulent behavior such as fraudulently issuing loans.
B. *Discrimination*: AI systems that have exhibited bias and perpetuated discrimination, resulting in unequal treatment.
C. *Invasion of Privacy:* AI systems used to gather personal data without the consent of individuals, leading to the violation of privacy rights.
D. *Malicious use:* AI systems used for cyberattacks, such as phishing attempts, social engineering and vulnerability identification and exploitation.
E. *Spreading misinformation:* AI systems used to create and distribute false or misleading information.

Unintentional misuse

A. *Bias and discrimination*: AI system used that has inadvertently reinforced existing biases or discriminatory patterns that have led to unfair treatment of certain individuals or groups.
B. *Privacy violations*: Unintentional exposure of sensitive or personal information through AI systems, either by providing the AI with restricted data or by failing to anonymize or encrypt the data properly before use.
C. *Inaccurate or misleading information:* AI-generated outputs that are incorrect, outdated, or misleading, and have led to poor decision-making, financial losses, or reputational damage for the organization.
D. *Inappropriate content:* AI systems that have generated content that is offensive, politically biased, or otherwise inappropriate for the intended audience or context.
E. *Over-Reliance on AI:* Agencies may unintentionally rely too heavily on AI systems, neglecting to apply their own judgment, expertise, or common sense. This could lead to the adoption of suboptimal solutions, overlooking valuable human insights, or exacerbating existing issues.

8. AI Development and External Collaboration

When developing AI tools or engaging with external collaborators, the following guidelines shall be adhered to:

A. Collaboration agreements must clearly disclose the manner and use of AI technology and define the roles, responsibilities, and expectations of all parties involved. They should also address ownership and usage rights of AI models and data.
B. All AI development and collaborations must be carried out in alignment with this Standard, as well as any other applicable policies or guidelines of the agency.
C. AI development shall prioritize safety, privacy, and ethical considerations.
D. The development process shall include measures to address potential biases in AI systems, such as conducting regular bias audits, seeking diverse input, and using fair and unbiased data for model training.
E. All external collaborators shall commit to upholding the principles of transparency, accountability, and respect for privacy as outlined in this Standard.
F. Prior to deployment, AI systems shall undergo rigorous testing to ensure their safety, accuracy, and reliability. They shall also be reviewed for potential ethical, legal, and societal implications.
G. AI development shall include a plan for ongoing monitoring and maintenance to ensure that the system continues to operate safely and effectively and that any new issues or concerns are promptly addressed.
H. Any potential risks or ethical concerns related to AI development or collaboration shall be reported to the agency's designated Ethics Officer, CIO, and State CIO.

9. Termination and Business Continuity

The State CIO or GTA OIS may order the termination of use of any AI tool that has been found to generate outputs that cause a risk to data privacy and security, reputational damage to the agency or State, or malicious consequences. The agency shall immediately decommission the tool, conduct an investigation to determine the appropriate actions and submit the corrective plan to GTA.

To minimize disruptions in service resulting from an immediate decommission, agencies shall develop and maintain, in accordance with enterprise standard *SS-08-045 Contingency Planning*, a contingency plan for business operations and functions utilizing AI tools.

10. Violations

Any state employee, contractor, or vendor found to violate this standard will be subject to disciplinary action, including but not limited to, termination of employment, contract, or agreement.

11. Audit & Monitoring

Agencies shall maintain and make available to GTA upon request, records of the agency's AI tool monitoring efforts. GTA shall conduct an audit of all AI tools determined to be out of compliance with this Standard.

REFERENCES

GTA Artificial Intelligence Program Webpage - https://gta.georgia.gov/artificial-intelligence

Artificial Intelligence Glossary - https://docs.google.com/spreadsheets/d/1OrCo3QSkZE6MRnmIt6sbAyE2GelhtYaeI5OVlP-2WpQ/edit?usp=sharing

RELATED ENTERPRISE POLICIES, STANDARDS AND GUIDELINES

Artificial Intelligence (AI) Responsible Use Policy PS-23-001
Artificial Intelligence Responsible Use Guidelines GS-23-001