Policy Standards for the Utilization of Artificial Intelligence by the Commonwealth of Virginia

Overview

Virginia is positioned to be a leader and innovator in the emerging artificial intelligence (AI) industry due to its strong data center industry, robust higher education institutions, an expanding workforce, as well as a state government in need of increased efficiency and effectiveness. The technology is already utilized by many Commonwealth agencies to process data, produce automated decisions, enhance customer services and increase government efficiency. The use of AI within government is only expected to increase moving forward, and there needs to be a better understanding by the public as to how AI may be utilized by the Commonwealth. Given the increasing use of AI systems by government, it is imperative to establish comprehensive policy standards for the responsible, ethical and transparent use of AI by all Commonwealth agencies.

Policy Standards

I. The Ethical Use of AI

AI ethics are a set of guiding principles used to ensure that Artificial Intelligence is developed and used responsibly.

a. Every Department, Agency, and Office in the Commonwealth is responsible for and must ensure that its use of Al-related capabilities and the resulting information, including generative Al, is trusted, safe and secure and that it is acting in a responsible, ethical, and transparent manner in implementing such capabilities.
b. AI models must be well documented and available for review
c. AI outcomes must be validated by humans for bias and unintended consequences
d. All Departments, Agencies, and Offices of the Commonwealth shall work to ensure that their use of generative AI and other AI-related capabilities is resilient, accountable, and explainable. Unexplainable (aka, blackbox) AI shall not be used for any decision-making or approval processes.

II. Business Case for the Use of AI

a. Any Department, Agency, or Office in the Commonwealth that considers generative Al and other Al-related capabilities should only deploy the capabilities if there is a positive outcome for the citizens of the Commonwealth, such as reducing wait times for service, removing barriers to access to government institutions and services, limiting bureaucracy and delays in government, cutting the cost in time and money of working with government for individuals and business alike, improving government services and their delivery to the citizens of the Commonwealth, and making Virginia a better, safer, more productive Commonwealth for all of our citizens, regardless of where they live and their socioeconomic status.

b. AI must be determined to be the optimal solution for the stated outcome. Other technology or process applications should be examined before an AI solution is selected. The examination

should include the completion of a regulatory impact analysis (RIA) to assess costs and benefits, including prevalent data, sources, methods, quantification of costs and benefits, and alternatives to proposed change.

c. A clear statement as to the intent of the AI application should be included in the registry discussed in section II. below including whether AI will be relied upon to make a recommendation to the user or if AI will be utilized to make a decision on behalf of the user.

III. Mandatory Approval Processes

a. To ensure the trusted, safe, and secure use of generative AI and other AI-related capabilities, all Commonwealth Departments, Agencies and Offices shall implement the approval process for both internal and external uses of AI systems.
    i. Internal AI systems are those systems employing generative AI or other AI-related capabilities that are only used internally by departments, agencies, and offices to increase efficiency, streamline internal processes, or otherwise improve the way such organizations function internally. They do not include any systems that produce a decision relating to an individual citizen or business within the Commonwealth.
    ii. External AI systems are those systems employing generative AI or other AI-related capabilities that are used to analyze data about individual citizens or businesses within the Commonwealth, make decisions relating to such individual citizens or businesses, produce outputs directly accessible by such individual citizens or businesses.
    iii. Any internal or external AI system which an agency or office seeks to develop, implement, employ, or procure must be entered into a registry maintained by VITA and undergo internal review and final approval by both the agency IT representative and information security officer (ISO). Following agency approvals, the planned AI use and necessary documentation will be reviewed for approval or disapproval by VITA / the CIO of the Commonwealth. The agency Secretariat shall then review and approve or disapprove the AI use. The Secretary of Administration, in consultation with the Governor's Chief Counsel, shall review the ethical issues involved in the AI systems as discussed in section I above. VITA shall retain records of the specific AI use and approvals thereof and ensure that appropriate notifications (such as upon submission to VITA and upon approvals) are sent to both the agency and agency Secretariat.

b. As part of the approval process, the responsible manager and where applicable, the agency IT representative and ISO responsible for approving the development, implementation, employment or procurement of an AI system shall at a minimum, consider the following:
    i. Verify that an AI system is fair and will not result in any unlawful discrimination against any individual or group of individuals, or has any unlawful disparate impact on any individual or group of individuals on the basis of any actual or perceived differentiating characteristic, including, but not limited to: age, genetic

information, color, ethnicity, race, creed, religion, national origin, ancestry, sex, gender identity or expression, sexual orientation, marital status, familial status, pregnancy, veteran status, disability or lawful source of income.

ii. The use of generative AI or other AI-related capabilities will benefit the citizens of the Commonwealth and promote the objectives and key results of the Department, Agency, or Office, in the context of the intended use or application of such capability.

iii. Assess the extent to which human interaction with, and oversight of, the AI system is part of the program utilization.

iv. The potential inherent risks associated with the use of generative AI or other AI-related capabilities in the specific context, including cybersecurity, data protection and privacy, and risks to the health and safety of individuals or businesses in the Commonwealth, and, to the extent such risks are identified, that steps have been taken to address them and mitigate such risk.

v. Additional risk mitigation efforts and guardrails are necessary to protect citizens and business in the Commonwealth, and if so, what efforts are being undertaken and what guardrails are being put in place.

vi. Appropriate stewardship of data held by the Commonwealth.

vii. Completion of a cost impact analysis to assess the costs and benefits, including prevalent data, sources, methods, and alternatives to proposed change.

viii. Whether the developer of the generative AI or other AI-related capabilities is providing or should provide any warranties or assurances regarding the safety and security of the capability, including its cybersecurity and resilience, as well as any warranties or assurances related to its output and, if so, what warranties or assurances are being provided by the developer, if any and any mitigation steps taken to address any lack of such warranties or assurances.

c. The approval process does not apply to the following:

i. AI used in defense of COV security systems.

ii. AI embedded within common commercial products.

iii. AI research and development (R&D) activities or instructional programs at public institutions of higher education.

IV. Mandatory Disclaimers

a. To ensure full transparency, it is the policy of the Commonwealth to require mandatory disclosure to the public when generative AI or other AI-related capabilities are utilized in any process or to produce any decision regarding individual citizens or businesses within the Commonwealth, used to make decisions relating to such individual citizens or businesses, or produce outputs directly accessible by such individual citizens or businesses.

b. Under this policy, all Departments, Agencies, and Offices shall include a disclaimer when an AI system is used in any process or to produce any decision or output.

i. The following disclaimer, or a similar statement, shall be included on any decision or output of an AI system:

"DISCLAIMER: This decision or output was generated by artificial intelligence."

ii. The following disclaimer, or a similar statement, shall be included when AI is used as part of a larger process, but not making the final decision or output:

"DISCLAIMER: This decision or output was created with assistance from artificial intelligence."

c. To add an increased level of transparency, AI systems making external decisions regarding citizens of the Commonwealth shall:
   i. disclose how AI is used to arrive at a decision;
   ii. to what extent human involvement played a role in validating and overseeing those decisions;
   iii. and clearly list options for individuals to appeal those decisions (if appealing is an option).
d. The disclaimer should also include the relevant information regarding any third-party AI products or programs including but not limited to:
   i. which data sets the AI product utilizes;
   ii. the specific cutoff date for the data set inputted into the program;
   iii. any relevant description of the biases embedded in the program;
   iv. any warranties specifically provided by the owner of the AI program and its data outputs.

V. Mitigating Third-Party Risks

a. To mitigate certain third-party risks as required by II(b)(iv) of this Policy, including potential data breaches, unauthorized access, or misuse of personal information, all Commonwealth Departments, Agencies, and Offices shall review and vet any third-party AI developers, system administrators, providers, or contractors. Reviewing and vetting third parties and mitigating third-party risks shall include:
   i. Rigorous and thorough vendor selection by evaluating their ability to deliver value to the Commonwealth and its citizens through the implementation of generative AI or other AI-related capabilities and their trust, safety, and security procedures relative to industry best practices.
   ii. Ensuring third-party vendors apply industry standard best practices when it comes to data collection and utilization, including protecting personally identifiable information and complying with any applicable laws or regulations of the Commonwealth related to such data or information, including as described further herein.
   iii. Review of the results of any testing or red-teaming conducted by the third-party vendor, including testing or red-teaming related to the efficacy of their generative AI or other AI-related capabilities, the cybersecurity and physical security of such capabilities, potential risks associated with such capabilities, and any potential biases that could result in unlawful discrimination, including

algorithmic bias, of such capabilities and work with the vendors to address any issues identified or testing that ought be conducted, including through the use of various methods to improve performance and remove bias that could result in unlawful discrimination, prior to implementation for public use by the Commonwealth

iv. Review of audit reports, product roadmaps, warranties, terms of service, end user license agreements, contracts, and other documentation from third-party vendors to assist in value assessment and risk management activities with respect to acquisition of such capabilities.

v. Maintaining an inventory of third-party material (hardware, open-source software, foundation models, open-source data, proprietary software, proprietary data, etc.) used or required for the use, implementation, and maintenance of all third-party provided generative AI or other AI-related capabilities.

vi. Verification that third-party AI resources and personnel undergo security audits and screenings. Risk indicators may include failure of third parties to provide relevant security information.

vii. Utilization of watermarking technologies for labeling Commonwealth materials produced by generative AI or other relevant AI-related capabilities as a deterrent to data and model extraction attacks.

viii. Consideration of legal and ethical frameworks by ensuring that any AI systems align with existing laws, regulations, and other guidelines, and regularly reviewing and updating these AI systems to ensure continued compliance as the technology and best practices evolve.

ix. Education of government employees and decision-makers about the benefits and risks of AI, including raising awareness about the potential biases and challenges.

VI. Protecting Citizens' Data

a. It shall be the policy of the Commonwealth to prioritize privacy and the protection of citizens' data as agencies and offices move to develop, implement, employ and procure AI systems. To ensure the highest data security and protection, all Commonwealth offices and agencies seeking to develop, implement, employ and procure AI systems shall:

i. Ensure that only the most necessary data is used in AI systems, i.e., ensure that AI systems do not have unrestricted access to vast amounts of personal data.

ii. Secure all data and only keep as long as necessary to complete the intended objective or goal of the AI systems, implementing a timeframe for data retention when feasible.

iii. Monitor for anomalies using approaches such as control limits, confidence intervals, and integrity constraints.

iv. Establish and track AI system security tests and metrics (e.g., red-teaming activities, frequency and rate of anomalous events, system down-time, incident response times, time-to-bypass, etc.)

v. Implement proper user controls, ensuring users know when their data is being used by the AI systems to produce outcomes and make automated decisions.

vi. Allow users the ability to consent to the use of their data by AI systems when possible.
vii. Sensitive, confidential, and protected data shall only be used in private AI systems that are solely accessible by Commonwealth users.