Commonwealth of Virginia

Enterprise Architecture Standard (EA-225)

Enterprise Solutions Architecture [ESA]

Artificial Intelligence

vita.virginia.gov

November 16, 2023

Revision History

ESA Artificial Intelligence: Version History

Revision	Date	Description
1.0	8/4/2023	Initial document created.
1.1	8/21/2023	Updated to insert new Objective 2, insert AI-011, and modify other requirements based on feedback from Gartner.
1.2	11/16/2023	Updated based on public comment received.

Review Process
This requirements document was posted on the Virginia Information Technologies Agency's (VITA) Online Review and Comment application (ORCA). All agencies, stakeholders, and the public were encouraged to provide their comments through ORCA. All comments were evaluated, and individual commenters were notified of action(s) taken.

Requirements and agency Exceptions
The requirements included within this document are mandatory. Agencies deviating from these requirements must request an exception for each desired deviation, and receive an approved *enterprise Architecture Exception* via Archer, prior to developing, procuring, or deploying such technology, or not complying with a requirement specified in this document. The instructions for completing and submitting an exception request are contained within the *Commonwealth Enterprise Architecture Policy*.

Glossary
As appropriate, terms and definitions used in this document are in the COV ITRM IT Glossary. The COV ITRM IT Glossary is available on the VITA website.

References/Links
Other documents referenced or linked in this document are additional resources that agencies or workers may consult to find best practices and guidelines or obtain increased understanding. Unless expressly stated in this document, references or links do not incorporate or require compliance with such other documents.

Contents

Introduction

Vision & Strategy	
Vision	
That Commonwealth of Virginia (COV) agencies and workers shall be able to leverage the creative and insightful capabilities of Artificial Intelligence (AI) while creating public trust by ensuring that the use of such AI does no harm to citizens of the Commonwealth, its guests, the business of the Commonwealth, any known business interest, or the environment.	
Strategy	
Objective 1	Utilize AI Systems to improve the operation of government, the delivery of services, and accessibility through the productivity enhancements from tools ranging from automatic language translation to virtual assistance, speech recognition, task automation, sentiment analysis, text summarization, and documentation.
Objective 2	Ensure that AI Systems are employed safely to win public trust, such that they cause no harm to people, property, businesses, or the environment.
Objective 3	Ensure that the accountability and risks associated with COV decisions must remain with humans to detect and mitigate any risk of bias and discrimination by AI Systems.
Objective 4	Maintain a healthy respect for data security and privacy such that data collected, leveraged, or trained in an AI System is carefully controlled and secured, and that such data is untraceable back to an individual.
Objective 5	Promote transparency as to the use of AI Systems in COV solutions so that their use, extent, and value provided are clear.
Objective 6	Ensure that the use of AI Systems is a sustainable enhancement to the productivity of COV agencies and workers, and never a path to overreliance and the loss of essential skills.

Background

Artificial intelligence (AI) is a subset of the broader disciplines of data analytics and data science. Traditional practices of data analytics study what has already happened to understand and explain the past in a descriptive or diagnostic manner. Descriptive and diagnostic data analytics answers the questions, “What happened? Why did it happen?”

Data science investigates what will happen in the future using predictive analytics (“What will happen?”), prescriptive analytics (“What should we do about what will happen?”), and cognitive/self-learning analytics (“What don’t I know and will need to learn to predict and advise?”).

Data analysis and analytics forms the basis of the scientific method and have long been around. What has recently changed is the economization and consumerization of the software and computational power necessary to provide AI as a consumer service. AI is typically defined as the ability of a machine to perform cognitive actions, such as perceiving, reasoning, learning, interacting with the environment and problem solving.

Purpose

This standard provides requirements for Commonwealth agencies and suppliers on the acceptable and ethical use of AI. It applies to both existing and new uses of AI; stand-alone, AI embedded and generative AI within other systems or applications; AI developed both by the agency or by third parties on behalf of

agencies for the fulfilment of specific agency missions, including relevant data inputs used to train AI and outputs used in support of decision making; and agencies' procurement of AI applications.

For further information on the perspectives identified in this document, please reference the VITA Enterprise Architecture Standard (EA-225).

Scope

This standard is applicable to all Commonwealth agencies (hereinafter collectively referred to as "agencies") that are responsible for the management, development, purchase and use of information technology resources in the Commonwealth of Virginia and applies to all COV AI Systems.

This standard does not apply to:

- AI used in defense or COV security systems
- AI embedded within common commercial products
- AI research and development (R&D) activities or instructional programs at public institutions of higher education

In addition to the requirements below all COV IT technology solutions shall comply with the standards found on VITA's Policies, Standards & Guidelines.

Authority

Code of Virginia, §2.2-2007	Powers of the CIO
Code of Virginia, §2.2-2007.1	Additional duties of the CIO relating to information technology planning and budgeting
Code of Virginia, §2.2-2009(A)	Additional duties of the CIO relating to security of government information
Code of Virginia, §2.2-2012(A)	Additional powers and duties related to the procurement of information technology

Solution Business Requirements

AI-001 COV AI Systems shall be employed for use by COV agencies and suppliers consistent with applicable law.

AI-002 The use of AI Systems by agencies and suppliers within COV shall be limited to those appearing on the Artificial Intelligence technology roadmap.

AI-003 COV AI Systems shall be enumerated in a central registry by the agencies and suppliers employing them and shall provide

- Name
- Purpose
- Sensitivity
- Public Safety
- Technology Used
- Model Architecture
- Model Input
- Model Output Data Type & Structure
- Model Algorithm
- Data Sets Used
- Operation

AI-004 COV AI Systems shall be reviewed for approval by both VITA and the agency Secretariat.

AI-005 COV AI Systems shall be designed and trained to be compliant with legal requirements with respect to consideration of race, color, national origin, sex, religion, genetic information, disability, or other factors where government consideration or use of such factors is regulated or limited by law. Human oversight shall be established and exercised in matters involving such factors.

AI-006 Workflows and decision trees used by COV AI Systems shall be documented in an Architecture Overview Document (AOD).

AI-007 Agencies shall provide periodic opportunities for public participation regarding the function and use of COV AI Systems where such systems involve information about individuals, to inform the public and obtain public input.

AI-008 Agencies shall provide training and awareness programs for employees involved in AI-related activities, covering ethical considerations, bias mitigation, data privacy, accountability, and security.

AI-009 VITA and the Office of Data Governance and Analytics shall establish a Center of Practice that shall assist agencies foster best practices and innovation with COV AI Systems and promote knowledge sharing and collaboration among agencies employing AI Systems.

AI-010 COV AI System data shall be of high quality, complete, accurate, consistent, timely, reliable, and relevant.

AI-011 Agencies and suppliers that intend to incorporate copyrighted material in AI training data shall assess how to do so in a legally compliant manner. Questions about license terms, fair use, or other aspects of using copyrighted material with AI shall be directed to the agency's legal resources.[1]

[1] Use of copyrighted material in AI training data is a developing legal area, as well as a developing technical one. Agencies and suppliers are advised to proceed with the use of such data cautiously.

AI-012 Agencies and suppliers shall establish feedback loops and processes for continuous learning and improvement of COV AI System models based on user feedback, new data, or changing business requirements.

AI-013 Agencies and suppliers shall have business continuity plans for when the supporting AI in a COV AI system fails.

AI-014 Agencies and suppliers shall ensure they have the technical resources to implement and support AI applications and models as well as succession plans for key staff members.

AI-015 Agencies and suppliers shall identify an exit strategy prior to AI implementation to anticipate marketplace volatility due to the rapidly evolving technology space.

Design/Architecture

AI-102 Use of an AI System within any COV Web System shall be labeled as such and displayed per Commonwealth of Virginia Design System (COV Design System) guidance.

AI-103 COV AI Systems must provide a clear, unambiguous, and accessible description of how the AI component is being used, the underlying data sets involved, and the value being provided, displayed per COV Design System guidance.

AI-104 COV AI Systems shall include fail-safe systems that are part of a comprehensive approach to system safety that would keep effective human oversight, resilience, and robustness embedded within the process.

AI-105 Agencies and suppliers shall employ a Human in the Loop where COV AI Systems are integrated into Mission Essential and Business Critical business processes.

AI-106 COV AI System data architecture shall support storing, versioning, monitoring, and deploying AI models throughout their lifecycle.

Availability/Performance

AI -201 COV AI systems that require real-time processing – such as real-time language translation, autonomous vehicles, or online customer support – shall be deployed in a highly available (99.95% or greater) configuration to meet demand.

AI -202 COV AI systems utilizing large datasets and or which have high user traffic must have a defined scalability plan.

AI -203 Agencies with COV AI systems that require real-time processing shall conduct benchmark testing on a bi-annual basis to confirm the system is supported with sufficient infrastructure to minimize latency and interruptions in service while maximizing available throughput.

Capacity

AI -301 COV AI Systems data architecture shall be capable of scaling as the volume of data grows, without performance degradation, with headroom to address immediate spikes in traffic.

AI -302 Agencies and suppliers shall create and maintain capacity plans for data storage required by COV AI Systems that shall be available to VITA on demand.

Continuity

AI-401 COV AI Systems shall be categorized as one of the following

- Mission essential
- Business critical
- Non-critical

AI -402 Mission essential COV AI Systems shall:

- Be deployed in a highly available configuration
- Have a Recovery Time Objective (RTO) and Recovery Point Objective (RPO) of less than 15 minutes
- Subscribe to Disaster Recovery services

AI -403 Business critical COV AI Systems shall have

- RTO of 4 hours
- RPO of 4 – 24 hours

AI -404 Non-critical COV AI Systems shall have an

- RTO of 24 hours
- RPO of 25 – 48 hours

AI -405 COV AI Systems integrated with operational systems —such as environmental controls, transportation systems, or public safety systems — shall include safety systems that manage both everyday scenarios and emergency situations.

Integration/Interoperability

AI-501 COV AI systems should support the following common data formats to enable the exchange of data with other applications and services, including

- CSV
- JSON
- XML

AI-502 COV AI systems shall follow the Artificial Intelligence File Exchange technology roadmap for AI models and formats to address the needs of specific AI domain types.

AI-503 Agencies and suppliers shall document their interoperability requirements in an AOD to ensure seamless integration with other analytics tools, data pipelines, or APIs.

AI-504 Agencies and suppliers shall create and publish guidelines for integrating COV AI System models with existing enterprise systems, data platforms, and processes.

Technology

AI-601 COV AI Systems shall be annually certified for trustworthiness according to the following characteristics defined by NIST AI 100-1:

- Valid and reliable
  - Validation is the confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled
  - Reliability is the ability of an item to perform as required, without failure, for a given time interval, under given conditions
- Safe — Under defined conditions, AI systems should not lead to a state in which human life, health, property, or the environment is endangered

- Secure and resilient — AI Systems can withstand unexpected adverse events or unexpected changes in their environment or use, maintaining their functions and structure in the face of internal and external change and degrade safely and gracefully when this is necessary
- Accountable and transparent
  - The extent to which information about an AI system and its outputs are available to individuals interacting with such a system
  - Confirmation that an AI System is not infringing intellectual property
- Explainable and interpretable
  - Explainability refers to a representation of the mechanisms underlying AI systems' operation
  - Interpretability refers to the meaning of AI systems' output in the context of their designed functional purposes
- Privacy-enhanced — Design support to help safeguard human autonomy, identity, and dignity.
- Fair with harmful bias managed — Concerns for equality and equity by addressing issues such as harmful bias and discrimination

AI-602 COV AI Systems shall be subject to Test, Evaluation, Verification, and Validation (TEVV) review throughout the AI System lifecycle to ensure the models are producing highly accurate responses.

- Internal and external validation of assumptions for system design, data collection, and measurements relative to the intended context of deployment or application
- Model validation including model drift and Algorithm Risk assessment
- System validation and integration in production, with testing, and recalibration for systems and process integration, user experience, and compliance with existing legal, regulatory, and ethical specifications
- Ongoing monitoring for periodic updates, testing, and subject matter expert (SME) recalibration of models, the tracking of incidents or errors reported and their management, the detection of emergent properties and related impacts, and processes for redress and response

AI-603 COV AI Systems shall support the ability to be disengaged or deactivated if the system demonstrates unintended or unexpected behavior.

AI-604 Agencies and suppliers shall ensure the availability of high-quality, diverse, and relevant data to train and validate AI models which have been copyright cleared for use by the COV.

AI-605 COV AI System data quality shall be ensured through appropriate data cleansing, validation, and imputation mechanisms.

AI-606 COV AI System data shall be governed by clear policies and procedures, including precise definitions of data access rights, categorization rules, and responsibility assignments for data accuracy and protection.

AI-607 COV AI System data shall be managed throughout its entire lifecycle, from creation to retirement, including processing, storage, archival, and deletion.

AI-608 COV AI System data shall be traceable, with clear records of its origins and changes over time, to ensure the authenticity and reliability of the data. documentation of data lineage, including where the data came from and how it was altered.

AI-609 COV AI System data shall be standardized across the enterprise in terms of structures, formats, and models to ensure uniformity and interoperability.

AI-610 COV AI System data shall be managed with version control systems to track changes over time to ensure the data used for training and testing AI models can be accurately reproduced and verified.

AI-611 Metadata associated with COV AI System data shall be effectively managed, including data content, format, source, and ownership.

AI-612 Agencies shall define data collection, preprocessing, and transformation standards to ensure data consistency and accuracy for COV AI System.

AI-613 Agencies shall implement processes for developing and validating COV AI System models, including model selection, training, evaluation, and hyperparameter tuning.

AI-614 Agencies shall define guidelines for COV AI System model deployment, versioning, and monitoring to ensure ongoing performance and accuracy.

AI-615 COV AI System models shall be capable of providing explanations or insights into their decision-making processes to facilitate trust and understanding.

AI-616 Agencies and suppliers shall establish methods for capturing and documenting the rationale behind COV AI System model outputs to support audits and regulatory compliance.

AI-617 Agencies and suppliers shall implement version control and documentation processes to track COV AI System model iterations, improvements, and updates.

Security

AI-701 COV AI Systems shall conform to Commonwealth security standards.

AI-702 COV AI System decision paths shall be logged for traceability and replay.

AI-703 Prior to use in a COV AI system, all data consumed by that system must undergo data registration with a metadata repository identifying it as a source of AI data and identified with approval from the data owner.

AI-704 COV AI Systems shall limit the use of facial recognition technology to the use of biometric authentication; it shall not be used to monitor, track, or maintain a record of an individual.

AI-705 COV AI Systems shall not disseminate Personally Identifying Information (PII) collected in any operation to third parties

AI-706 Only Commonwealth public data (data already being shared with the public), such as public data being shared on the Commonwealth Data Trust, shall be used in a publicly available AI System.

AI-707 All public datasets shall be registered with the COV Data Catalog.

AI-708 Commonwealth non-public data, regardless of its level of confidentiality, shall only be used with approved Commonwealth AI applications and services running in Commonwealth public and private cloud environments.

AI-709 COV AI Systems shall not use data sets that include confidential data and shall ensure that any data derived from individuals is appropriately.

AI-710 COV AI Systems shall employ data quality tools that can score the data and shall meet a minimum accuracy threshold of 90%.

AI-711	COV AI Systems that include address, email, and phone numbers shall employ validation software specific to those data types.
AI-712	Agencies and suppliers shall comply with robust security measures and privacy methods to protect sensitive data used by AI models through encryption, anonymization, or data minimization techniques.
AI-713	AI monitoring shall detect model degradation, concept drift, or biases and trigger alerts or retraining processes.
AI-714	AI monitoring shall include resource utilization and resource utilization patterns to ensure efficient and cost-effective operations.
AI-715	AI monitoring shall include metrics and thresholds for model performance and accuracy.
AI-716	Agencies and suppliers shall implement measures to protect AI models and algorithms from unauthorized access, tampering, or theft.
AI-717	Agencies and suppliers shall enforce strict access controls and authentication mechanisms to prevent unauthorized use or modifications of AI systems.
AI-718	COV AI System data shall be managed in compliance with applicable data protection laws and regulations.

Definitions

As appropriate, terms and definitions used in this document are included in the COV ITRM IT Glossary.

Anonymization	A process that removes the association between the identifying data and the data subject through a combination of masking and de-identification.
Architecture Overview Document (AOD)	A document format used by VITA for agencies and suppliers present their architectures to demonstrate that they are in compliance with approved design patterns and standards. Architecture Overview Document (AOD) is split into 3 sections to be completed at different times in the lifecycle of service deployment. Each architecture will include a High-Level Section (HLS), Detailed Design Section (DDS), and an As Built Section (ABS).
Artificial Intelligence (AI)	The simulation of human intelligence processes by machines, especially computer systems, such that it can adapt and learn on its own using machine learning algorithms that can analyze large volumes of training data to identify correlations, patterns, and other metadata that can be used to develop a model that can make predictions or recommendations based on future data inputs.
AI Bias	AI Systems that systematically and unjustifiably yield less favorable, unfair, or harmful outcomes to members of specific demographic groups.
AI System	An engineered or machine-based artificial intelligence that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. They are designed to operate with varying levels of autonomy. AI Systems include solutions that exhibit Strong AI, such as Generative Artificial Intelligence, and Weak AI, such as Robotic

	Process Automation.
Automation Complacency	A state that occurs under conditions of multiple-task load, when manual tasks compete with the automated task for the operator's attention. Automation complacency is found in both naive and expert participants and cannot be overcome with simple practice
Business Critical	Business processes or systems that are essential for an agency to keep running, for which an outage produces significant disruption to the agency's mission. They are core to the agency's function, but which are not necessary to ensure immediate survival during outages and other disasters.
COV AI System	An AI System that has been registered with Commonwealth Security and Risk Management (CSRM) and has been approved for use.
De-identification	Reducing the risk of identifying a data subject to a very small level by applying a set of data transformation techniques such that the resulting data retains very high analytic utility.
Deskilling	The process by which skilled labor within an industry or economy is eliminated by the introduction of technologies operated by semi- or unskilled workers.
Human-in-the-Loop	A model requiring human interaction, used in the sense of humans aiding the computer in making the correct decisions.
Masking	Reducing the risk of identifying a data subject to a very small level by applying a set of data transformation techniques without any concern for the analytic utility of the data.
Mission Essential	A business process or system that must function continuously for an agency to be successful, for which the impact of an outage is immediate and catastrophic to the agency's mission.
Non-critical	A business processes or system for which an outage will not significantly affect an agency's mission.
Object Character Recognition (OCR)	A technology that converts any kind of images containing written or printed text into a machine-readable format.
Personally Identifiable Information (PII)	Any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means.
Privacy	Norms and practices that help to safeguard human autonomy, identity, and dignity. These norms and practices typically address freedom from intrusion, limiting observation, or individual's agency to consent to disclosure or control of facets of their identities.
Robotic Process Automation (RPA)	A form of business process automation that is based on software robots or artificial intelligence agents. It is sometimes referred to as software robotics.
Strong AI	An artificial intelligence that constructs mental abilities, thought processes, and functions that are impersonated from the human brain. It includes models for Artificial general intelligence, Human-

	level intelligence, Superintelligence, and Artificial Consciousness.
Test, Evaluation, Verification, and Validation (TEVV)	A lifecycle model developed by the US Department of Defense (DOD) for machine learning and deep learning systems defined in 2019.
Weak AI	Artificial intelligence that is focused on mimicking how humans perform basic actions such as remembering things, perceiving things, and solving simple problems.

References

- M-21-06 Guidance for Regulation of Artificial Intelligence Applications
- NIST AI 100-1 Artificial Intelligence Risk Management Framework (AI RMF 1.0)

Appendix – Definitions of Artificial Intelligence

Artificial intelligence (AI) is a subset of the broader disciplines of data analytics and data science. Traditional practices of data analytics study what has already happened to understand and explain the past in a descriptive or diagnostic manner. Descriptive and diagnostic data analytics answers the questions, "What happened? Why did it happen?" Data science investigates what will happen in the future using predictive analytics ("What will happen?"), prescriptive analytics ("What should we do about what will happen?"), and cognitive/self-learning analytics ("What don't I know and will need to learn to predict and advise?"). Data analysis and analytics forms the basis of the scientific method and have long been around. What has recently changed is the economization and consumerization of the software and computational power necessarily to provide AI as a consumer service.

As the highest form of data analytics, AI is typically defined as the ability of a machine to perform cognitive actions, such as perceiving, reasoning, learning, interacting with the environment and problem solving. Artificial intelligence can be generally classified into three types: Artificial Narrow Intelligence, Artificial General Intelligence and Artificial Super Intelligence. The differences between them all are based on their perception capabilities. Artificial Narrow Intelligence (ANI) can only perform a specific task autonomously and has limited, narrow capabilities to perceive the systems they've been programmed to interact with. Artificial General Intelligence (AGI): expands beyond a single system and can perceive and function completely like a human being by using multiple capabilities and forming connections to multiple topic areas. Artificial Superintelligence (ASI) will be exceedingly more capable because of a significantly greater memory along with faster data processing and analysis to perceive and understand the world around the system in real time.

Currently the industry is making very early steps towards AGI with applications like ChatGPT being one of the most relevant examples of an implementation making its way towards AGI. Though the ChatGPT model, conversational user interface, and prose response can seem very much like interacting with a person; ultimately the model used is still only a large language model (with over 500 billion words and word combinations). When requesting items such as mathematical answers or other solutions to logic issues the model may not provide accurate responses. ChatGPT also has a disturbing tendency to "hallucinate" and provide irrelevant and inaccurate responses in some situations.

Machine Learning

Prior to the recent focus on artificial intelligence a large focus was put on a related field called machine learning (ML). These two terms are often used interchangeably however they have a different meaning. Machine learning (ML) focuses on the development of algorithms and models that enable computers to learn from data and make predictions or decisions without being explicitly programmed. ML systems use statistical techniques to learn patterns and relationships from data. They rely on training data to train models that can generalize and make predictions on previously unknown data.

Deep Learning

Deep learning is a type of machine learning that can process more than just text-based data types. Deep learning can process images and requires less human involvement with more accurate results than traditional machine learning. Deep learning techniques leverage a neural network model where data is processed through multiple iterations that learn features of the data. The neural network can then use what it learned to classify new data and determine if an object meets the learned classification criteria. For example, the model learns what a stool looks like, it can recognize the object in a new image.

Generative Artificial Intelligence

TechRepublic defines generative AI as a subfield of artificial intelligence in which computer algorithms are used to generate outputs that resemble human-created content, be it text, images, graphics, music, computer code or otherwise. Generative AI is the type of AI that has been the area where the most significant and publicized progress has been made recently. In the case of ChatGPT, generative AI is classified as a model that can generate narrative content in response to a prompt. A prompt is either a

question or reference to a subject the user is interested in finding information on. Crafting a prompt which returns the most efficient set of results requires some testing and configuring from the user. Some speculations indicate prompt engineering will be incorporated into job duties when generative AI becomes more pervasive in the workplace. Tools like ChatGPT and DALL-E (a tool that generates art) have the potential to change the way people access information. The full breadth of that change is still being debated, however the initial areas where this technology is being applied is to search and analyze online data including both public data and private enterprise data.