Information Technology Policy

Artificial Intelligence General Policy

Number ITP-BUS012	**Effective Date** September 26, 2018
Category Business	**Supersedes** All Prior Versions
Contact RA-ITCentral@pa.gov	**Scheduled Review** August 2023

1. Purpose

Establishes an overview and guidelines for the integration of artificial intelligence (AI) technologies and capabilities into Commonwealth business and decision processes.

2. Scope

This ITP applies to all offices, departments, boards, commissions and councils under the Governor's jurisdiction (hereinafter referred to as "agencies"). Agencies not under the Governor's jurisdiction are strongly encouraged to follow this ITP. Third-party vendors, licensors, contractors, or suppliers shall meet the policy requirements of this ITP that are applicable to the products and services provided to the Commonwealth.

3. Background

AI technologies allow computers and machines to function in an intelligent manner by creating self-learning machines that are capable of reasoning, planning, solving problems, thinking abstractly, comprehending complex ideas, and learning. Agency business areas are leveraging AI technologies to advance discovery, , transform decision making, and improve business outcomes. A risk with AI systems is that Commonwealth Data could be used in unintended ways leading to potentially negative consequences. Establishing the appropriate governing oversight, combined with compliance to the necessary policies, can improve the services provided to both internal and external customers.

4. Definitions

4.1 **Anomaly:** An unplanned unexpected variable that differs from expectations

4.2 **Bias:** Refers to the gap between a predicted value and the actual value.

4.3 **Material Decisions:** A decision that has a significant legal, financial, human resource, legislative, organizational, or regulatory impact. This includes but is not limited to, program eligibility, benefits determinations, and decisions impacting the health, safety, and welfare of Commonwealth citizens or employees.

4.4 **Overfitting:** A prediction-based outcome that has low Bias and high Variance.

4.5 **Right-To-Explain (RTE):** A concept that requires an AI service provider to satisfactorily detail an AI algorithm's use of a user's input data to formulate output data.

4.6 **Robotic Process Automation (RPA):** A technique utilizing automation and AI technologies to handle high-volume, repeatable tasks to streamline business processes.

4.7 **Underfitting:** A prediction-based outcome that has high Bias and low Variance.

4.8 **Variance:** A statistical measurement used to determine the spread of a set of random datapoints from the average value. "Low" Variance datapoints are grouped tightly (densely) together. "High" Variance datapoints are grouped loosely (spread out).

5. Objective

To ensure the proper due diligence is established and exercised for business automation solutions enabled with AI technologies relative to governance, data management, modeling, architecture frameworks, testing/validation, risk management, and security.

6. Policy

6.1 Key Considerations

Before agencies begin an AI initiative, they shall evaluate the following key considerations:

Decision Making

Agencies shall evaluate how much decision-making they allow the AI solution to make. Material decisions which result in a programmatic or financial outcome should be carefully evaluated as to avoid unintended consequences if transferred to an AI solution. (Refer to *RFD-BUS012B, Artificial Intelligence IT Policy Guideline* for further details)

Data Sets

AI solutions will rely on a pre-determined data set to complete the processes they have been assigned. Data sets identified should be reviewed to ensure quality and integrity of the data sets to minimize Overfitting and Underfitting due to Bias and Variance errors in the model that could alter the solution's decision-making patterns.

Methodology

Development and maintenance of an AI solution is different than a traditional IT system. AI solutions need regular and ongoing review to ensure the algorithms and models used yield the best possible result and not producing unintended consequences as changes in the data or business processes occur. As a result, business and technical decision makers need to apply a machine learning test-and-learn mentality to establish successful data analysis and determine the best model to use.

Audit

Agencies shall ensure they understand the algorithm and model established for the AI solutions decision making patterns. Sample data used to test and validate the algorithm shall be retained in the event an audit takes place.

Note: AI solutions that are used to formulate decisions regarding the topics listed below will be subject to audits:

a) Direct or indirect material financial interests or transactions;

b) Administrative policy and program changes;

c) Benefits eligibility and determinations;

d) Life-changing, and;

e) Health, safety, and welfare of citizens or Commonwealth employees.

Disclosure

When a customer is interacting with an AI solution on behalf of an agency, the solution shall disclose to the customer that they are interacting with an AI solution

6.2 Readiness Assessment

Prior to the adoption of any AI enabled solutions, agencies shall conduct a readiness assessment and have a general understanding of AI. (Refer to *RFD-BUS012A, Commonwealth Artificial Intelligence Assessment Tool* and *RFD-BUS012B, Artificial Intelligence IT Policy Guidelines* for further details)

6.3 Governance

Proper governance of AI solutions is required prior to deployment of any AI solutions into the enterprise. Governance and appropriate oversight mitigate risks associated with emerging technologies. Agencies shall use existing governance bodies to ensure overall impact to business and technology operations are not negatively impacted by the integration of AI solutions. Governing bodies are responsible for the continuous monitoring and outcomes of AI solutions to ensure alignment with business and technology strategic objectives.

Governing bodies are recommended to provide oversight for the following:

- Examine the social, economic, and legal impacts of AI adoption on the workforce, citizens, and business operations.
- Determining the conditions and constraints in which supervised and unsupervised techniques will be used for training AI and algorithmic decision-making systems.
- Legal reviews required for use of third-party AI services, contracts, licenses, agreements, and specific use cases of AI solutions with potential impacts to workforce.
- **Note:** it is important to understand the potential liabilities with intellectual

property and data ownership associated with third party entities. Completion of the Cloud Use Case Review process for cloud-based AI solution (refer to *ITP-SEC040, IT Service Organization Management Cloud Requirements*)

- Legal requirements regarding transparency and disclaimers for public engagement and use of AI systems. As well as the RTE that will obligate Commonwealth agencies to explain the purpose of an algorithm and the kind of data it uses when making automated decisions. This includes third-party AI solutions. Agencies shall validate (understand) the functionality of third-party AI algorithms and how the data collected and utilized is managed by the third-party solution. The following elements shall be captured for any AI solution to satisfactorily comply with an RTE request:
  - Technical/design details of the AI system and algorithms
  - How the AI system was trained (including personnel and documentation)
  - How the AI system works (i.e. what are the inputs and outputs)
  - Data sources (documentation of all data sources)
  - Audit Logging (refer to *ITP-SEC040, IT Service Organization Management Cloud Requirements* for audit logging requirements)
  - Change Management details and documentation that impact the AI system algorithms (i.e. decisions, inputs, outputs)
  - Testing and validation results
  - Timeframe documentation (captures time periods of testing, validations, governance approval, deployment, and other critical milestones the of AI solution)
  - Human elements (any personnel information that will assist in the RTE request)
- Decisions made by AI that have legal, financial, human resource, legislative, organizational, or regulatory impact must include a human verification process.
- Evaluate and authorize AI: technology architecture frameworks, software, platforms, libraries, software as a service (SaaS), platform as a service (PaaS), and infrastructure, and relevant tools that can be properly integrated and supported in our IT ecosystem and securely interface with our back-end services/systems.
- Protocols and procedures for assessing and handling inquiries or accidental events regarding AI system anomalies with priority given for decisions that have potential implications for public safety or perceived workforce/labor discriminatory practices.

6.4 Data Management

Machine learning algorithms learn from data. It is critical to subject them to the right data for the problem to be solved. Even if there is a reliable and relevant data source, it is imperative to develop proper methods for data evaluations and preparedness to make sure that it is in a useful state, scale, format, composition, and representative to the problem being solved.

- Institute proper data and information management controls, procedures, and processes for data set selection, evaluation, and preparation for use with AI solutions.
- Data availability, quality, and integrity are critical for AI systems. AI systems should not be trained with data that is biased, inaccurate, incomplete, or

misleading. All AI training shall be vetted through the appropriate governing processes.

- Create procedures for properly parsing data sources used with AI systems models into multiple randomized data sets consisting of training, cross-validation, and test data.
- AI systems should have access to and use only what data sources they need.
- Establish data validation procedures and processes to select, analyze, clean, and certify the quality and integrity of the data sources that will be used for AI automation solutions.
- Institute processes and procedures for preprocessing and transforming the selected data set to format, clean, sample, decompose, and aggregate the data to ensure alignment with the model and the problem being solved.

6.5 Model Testing and Validation

Machine algorithms are complex and requires expertise and practical experience in determining and implementing the best machine learning algorithms to solve the problem and form accurate outcomes. Equally important is the proper testing and validation of the model to determine the degree of Underfitting, Overfitting, and errors related to Bias and Variance. Modeling and testing methods shall be established to:

- Use AI measurement methods (accuracy, recall, and precision metrics) to evaluate each model's performance and to choose the best model to solve the problem and produce the best results.
- Use multiple randomized data sets consisting of training, cross-validation, and test data to determine the best model and minimize potential Underfitting and Overfitting resulting from Bias and Variance errors.
- Define, validate, and document execution of hand-off criteria as to when judgment and decisions from an AI system are transitioned to a human.

6.6 Security

Safety and security must be considered regarding full disclosure and transparency of machine designs, algorithmic models, and decisions. The following shall be considered for designing all AI systems:

- Evaluate the level of risk that AI systems are exploited by malicious actors and determine appropriate risk controls.
- Establish controls to prevent adversarial learning to include attacks that try to influence the training data of spam filters or systems for abnormal network traffic detection, designed to mislead the learning algorithm for subsequent exploitation.
- AI systems vulnerability scanning methods and techniques need to be enhanced for the discovery and categorization of security vulnerabilities or other design flaws and appropriate mitigation or resolution requirements to address known vulnerabilities. (Refer to ITP-SEC005, *Commonwealth Application Certification and Accreditation (CA^2)*)
- Expand incident management procedures and processes for proper handling of AI systems cybersecurity attacks or security findings to those who are in the best position to fix the problem. (Refer to *ITP-SEC024, IT Security Incident Reporting Policy*)

- AI systems are required to comply with all Commonwealth IT Policies, Management Directives, and any other applicable regulations.
- AI systems should only collect, use, share, and store data in accordance with privacy and personal data laws and best practices. (Refer to Section 8 for list of Security-based ITPs)
- Establishing AI solution risk profiles based on a set of criteria to categorize and regulate the degree of oversight, review, controls, testing, documentation, and validation required pre and post deployment of AI solutions into our business and technical ecosystems. The NIST AI Risk Management Framework is currently being developed. An initial draft is available at: https://www.nist.gov/itl/ai-risk-management-framework

7. Responsibilities

7.1 Agencies shall

- Ensure the proper due diligence is exercised in the design, development, testing, validation, adoption, and deployment of business automation solutions and services that integrate AI technologies.
- Institute industry best practices and implement controls and processes to comply with the requirements outlined in this ITP.

7.2 Office of Administration, Office for Information Technology shall

Comply with the requirements as outlined in this ITP.

7.3 Third-party vendors, licensors, contractors, or suppliers shall

Comply with the requirements as outlined in this ITP that are applicable to the products or services they are providing to the Commonwealth. If the products or services being provided by the third-party vendor, licensor, contractor, or supplier do not fall within the scope of this ITP, compliance is implied. If the third-party vendor, licensor, contractor, or supplier subsequently deploys products or services that fall within the scope of this ITP in the future, compliance with the policy is required.

8. Related ITPs/Other References

- Definitions of associated terms of this policy are published on the Office of Administration’s public portal: http://www.oa.pa.gov/Policies/Pages/Glossary.aspx
- Commonwealth policies, including Executive Orders, Management Directives, and IT Policies are published on the Office of Administration’s public portal: http://www.oa.pa.gov/Policies/Pages/default.aspx
- *Management Directive 205.34 Amended, Commonwealth of Pennsylvania Information Technology Acceptable Use Policy*
- *RFD-BUS012A, Commonwealth Artificial Intelligence Assessment Tool*
- *RFD-BUS012B, Commonwealth Artificial Intelligence Guidelines*
- *ITP-ACC001, Digital Accessibility Policy*
- *ITP-INF000, Enterprise Data and Information Management Policy*

- *ITP-INF001, Database Management Systems*
- *ITP-INFRM005, System Design Review of Electronic Systems*
- *ITP-SEC000, Information Security Policy*
- *ITP-SEC005, Commonwealth Application Certification and Accreditation (CA^2)*
- *ITP-SEC019, Policy and Procedures for Protecting Commonwealth Electronic Data*
- *ITP-SEC024, IT Security Incident Reporting Policy*
- *ITP-SEC025, Proper Use and Disclosure of Personally Identifiable Information (PII)*
- *ITP-SEC031, Encryption Standards for Data in Transit*
- *ITP-SEC040, IT Service Organization Management Cloud Requirements*
- *ITP-SFT000, Software Development Life Cycle (SDLC) Policy*
- *ITP-SEC041, Commonwealth IT Resources Patching Policy*
- *NIST AI Risk Management Framework (initial draft)*

9. Authority

Executive Order 2016-06, Enterprise Information Technology Governance

10. Publication Version Control

It is the Authorized User's responsibility to ensure they have the latest version of this publication, which appears on https://itcentral.pa.gov for Commonwealth personnel and on the Office of Administration public portal: http://www.oa.pa.gov/Policies/Pages/default.aspx. Questions regarding this publication shall be directed to RA-ITCentral@pa.gov.

11. Exemption from this Policy

In the event an agency chooses to seek an exemption from the guidance within this ITP, a request for a policy waiver shall be submitted via the enterprise IT policy waiver process. Refer to *ITP-BUS004, IT Policy Waiver Review Process* for guidance.

This chart contains a history of this publication's revisions. Redline documents detail the revisions and are available to CWOPA users only.

Version	Date	Purpose of Revision	Redline Link
Original	09/26/2018	Base Document	N/A
Revision	09/9/2020	Policy Refresh - No changes needed	N/A
Revision	08/29/2022	ITP Refresh Updated references and links Updated scope to include third parties if providing applicable products or services to the Commonwealth Updated definitions and replaced with links to glossary where applicable Minor clarifications and grammatical updates	Revised IT Policy Redline <08/29/2022>

		Added reference to the initial draft of the NIST AI Risk Management Framework	