Information Technology Policy

Artificial Intelligence IT Policy Guideline

ITP Number	**Effective Date**
RFD-BUS012B	September 26, 2018
Category	**Supersedes**
Business	None
Contact	**Scheduled Review**
RA-ITCentral@pa.gov	September 2021

Table of Contents

Introduction

This reference document is to be used as a primer for commonwealth agencies exploring artificial intelligence and machine learning solutions. The guide is meant to supplement ITP-BUS012 *Artificial Intelligence General Policy* and is to be consulted during any strategic planning and designed of artificial intelligence/machine learning solutions. Artificial intelligence is a disruptive technology and its complexity prevents a comprehensive coverage within this document. The document focuses on high-level readiness, chatbots, robotic process automation, machine learning, an overview of evaluating models, and considerations when exploring artificial intelligence vendors.

Am I ready? (Criteria Checklist)

- Mechanisms in place for investigating innovative strategic and tactical options of artificial intelligence (AI) solutions ensuring value of AI opportunity, mission alignment, and total cost of ownership (TCO) are reasonable?
- Established metrics and measure for AI solutions and process performance that are important to and directly serve your end-users and/or citizens?
- Do application security scanning capabilities exist to identify areas of risk for known vulnerabilities?
- Are the technologies being used in the solution align with the adopted AI architecture frameworks and standards?

- Do you have mechanisms in place to validate and measure performance consistently and comprehensively for AI-enabled solutions?
- What is your capability to build and support data driven applications that run on large streams of big data?
- Do contractual terms and conditions exist specifying the obligation of each party for AI solutions developed and maintained using third party entities?
- Do you have data/information management governance, processes, and procedures in place regarding the analysis, protection, selection, parsing, cleansing, testing, validation, and accessing the integrity of data sets for the development and continued use with AI solutions?
- Do you have AI and process design strategies in place according to best practice guidelines?
- Have you established procedures for AI solution adoption and assessing the potential impacts on existing business model, services, processes, roles, and functions as well as determining if the organization is ready to make the necessary changes?
- Do you have quality assurance and risk management programs in place for understanding how AI systems/models must be validated and developed?
- Are proper security detection and safeguards in place for business solutions with embedded AI components?
- Is there a high level of maturity and experience with agile-scrum project management, solution development and delivery?
- Are mechanisms in place to evaluate the risks relative to privacy, ethical, labor, and social impacts to end-users or citizens for solution with AI capabilities?
- Do you have effective end-point management to deliver high-performing and secure AI mobile experiences?
- Do you have the capability to provide dynamic user experiences on the application layer for AI-enabled solutions that require cloud maturity, including a low latency network, cloud-based engines for in-stream big data processing and secure data storage?
- Is a strong identity and access management (IAM) foundation in place to provide the necessary bridge between legacy and smart apps to automate user-specific provisioning and enable single sign-on capabilities?
- Do you have a dynamic team model approach with accessibility and availability to AI specialists for the design, development, and deployment of AI-enabled solutions?
- Do you have adequate business and IT governance frameworks in place to provide oversight, arbitrate, and render decisions regarding AI-enabled solutions?
- Do you have the necessary back-end services, compute, and infrastructure resources?
- Are mechanisms in place to address the legal requirements regarding the “right-to-explain” 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.
- Do protocols and procedures exist for assessing and handling inquiries and/or accidental events regarding AI system anomalies with priority given for decisions that have implications for public safety or result in discriminatory practices?
- Is there documentation covering procedural compliance, accountability, incident/problem resolution regarding the use of government contractors, open source or proprietary AI tools, applications, and/or services?
- Do have the capability to evaluate the level of risk that AI systems are exploited by malicious actors and determine appropriate risk controls?

When Should I Use a Chatbot?

Organizations should not to rush to build chatbots to ensure they're not behind the curve, but it's important to consider whether the chatbot should exist in the first place. Chatbots are computer algorithms that are integrated into organizations websites, mobile applications, or messaging platforms to better assist and enrich the end-user experience. Computer algorithms excel at finding information quickly, which is precisely what most chatbots do. They look for keywords in message inputs and use them to respond with the information a user need. Most chatbots stumble when it comes to answering complex questions, but that doesn't mean they are not useful. Chatbots are often used for the following:

- Process Automation: automation of simple high-volume routine processes that that have a known precise output for a specific input that are the same each time.
- Handling Queries: assisting end-users with routine simple requests, searches, repetitive queries, or frequently asked questions
- User Support: when running a user service website, a chatbot can save the service desk from having to be on-call 24/7 by offering prepared answers to typical requests during live chat sessions.
- Data Collection: websites that include signup forms to be used to opt into mailing lists. A chatbot can replace those forms and provide a more engaging experience.
- Intermediary/Assistant: act as a gateway to redirect and provide referrals to end-users based on replies to other agency resources or other enterprise chatbots (also known as concierge bots).

Citizens will use chatbots to accomplish a certain task, whether it's checking the status, ordering a custom license plate, resetting their password, or listening to an informational recording about a specific agency program or service offering. Integrating Natural Language Processing (NLP) capability in the chatbot can spark an amazing user experience, but it's essentially useless, unless the underlying chatbot can help people accomplish the task. The chatbot should be designed to improve business operations and lower friction for consumers who aren't sure what to do, or where to go, or the bot makes it easier for the end-user to complete an online transaction because the bot has instantaneous access to the right information and the back-end services. Before an agency sets off to build the ultimate chatbot for their end-users/citizens, they should consider six simple questions:

1) What task will end-users accomplish using a chatbot?

   Set a goal and concrete objectives for how the chatbot will serve the end-users. Think about common pain points for the agency or users, and brainstorm ways the chatbot can offer solutions using design thinking techniques and human-centered design approaches. Determine the appropriate metrics to see whether your chatbot has met the primary goal and objectives.

   Note: Users might not be excited about the chatbot and/or may need some convincing that the chatbot is useful. User engagement and feedback is very important. It is recommended that agencies obtain the end-user's perspective up-front as well as beta test the chatbot with real users monitoring reaction and gathering data on what the users are doing, interactions with chatbot, and validating the chatbot responses to ensure that it is working correctly.

1) What process or interaction specifically is to be automated and why?

Chatbots are ideal for the automation of frequent routine repetitive tasks that directly engage or interact with the user; saving time, improving user experience, and increasing productivity. They are new tools agencies can use to improve operational efficiencies and build user experiences in a new environment beyond the web and apps. Agencies should carefully choose and analyze the process they wish to automate relative to their respective inputs, outputs, complexity of activities, decision points, time durations, and resources used in accomplishing the process activities. Starting out, agencies should focus on low risk, high volume, simple routine processes that have a known precise output for a specific input that is the same each time. Agencies should consult with AI subject matter experts (SMEs) to explore solution alternatives, conceptualize, and examine the future target end-state process enabled with AI solution. Define the business case outlining goals and objectives for the automation initiative with critical success factors and key performance indicators. Establish metrics based on hand-off counts, successful engagements, response times, chatbot versus human interaction/transaction counts (before and after), and user feedback to determine the effectiveness of the chatbot, improve chatbot's capabilities and user experience, as well as the determine if agency's automation objectives have been met.

Consider whether a chatbot with integrated Natural Language Processing (NLP) would improve that process. According to Gartner, a third of all user service interactions still require another human to perform the desired task. NLP might help users feel as if they are interacting with another human, but it will not change the functional capabilities of the chatbot. Make sure the agency is offering a better or at least different value-add solution.

3) Why is a chatbot better than other interfaces?

The more value-add functionality you inject into the user experience, the more likely users will engage with the chatbot. You cannot expect a user to use your chatbot unless it offers a better or unique way of accessing information. An agency might let citizen check status of a claim or obtain recommendations for services from a chatbot, but it probably should not let the citizen open a new account from that interface since the chatbot may not be secure enough for that type of interaction. It is important to identify the context in which users may prefer a chatbot to another solution and focus on optimizing the chatbot for that situation.

4) What type of chatbot, kinds of content, and forms of media will be best to communicate to the target audience?

A chatbot is an interactive conversational platform. It is important to evaluate the type of content and forms of media that will be used to resonate with the target audience and encourage meaningful two-way interactions (equal parts user and chatbot). The interaction should not lead users through a lengthy conversation without an appropriate end-point. The designers should script ways for the chatbot to drive the conversation back on track if it deviates from the original function as well as realize the right time to hand the user off to another agent or human for those instances where the chatbot will not have the answer. Be sure the responses provided by the chatbot are clear, accurate, and helpful. Consider what type and category of chatbot will best align with users' needs relative the level of interaction and sophistication. If the chatbot engagement with the user is dependent on a complex communications dialog and memorizing sequences, then you need an AI powered chatbot. These use NLP to decipher human speech and language in voice or text, through pattern recognition. With machine learning (ML), the chatbot can learn over time and get smarter at conversations.

5) How sophisticated is the chatbot?

Stateless chatbots are the simplest type of bot. The bot is a retrieval-based machine model that uses a repository of predefined responses, often with a heuristic algorithm to pick an appropriate response based on the input and context. It does not maintain a state of the conversation and operates in a closed domain (short-text conversations, possible inputs and outputs are limited to fulfill a specific tasks/topic). Its users engage in very simple conversations, sending text to the bot and it will process and then send a reply based on a very limited set of questions on a specific topic. These are the least complex and easiest to build. The stateless bot approach is a great first step for a chatbot because it is contained, may not require the complexity of smart machines and can deliver both business and user value.

Semi-stateful and stateful chatbots are more sophisticated by integrating smart machines using neural networks (deep learning architectures), machine translation techniques, and NLP, enabling more advanced capabilities such as:

- allowing the chatbots to keep the entire history of each conversation they had with the user;
- recall what the user asked for before and can adjust their response based on that;
- self-learning and discovery of certain user patterns which can reveal deeper intentions or preferences of the users;
- long conversational threads with multiple turns, answering various questions and tracking previous conversations.

Stateful chatbots often use generative-based models which do not rely on predefined responses, they can generate new responses using machine translation techniques (translating from one input to an output response).

Stateful chatbot conversations can operate in both closed and open domains. Open domains are not limited to specific tasks/topics as in closed domain. The user can take the conversation anywhere. Questions are asked and the stateful chatbot generates responses to both the common questions and some unforeseen cases for which there are no predefined responses. Stateful chatbot can handle longer conversations and appear to be more human-like. However, generative-based models will increase the chatbot design and learning complexity significantly and require huge amounts of training data.

Note: Generative open-domain systems are true Artificial General Intelligence (AGI) systems because they need to be able to handle all possible scenarios. We are far away from developing these types of systems (but a lot of research is going on in that area). This leaves agencies with providing solutions in restricted domains where both generative and retrieval-based methods are appropriate. The longer the conversations and the more important the context, the more complex and difficult the problem becomes.

6) Do you have the necessary back-end services, compute, and infrastructure resources?

If there is no existing application programing interface (API) that powers existing interfaces, then adding a chatbot will require more technical and administrative work, as well as complexity in the technical environments. Look at existing APIs to see what can be folded into the system, or work to build from the ground up. Developing a chatbot without a strong

back-end system is a recipe for failure. No matter what time of the day it is or how many people are contacting your website or mobile app, every single one of them need to be answered instantly. This can have enormous implications for the enterprise computing platforms and infrastructure resources. It is important to understand the compute capacity and availability required for AI-enabled solutions and intelligent ecosystems as well as to handle unpredictable changes in user demand and to optimally balance demand across computing platforms, storage, and network bandwidth. In addition to increased demand on enterprise data center due to the evolution of traditional web and mobile applications from information-sharing portals to advanced conversational interfaces and complex digital communications channels that are linked to intelligent ecosystems and integrating into legacy enterprise platforms and/or third-party cloud services.

Agencies will need AI management systems designed to monitor intelligent ecosystems, balance IT assets and to optimize their performance. These systems will track a wide range of conversation metrics and other data points and employ advanced analytics and machine control to tweak both the chatbot software and the back-office systems it engages with. This enables continuous improvement for both user and employee/partner interactions but requires a reliable robust high-density virtualized compute and infrastructure architectures.

When Should I Use Robotic Process Automation (RPA)

RPA is typically a tool used for automating manual, time-consuming, complex, rule-based workflows, and functions for back-end administrative IT work. RPAs also provide organizations with the ability to reduce staffing costs and human error. RPA robots are logic and algorithm driven and they execute pre-programmed rules on structured data and mimic humans in their logical processing and decision making. However, RPA robots lack complex, nonlinear process support, can replicate errors hundreds, if not thousands of times due to their reliance on programming scripts, can come with hidden costs, and have been known to negatively impact employee morale because they are meant to replace or reduce human labor. Chatbots, on the other hand, facilitate intelligent dialogue between people, systems, and things, which creates a back-and-forth exchange of information initiated by a human, rather than pre-programmed rules. Chatbots can live in virtually any channel, from websites and mobile apps, to emails, messaging platforms, collaboration tools, and more. This conversational UI not only widens the use case and task applications for chatbots far beyond that of RPAs, but it helps simplify complex interfaces for users and makes digital interactions more human and positive for employees and users.

RPA is an application of technology, governed by business logic and structured inputs, aimed at automating business processes. Using RPA tools, a company can configure software, or a robot, to capture and interpret applications for processing a transaction, manipulating data, triggering responses and communicating with other digital systems. RPA scenarios range from generating an automatic response to an email to deploying thousands of bots, each programmed to automate jobs in mainstream ERP, CRM, and HR information systems.

The RPA target audiences are not necessarily external users. The users can be internal (staff) that employs the use of RPA to perform office tasks (business or IT processes). For example, an agency's licensing/permitting process. The user is not aware of the internal processes, but the speed of the processing may positively impact the user experience. If the RPA was trained to reduce the manual process by automating 70–75% of the process, the process becomes faster and user experience is impacted positively. To the users, the agency is improving the user experience by reducing wait time. Furthermore, external and internal users don not chat with

RPA (or RPA-Bot); the instructions to perform a set of tasks is given as a rule-based instruction.

Flexibility can be a constraint for the RPA-bot operating in environments that are frequently changing. Problems can arise in business/IT environments and platforms on which RPA-bots interact often change. Moreover, a new regulation requiring minor changes to an application form or workflow can delay work in the back office on an RPA initiative. Before an agency sets off to build the RPA they should consider these elements:

1) Set and manage expectations

Quick wins are possible with RPA, but propelling RPA to run at scale across all the lines of business/program areas without an enterprise automation strategy and roadmap is not prudent and can result in poor expectations. In addition, claims about RPA from vendors and implementation consultants can be overhyped. It is important for business and IT leaders to have a clear vision of the automation solution and its capabilities and proceed with a cautiously. Trust but verify when comes to AI automate initiatives and be willing to pull the plug if the results are not meeting expectations.

2) Understand the why and business impacts of the process you plan to automate

RPA is often propped up as a mechanism to bolster return on investment or reduce costs. But the primary focus should be use RPA to improve user experience. Agencies employ many user service agents, but citizens are still waiting in the queue to have their call fielded. A chatbot could help alleviate some wait time by handling the routine inquiries that do not require human intervention. By automating manual backend processes via RPA-bot solution, the process becomes faster and citizen experience is impacted positively and the user service agents can be repurposed to support other business program needs. To the citizen, the agency is improving user experience by reducing the wait time to field the call. Business leaders need to look for automation opportunities and fully understand how bots will be transformative for their operations as well as the ethical impacts on the workforce and improving outcomes for its citizens.

3) Business and IT need to be aligned early and often

When automating manual, time-consuming, complex workflows and functions for backend administrative work for business units, it is important for the business to engage with IT early on due to the inherent complexities associated with integration and support of complex and diverse IT environments.

4) Poor design and change management processes can be devastating

Many implementations fail because design and change are poorly managed. In the rush to get something deployed, organizations may overlook communication exchanges between chatbots, which can break a business process. Before implementation, think about the operating model design. It is important to map out how the expectations of the various bots to work together. Consider the changes any new operations will have on an organization's business processes, workflows and systems that the RPA-bots interact with well in advance to avoid business disruption and costly rework.

5) Command and control are paramount

Problems arise in business/IT environments and platforms on which RPA-bots interact often change. A new regulation requiring minor changes to an application form or workflow could

delay the work on an RPA initiative. Another problem is the lack of proper governance and oversight, which can lead to failures and significant disruptions in business operations and lack of formal reviews, insights to changing business and IT ecosystems impacts on new automation initiatives and existing solution powered by AI.

Agencies must constantly check for chokepoints where their RPA solution can bog down, or at least, install a monitoring and alert system to watch for hiccups impacting performance. It is important for agencies have the proper governance and controls in place to manage all the of the bots and understand the potential impacts due to changes in the business and/or IT ecosystems. AI solutions cannot be built and set free and unmonitored; agencies need command and control to manage the solutions portfolio.

6) Impacts on backend services, compute, and infrastructure resources

Deploying AI solutions to automate manual data entry or to monitor software operations can generate large quantities of data. This can have implications for the enterprise computing platforms and infrastructure resources. It is important to understand the compute capacity required for AI-enabled solutions and intelligent ecosystems as well as how to handle unpredictable changes in user demand and to optimally balance demand across computing platforms, storage, and network bandwidth. Agencies will need AI management systems designed to monitor intelligent ecosystems, balance IT resources, and to optimize their performance.

When should I use a Machine Learning?

Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. Machines operate based on statistical algorithms that iteratively learn from data that allows computers to find insightful information without being programmed where to look for a piece of information. Machine learning does not change the code, but it might change its execution path and decision based on previous data or new gathered data.

Machine learning solutions have been applied in web search results, real-time ads on web pages, email spam filtering, network intrusion detection, and pattern and image recognition. It is important to know when and how to use a machine learning algorithm. Machine learning initiatives should be considered, not only as strategic initiatives, but for their possible effect on other business strategies and operations. Deployment can carry business risk, so investment decisions should be approached with care.

Machine Learning can be used to:

- Discover patterns and trends from increasingly large and diverse datasets as well as enables them to automate analysis that has traditionally been done by humans, to learn from business-related interactions and deliver evidence-based responses;
- Discovery of relevant features in otherwise disordered datasets;
- Identify events: use the machine learning process to understand how various objects might add up to an event;
- Creation of intelligent chatbots (i.e., semi-stateful and stateful) by integrating smart machines using neural networks (deep learning architectures), machine translation techniques, and natural language processing enabling more advanced capabilities.

Machine learning systems are made up of four major parts, which are:

- **Model:** The system that makes predictions or identifications. Algorithm selection for the model is critical to both the type of problem to be solved and forming accurate outcomes.
- **Parameters:** The signals or factors used by the model to form its decisions and the hyperparameters are settings used to tune the algorithm used in the model.
- **Learner:** The system that adjusts the parameters, and in turn the model, by looking at differences in predictions versus actual outcome.
- **Data:** Input data is required to train and test/validate the model and produce actionable results/outputs based on the parameters of the model consisting of statistical based rules. Data sources should be scrubbed, cleaned, and integrity examined, then split into training and test datasets. In cases when the dataset is too small or a portion of the training/test partition of the data is not appropriate, an additional cross-validation data set is required to properly determine the level of accuracy of the model.

Machine learning solutions are heavily dependent on programming, statistical modeling using algorithms, and input data. Data is fed into the machine model (an algorithm), the hyperparameters (settings) are configured and adjusted, the machine proceeds to learn, analyze and decipher patterns found in the data through the process of trial and error. Once the machine has been properly trained (via supervised, unsupervised, or reinforced learning techniques), it can then independently apply its training in making decisions based on past experiences.

How much data do I need?

In general, machine learning works best when the training dataset includes a full range of feature combinations.

For example, if you have dataset about SNAP Eligibility categorized by the following features:

- Salary (X)
- Dependents (X)
- Employment status (X)
- Age (X)
- Eligibility Status (Y)

To properly assess the relationship that the first four features (X) have on the SNAP eligibility (Y), you need to have a dataset that includes the y value for each combination of features. We would need to know the SNAP eligibility status for citizens that are employed with a specific salary that do not have dependents and those that do have dependents. The more available combinations, the more effective the model will be at capturing how each attribute affect Y (eligibility status). At a minimum, the dataset should have ten times as many data points as the total number of features. For the example above with four features, a small training dataset should have at least forty rows of relevant data that covers all combinations. In some cases, it may not be possible to or cost effective to obtain source data for every possible combination. In these cases, you will need to make do with the data you have available but understand the potential limitations on training and overall accuracy of the model.

How can I Evaluate the Accuracy of the Model?

Machine algorithms are complex and requires expertise and practical experience in determining and implementing the best machine learning algorithms to solve the problem. Algorithm

selection for the model is critical to both the type of problem to be solved and forming accurate outcomes. Each algorithm can produce vastly different models and can lead to dramatically different results and degrees of accuracy based on the hyperparameters provided and how they are configured. A constant challenge in machine learning is the balancing to underfitting and overfitting of the model (describes how closely the model follows the actual patterns of the dataset). Both underfitting and overfitting are due errors related to bias and variance.

Bias refers to the gap between the predicted value and the actual value. Variance refers to how scattered the predicted values are. This is best illustrated in Figure 1. The center bulls-eye of the target perfectly predicts the results from the machine model and each dot on the targets represents a specific result from the machine model based on the training dataset.

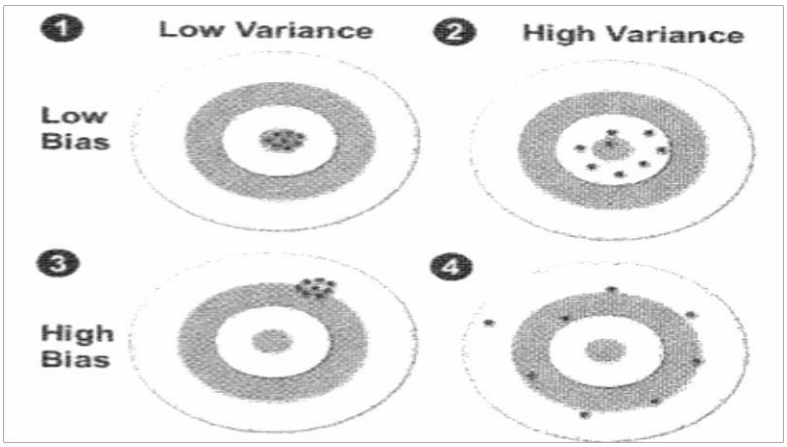

Figure 1. Source: (Machine Learning; Oliver Theobald, Page 87)

Target 1: Has low bias and low variance: Bias is low because the dots are closely aligned to the center bulls-eye and low variance because the dots are densely populated in one location.

Target 2: Has low bias and high variance: Bias relatively low because the dots are near to the center bulls-eye and high variance due to the dots are spread out from each other.

Target 3: Has high bias and low variance: Bias high because the dots are closely aligned far from the center bulls-eye and low variance because the dots are densely populated in one location.

Target 4: Has high bias and high variance: Bias high because the dots are not closely aligned and scattered from the center bulls-eye and low variance because the dots are spread out from each other.

The more the dots deviate from the center, the higher the bias and the less accurate the model will be overall. The more densely positioned the dots are helps determine the degree of variance the results are to the actual data. Ideally, the model should have low bias and low variance.

However, it is not possible to achieve an ideal model. It is the ***prediction error*** you want to minimize, which requires analyzing and determining the acceptable trade-off between the optimal bias and variance of the model. AI engineers and programmers can plot and compare the training and cross-validation data (or test data) results using a model complexity graph to determine the trade-off point where the model has the lowest prediction error. The graph will also illustrate the degree of ***underfitting*** (due to High Bias and Low Variance) and ***overfitting*** (due to Low Bias and High Variance) as shown in Figure 2.

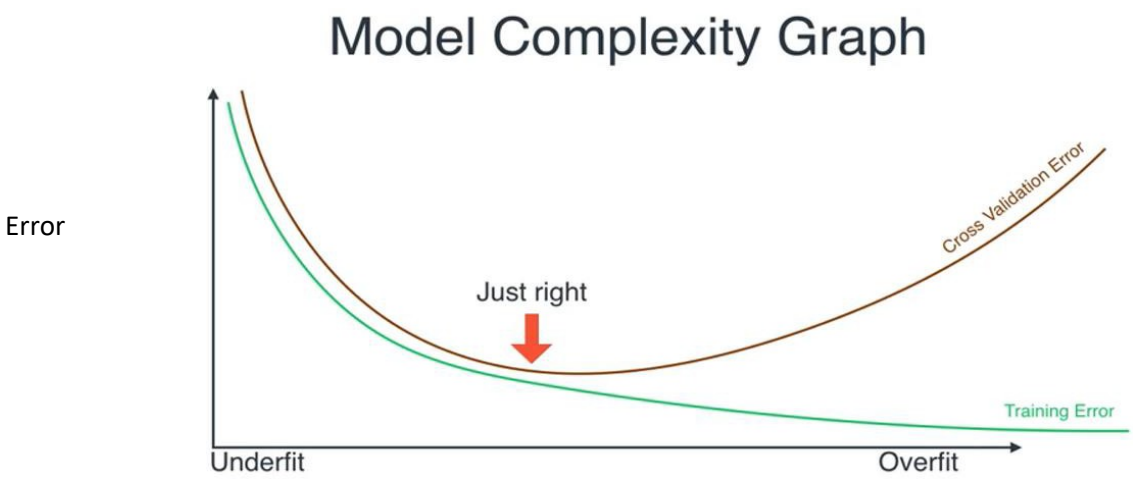

Figure 2. Source: (Gartner)

Risk of underfitting and overfitting of a machine model (due to bias and variance errors) are highly dependent on dataset size, completeness, integrity, quality of data used in training and testing machine models. It is important to use multiple randomize training, testing, and validating datasets to select and verify the best model the best model that produces the best results (Figure 3).

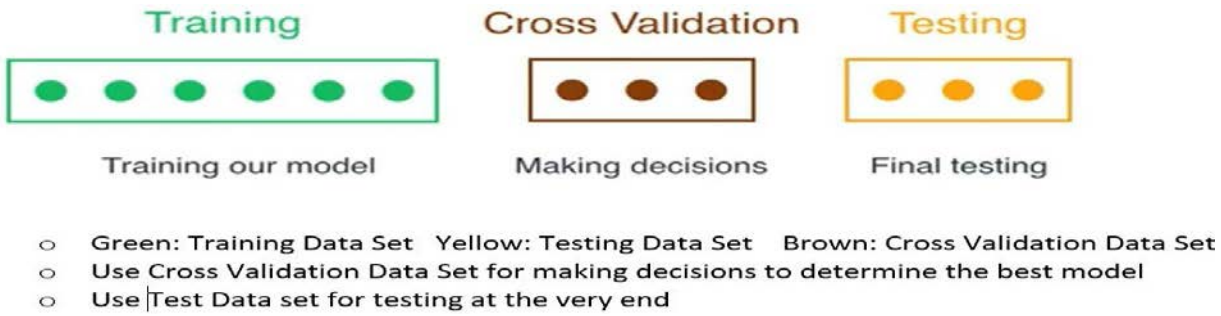

Figure 3. Source: (Gartner)

Establish machine learning model certification criteria and examine and validate results using a validation metric (accuracy, recall, and precision). Minimize errors due to overfitting or underfitting. Determine the platform to execute the ML models and associated algorithms.

It is important to note that accuracy is not the best or only metric to use when determining the best or verifying the machine model.

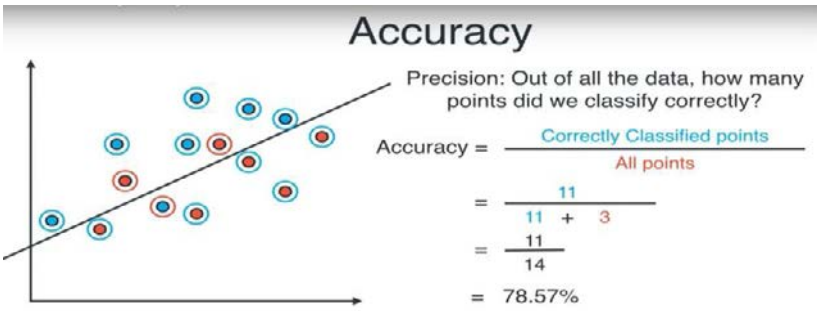

Figure 4. Source: (Gartner)

Recall and precision metrics are more reliable measurements to use to when determining the best or verifying the machine model (Figures 5 & 6).

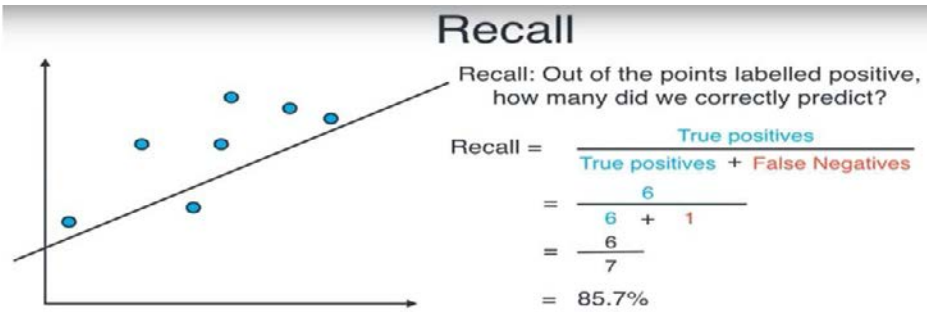

Figure 5. Source: (Gartner)

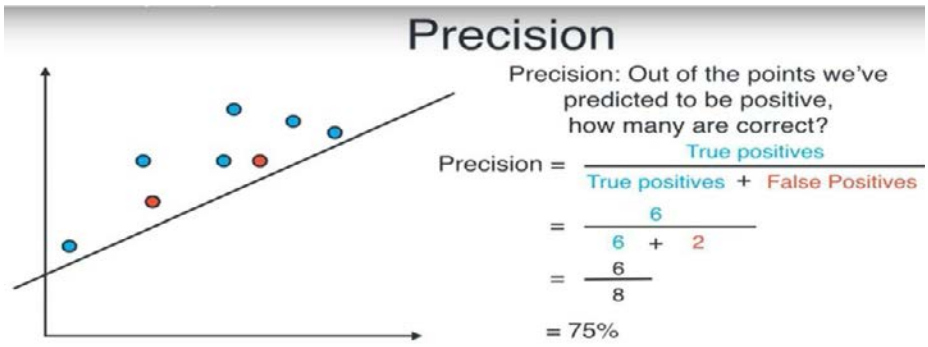

Figure 6. Source: (Gartner)

Use harmonic mean with the values of the recall and precision metrics to calculate the F1 Score of a model {FI Score = Harmonic Mean (Precision, Recall)}. (Figure 7).

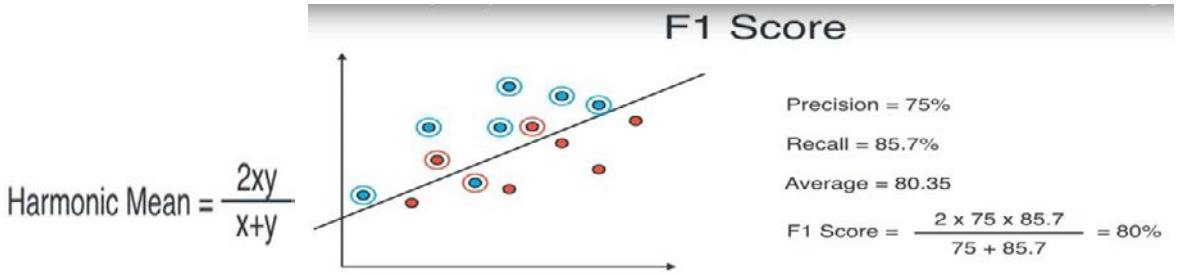

Figure 7. Source: (Gartner)

Machine Model Selection Examples:

Figure 8. Source: (Gartner)

Use the training datasets to train all the models and use cross validation data to calculate the F1 score. As the final step, use test data set to verify model. Parameters are the coefficients of the polynomials. The hyperparameters is the degree of the polynomial. The model with the highest F1 score is best.

Figure 9. Source: (Gartner)

Use the training datasets to train all the models and use cross validation data to calculate the F1 score (choose the model with the highest F1-Score). As the final step, use test data set to verify model. The hyperparameters is the depth of decision tree (number of trees). This decision tree model only has one hyperparameter = Depth. The model with the highest F1 score is best.

Precautions

1) Because machine learning is time and data-intensive, a critical assessment of the applicability of existing analytical models / approaches or alternative solutions is also appropriate. This ensures that the potential value is relative to the effort.

2) Machine learning works especially well for applications where applicable associations or rules are intuitively captured but cannot be easily described by logical rules. When accuracy is more important than interpretability, or when the data is problematic or too complex for traditional analytic techniques.

3) Machine learning in practice requires the application of the scientific methods, human centered design, and effective communication skills by humans. Successful organization have the analytical infrastructure, know-how and close collaboration between

technologists, analysts, and business professionals to translate these synergies into meaningful business solutions and outcomes.

4) It is imperative to have dynamic team model and iterative processes that provides maximum flexibility and agility allows for faster evaluation of progress and to determine whether an alternative approach is needed to best solve the business problem.

5) Successful outcomes are highly dependent on the ability of Business Relationship Managers and AI architects/engineers, who can understand the business problem and/or challenges in the context of the business and IT ecosystems as well as translate between the wants, the mathematicians/statisticians and the managers. If there is no link, then misunderstandings and misinterpretations are highly plausible, and the danger of failure is increased.

6) Machine learning often challenges traditional approaches to quality assurance and risk management. At some point, the training or test data must be replaced by productive data. Hence, true validation only results against new data.

7) Important to establish criteria for what is “good enough” to understanding how models must be validated and developed aligned with quality assurance and risk management frameworks.

8) Dynamic team model is comprised of various experts with business, data and technical expertise. This includes data scientists and other experts who can assess the required data and bring it on board. Business experts who can validate integrity, explain the context, and assess implications (business, social, ethical, potential bias) of AI solutions. It also requires IT staff capable of deploying and maintaining the technical ecosystems.

9) Ensure AI initiatives follow existing governance frameworks, internal controls, policies, and regulations.

10) The introduction of non-traditional (large) data sources, including unstructured text, speech, or images may also require new data management capabilities.

11) Maintaining the model is a critical, ongoing process that must be carried out in the same way as the initial model development.

12) Ensure mechanisms are in place to examine and remedy combinational explosion computing problems, resulting from the number of combinations (affected by inputs, constraints, and bounds of the problem) that one must examine grows exponentially, so fast that even the fastest computers will require an intolerable amount of time to examine them.

13) AI solutions require iterative modeling and tuning pre and post deployment. As with any good experiment, some hypotheses will initially turn out to be wrong. Model parameters may need to be fine-tuned, new data may need to be procured or generated, or the problem description rewritten based on what is found. As a result, decision makers and team members alike need to apply a machine learning test-and-learn mentality to

establish successful data analysis and determine the best model to use.

14) 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 should be vetted through the appropriate governing processes.

15) Training and testing the model using the same data set, results in an overly trained model, with high accuracy (% of correct predictions using the training data set) and low out of sample accuracy (overfit), unable to make correct predictions on unknown data sets.

16) AI systems are required to comply with existing security policies regarding the protection of commonwealth data assets.

17) Time required in researching, selecting, constructing, distilling, testing/validating, summarizing, and documenting algorithms before they can be implemented, can be significant investment in time and resources.

18) Algorithms are complex and requires expertise and practical experience in determining and implementing the best machine learning algorithms to solve the problem.

19) Be sure to validate the bot before testing with end-users. It is important to start seeing if statements/declarations are matching the intended task correctly. Validate intent recognition for all pre-programmed synonyms and phrases that way, when users talk to the chatbot, any failed statements can be quickly resolved with machine learning.

20) Deploying bots to automate manual data entry or to monitor software operations can generate a ton of data. This can lure business into an unfortunate scenario where they are looking to leverage the data using ML on the data their bots generate, and then throw another bot on the front to enable users to easily query the data. Suddenly, the RPA project has become an ML project that has not been properly scoped as an ML project. Agencies should be careful not to fall down the data rabbit hole and effectively manage the scope of bot automation projects to ensure they do not evolve into something unwieldy.

21) When generating responses, the bot should ideally produce consistent answers to semantically identical inputs. Many systems learn to generate linguistic plausible responses, but they are not trained to generate semantically consistent ones. Usually that is because they are trained on a lot of data from multiple different users.

22) Generative open-domain systems are true Artificial General Intelligence (AGI) systems because they need to be able to handle all possible scenarios. We are far away from that as well (but a lot of research is going on in that area). This leaves us with developing solutions in restricted domains where both generative and retrieval-based methods are appropriate. The longer the conversations and the more important the context, the more complex and difficult the problem becomes.

Considerations for Selecting the AI Vendor

1) First define what exactly is being procured. This will depend on the problem or opportunity, solution alternatives, and technical approach that will be used (if known). If the problem or opportunity is clearly understood and the goals and objectives defined but solution alternatives or technical approach is not known; be sure the first phase of the engagement outlined in your Statement of Work (SOW) requires the vendor to evaluate and propose viable solution alternatives (pros, cons, alignment/integration with business and technical ecosystems, risks, costs, level of effort, resources, timeframes, conceptual illustrations, sustainability, etc.) for agency and/or governing entity review and approval before proceeding to the next phase. If the vendor and/or agency are not confident or in agreement, then Phase 2 should require a structured proof of concept or pilot to assist in choosing the best solution alternatives and/or technical approach. Alternatively, depending on the scope, visibility, and potential risks of the initiative; it may be prudent to require that the vendor recommended solution and/or technical approach be initially validated via a proof of concept or pilot prior to jumping all in.

2) Because of hype in the market, vendors will exaggerate their AI credentials and capabilities. Vendors will boast their strong partnerships with AI technology providers in this space, which can range from a loose connection to a long-term strategic alliance. Some vendors will have only minimal AI capability and experience as a part of their product or services offerings. Some vendors may have different methodologies/approaches to AI solution development, testing, validation, and/or dealing with common challenges relative to data knowledge and anomalies, AI machine model overfitting, underfitting due to bias or variance errors in the data sets. It is important to obtain some proof of their capabilities and experiences regarding AI engagements of similar scope and complexity. Examining vendor methodologies/approaches, references, technology partnerships with providers, and requiring the vendor to conduct proof of concept or pilots as outlined in step 1 above. A vigilant, trust but verify approach regardless of the vendor's professed reputation in the industry.

3) Agencies need to have proof of value. The agency should have some business plan outlined (goals and objectives linked to CSFs and KPIs) to be able to understand how to determine success and overall value of the solution. The vendor should have a model and mechanism in which they can validate the agency's business case and desired outcomes for the solution implemented. If the vendor does not have a model or cannot articulate how they will ensure alignment to the business case and measure outcomes, they should not be selected.

4) Understand the agency's capabilities internally and required dependencies on third party service providers (existing and new) for solution support and long-term sustainability in your business and technical ecosystems. Ensure there is proper documentation, and transition of ownership, and knowledge transfer of the AI solution with appropriate internal stakeholders.

5) Pricing models in this space can vary significantly. It is important for agencies to have a good understanding of engagement types (consultants, engineers, SaaS, PaaS, software, platforms, libraries, and other services) and their pricing models. Recommend the use of Gartner or other industry research consulting firms to improve understanding and what is reasonable for the AI technology product and services this space. It is important to

establish enterprise architectural standards and technical specifications for AI products and services used in the commonwealth agencies.

6) Agencies should work closely with their legal counsel and DGS procurement to protect the commonwealth's best interest due to the special nature and risks associated with AI engagements and agreements.

   Note: Standard contracts for AI solutions can result in leaving user's critical intellectual property and data ownership at risk. Secondly, AI products software licensing and AI SaaS service providers user agreements need to be closely examined regarding pricing models, end-user consents, IP ownership, and indirect access rules, as well as language pertaining to transparency, accountability, and "explainability" for algorithms and machine learning solutions.