DRAFT--ALGORITHMIC TRANSPARENCY AND VENDOR ACCOUNTABLITY

ADS/AI Assessment and Audit Best Practices

ABSTRACT

USAGM OCIO IT Risk Management Division reviewed Automated Decision Systems (ADS) and/or Artificial Intelligence (AI) assessment/audit toolkits that were designed to inform the user to first assess and then manage the algorithm risk. These artifacts delve deeply into potential harms and human rights/privacy considerations than most of the other tools. The main goal of this review is to elicit internal agency conversation, raising questions and offer recommendations for considerations to mature the organizations approach to their adoption and integration.

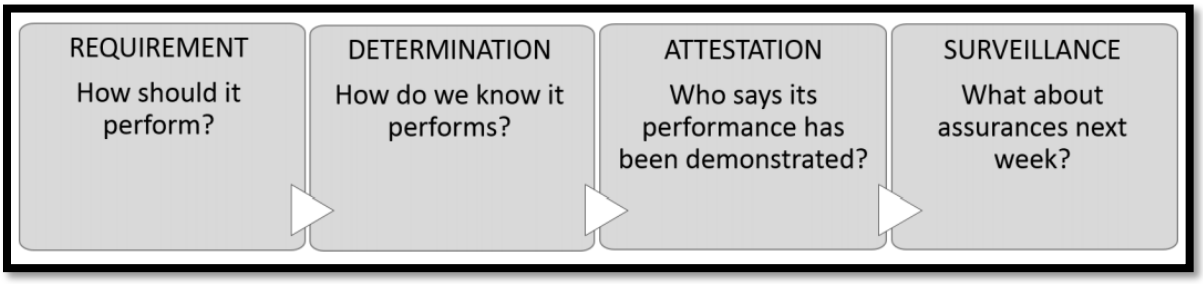
Ashley Moore

USAGM/OCIO AI In Broadcasting and Journalism Program

Contents

Background—Federal Agency Use of Conformity Assessments/Audits Processes

Under the National Technology Transfer and Advancement Act (NTTAA), *Conformity Assessment* is the demonstration that specified requirements relating to a product, process, system, person or body are fulfilled. In addition, conformity assessment is used in Federal programs to provide confidence that requirements in legislation, regulation, policy, and procurement are met. Conformity assessment can include; *supplier's declaration of conformity, sampling and testing, inspection, certification, management system assessment and registration, the accreditation of the competence of those activities, and recognition of an accreditation program's capability*.



[1]

Figure 1 A concept for federal agency conformity assessment programs

There are a number of factors that affect a Federal agency conformity assessment program, including the agency's mission; regulations underpinning the need for the conformity assessment program; the dynamics of specific markets and sectors in the U.S. and internationally; and the current state of conformity assessment programs within the area of interest. The Office of Management and Budget (OMB) Revised Circular A-119 (2016) Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities[2] (the Circular) establishes policies on Federal use and development of voluntary consensus standards and conformity assessment in regulatory, procurement, and program activities. The Circular provides factors for agencies to consider when assessing the effectiveness of conformity assessment options and determining the type(s) of conformity assessment activities to employ.

There are 3 general effective considerations used when designing, developing, and operating a conformity assessment program.

1) **Engage Stakeholders**: Engage stakeholder groups during each phase of program design, development, and operations. Stakeholders bring expertise, knowledge, and perspective that help inform agency decisions. To the extent feasible, a collaborative approach with the private sector (and other agencies with like requirements) often results in agency decisions and policies that are better understood and accepted, and overall program success. Experience and expertise in Federal agencies exist through the Interagency Committee on Standards Policy (ICSP), the ICSP

[1] NIST Special Publication (SP) 2000-01 ABC's of Conformity Assessment
[2] https://www.nist.gov/standardsgov/what-we-do/federal-policy-standards/key-federal-directives

Conformity Assessment Working Group (CAWG), the Trade Policy Staff Committee Subcommittee on TBT (TPSC), and each respective Agency Standards Executive.

2) **Maximize Transparency**: Transparency in design, development, and operations fosters greater acceptance and understanding of agency decisions and contributes to the conformity assessment program's success. The program development process should be as transparent as possible with an outreach strategy that seeks to engage with stakeholder groups through activities such as workshops, requests for information, blog posts, and other social media opportunities. In ongoing programs, access to policies, procedures, updates, plans, etc. allows stakeholders to adapt their own processes and activities.
3) **Leverage Existing Efforts**: *Leverage, if possible, existing conformity assessment programs, activities, results, and other output to reduce regulatory burden and stakeholder costs. Use of public- and private-sector conformity assessment programs already operating.*

Federal agency conformity assessment programs have four elements:

1) **Objectives and Goals** – A conformity assessment program exists to meet agency objectives as part of its mission, or as defined in regulation, procurement, or other programmatic needs.
2) **Conformity Assessment Scheme and Oversight**– A conformity assessment scheme consists of the program's requirements, activities, roles, and interactions of participants. The oversight function ensures the integrity, consistency, and correctness of the scheme implementation.
3) **Requirements for the Object of Conformity**– Requirements are the attributes or characteristics exhibited by the object of conformity (i.e., the product, process, system, person or body) necessary for meeting agency objectives. Requirements are often in the form of voluntary consensus standards.
4) **Program Management** – Program management policies and improvement strategies support a conformity assessment program that is effective and efficient while being adaptive to changes in markets, technology, regulation and policy.

Federal Conformity and its Application to Agency ADS/AI Integration:

Despite the importance of the uses and decisions as described above, government agencies frequently procure, develop, and implement algorithmic systems with minimal to no transparency, public notice, community input, oversight, or accountability measures. Procurement officers and agency staff often lack technical expertise to evaluate algorithmic systems, their capabilities, and potential consequences.

This creates a knowledge imbalance in contracting, particularly because many algorithmic systems vendors almost exclusively sell to government agencies. Consequently, vendors can oversell the utility and value of a system or offer the system at reduced costs, which is difficult for resource constrained agencies to turn down.

Algorithms are fallible human creations, so they are embedded with errors and bias like human processes. When algorithmic tools are adopted by government agencies without adequate transparency, accountability, and oversight, their use can threaten civil liberties and exacerbate existing issues within government agencies (e.g. bias, inefficiencies, opacity regarding decision making). We know that federal, state and local governments are increasingly implementing algorithmic systems in their daily practices, but we still do not know how widespread and integrated such algorithmic systems are used at any level of government.

With the above in mind, USAGM OCIO strides to be inline and use federal guidance and Public/Private Sector best practices when developing and using conformity and audit practices associated with the development, use, and procurement of Automated Decision Systems (ADS) also described as Artificial Intelligence (AI) an umbrella term for; *deep learning, machine learning, image recognition, natural language processing, cognitive computing, intelligence amplification, cognitive augmentation, machine augmented intelligence, and augmented intelligence*.

Algorithms, Automated Decision Systems and Artificial Intelligence Defined

- **What are Algorithms?**

An *Algorithm* is generally regarded as the mathematical logic behind any type of system that performs tasks or makes decisions. For example, how Facebook sorts what posts a user sees in their Facebook feed is an "algorithm." The logic used in a software program to assign criminal defendants a public safety risk score is also an "algorithm." "Algorithms" do not have to be based in software on computers. However, in the case of many types of risk assessments used in courts or human services agencies, the "algorithm" can be represented by a piece of paper that outlines the steps a human should take to evaluate a particular case.

- **What are Automated Decision Systems (ADS)?**

An Automated Decision[-making/-support] System is a system that uses automated reasoning to aid or replace a decision-making process that would otherwise be performed by humans. Oftentimes an automated decision system refers to a particular piece of software: an example would be a computer program that takes as its input the school choice preferences of students and outputs school placements. All automated decision systems are designed by humans and involve some degree of human involvement in their operation. Humans are ultimately responsible for how a system receives its inputs (e.g. who collects the data that feeds into a system), how the system is used, and how a system's outputs are interpreted and acted on. When talking about automated systems used in government, you might hear people refer to "algorithms," "automated decision systems," or "algorithmic systems" loosely and interchangeably. "Automated decision system" was the phrase used in New York City for its algorithmic accountability task force, so we stick with that when talking about a complete end-to-end system used in government, from design, testing, and actual use, including the human operators.

- **What exactly does "Artificial Intelligence" (AI) mean?**

Artificial Intelligence (AI) has many definitions and can include a wide range of methods and tools, including machine learning, facial recognition, and natural language processing. But more importantly, AI should be understood as more than just technical approaches. It is also developed out of the dominant social practices of engineers and computer scientists who design the systems, and the industrial infrastructure and companies that run those systems. Thus, a more complete definition of AI includes technical approaches, social practices and industrial power.

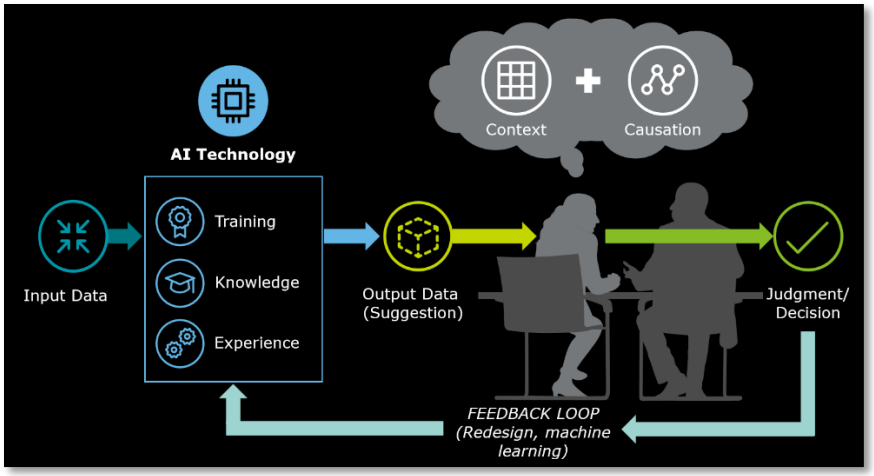


Figure 2 Human-In-The Loop

Executive Summary

There are many terms related to Automated Decision Systems and Artificial Intelligence (ADS/AI) capabilities; such as they include but not limited to the following: deep learning, machine learning, image recognition, natural language processing, cognitive computing, intelligence amplification, cognitive augmentation, machine augmented intelligence, and augmented intelligence. The USAGM OCIC AI in Broadcasting and Journalism program solicited input from numerous public and private sector entities that have specialized in the areas of ADS/AI auditing and assessing the procurement and integration of these algorithmic capabilities. The main goal of this review is to elicit internal agency conversation, raising questions and offer recommendations for considerations to mature the organizations approach to their adoption and integration.

As part of this body of work, the review uncovered that there are a few noteworthy tools that we couldn't access, because they are available only to customers or employees of a company. This includes *Microsoft, Fairness Flow from Facebook*, like *Google's What-If Tool*, which helps Facebook employees assess if their datasets and predictions are fair. Another is the *Fairness Tool from Accenture* which allows its staff and clients to determine how their prediction algorithms may be relying on sensitive variables like race and gender, along with correlated variables such as location, occupation, or others. Another toolkit is *Weights and Biases* which focuses on bias and transparency, though a key focus of the tool is hyperparameter tuning and experiment tracking. This is a tool which is free to use by an individual but has a team or company fee.

Review/Assessment Criteria:

The tools available to support ADS/AI assessments should be designed to ensure that the development, deployment and use of AI systems meets the seven key requirements for Trustworthy AI:

1) Human agency and oversight,
2) Technical robustness and safety,
3) Privacy and data governance,
4) Transparency,
5) Diversity, non-discrimination and fairness,

6) Environmental and societal well-being and
7) Accountability.

In addition, these ADS/AI assessment and auditing tools should consider technical and non-technical methods to ensure the implementation of those requirements.

Though this report primarily focuses on multi-functional toolkits, their design, feasibility of use and inclusion of the above criteria/principles. In addition, the review search for existing AI Service Level Agreements which include ADS/AI Governance; ADS/AI Impact Assessments, sandbox, human-in-the loop (*human intervention in every decision cycle of the system*) and corrective actions associated with—constructed biases, inaccuracies, errors and mistakes.

When USAGM employees desire to acquire ADS/AI capabilities they should able to spot the low-hanging fruit where ADS/AI could make our organization more efficient—but only if they have some minimum knowledge of ADS/AI. Any member of our organization should be able to answer the following questions:

- **How does it work**? Team members who aren't responsible for building an AI system should nonetheless know how it processes information and answers questions. Understanding data—the fuel of AI—helps people understand what AI is good at.
- **What is it good at?** Machine learning tools excel when they can be trained to solve a problem using vast quantities of reliable data, and to give answers within clear parameters that people have defined for them. Help your employees understand this difference by showing them tools they already use that are powered by AI, either within the organization or outside it.
- **What should it never do?** Just because machine learning can solve a problem does not mean it should. If employees understand the ethical limitations of AI, they can be important guards against its misuse.

In context to USAGM OCIO roles and responsibilities, any internal assessor or auditor must have enough knowledge of key information technology risks and controls and available technology-based audit techniques to perform their assigned work. However, not all internal assessors or auditors are expected to have the expertise of an internal auditor whose primary responsibility is information technology auditing.

Therefore, the internal assessment and/or auditing activity must evaluate risk exposures relating to the organization's governance, operations, and information systems regarding the:

- Achievement of the organization's objectives
- Reliability and integrity of financial and operational information
- Effectiveness and efficiency of operations and programs
- Safeguarding of assets
- Compliance with laws, regulations, policies, procedures, and contracts (to include service level agreements for ADS/AI Cloud services)

ADS/AI Assessment and Audit Toolkits Review

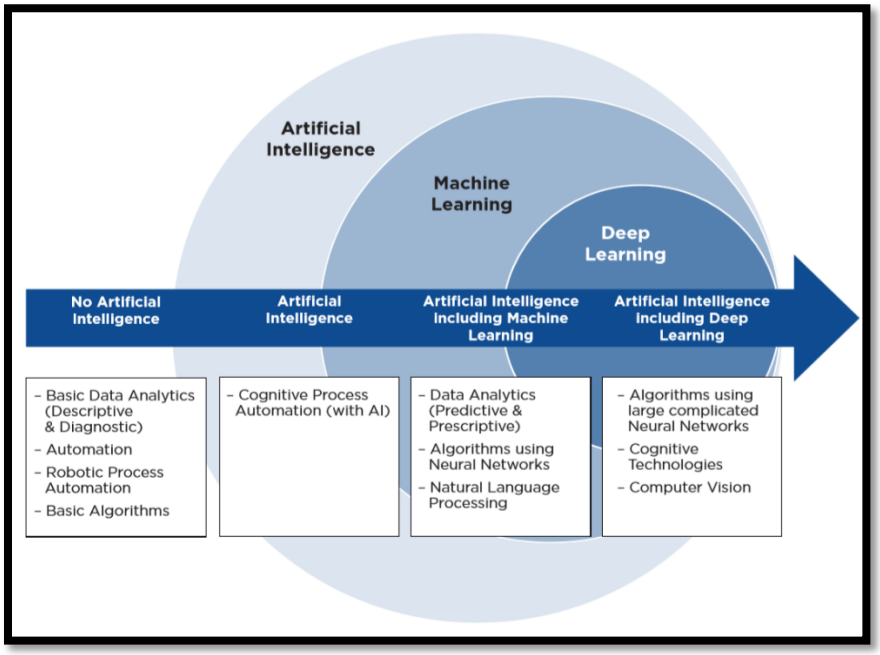


Figure 3 Assessing and Auditing ADS/AI

Deloitte—Artificial Intelligence Auditing

Deloitte's open source material on their approach to assessing ADS/AI seems sound and based on a set of 7 AI Governance Components

1) Technique: Specific technique or combination of techniques that is used to address a specific use case or business problem (e.g., language processing, neural network, image recognition)
2) Data: Data sets (internal or external) used to build and train AI models/algorithms, and their level of curation and fit-for-use (i.e., availability of vectors, weights, results) 3. Policies, Standards & Controls Organizational constructs that establish the design principles and guardrails for the development, deployment, and dispositioning of AI models/algorithms
3) Validation & Testing: Mechanism to review, test, and monitor the development and deployment of AI models/algorithms
4) Data Science Platform & Infrastructure: Operational and technological resources leveraged to build, operate and/or monitor AI models/algorithms
5) Talent & Workforce: Skills and people required to drive and sustain the development, operation, and monitoring of AI
6) Industry & Regulatory Alignment: Awareness and alignment with relevant regulations and/or industry standards related with the use of AI models/algorithms

Deloitte provides insight on “Notoriously Tough” Problems worthy of considering:

	Challenges / Questions	Industry Trends / Approaches
Vendor Black Boxes	• Who is responsible for testing vendor AI? • Do we allow the use of AI-enabled vendor “black boxes”? • How should (can?) we test vendor AI	• Explicit AI Governance expectations written in vendor contracts • Maintain inventory of AI usage by vendors • Periodic testing schedule over sample of vendor AI, prioritized by risk level
Role of Policy & Controls	• What is the right balance between policy and controls? • What existing structures can be leveraged for AI (e.g., MRM)	• AI should be added to existing policies; however, unique control expectations must also be developed for AI • MRM infrastructure may be leveraged, but specific testing procedures often lack necessary sophistication to govern AI
Use of Human and Non-Human Decisioning	• When and how frequent should human review be required? • Should we invest in AI challenger models?	• Use a spectrum to determine appropriate control structures for range of AI use cases • Challenger systems and bias detection monitoring are leading practice
Operating Model	• What is the right operating model (centralized vs. decentralized)? • How do we solve for the skills/knowledge gap in second and third lines of defense	• Centralized governance model, including a central pool of data scientists; performing always-on monitoring from intake to disposition • Institute review and gating processes
New Control Structures	• What new control structures/ controls do we need to manage the new risks driven by AI?	• Considerations of surveillance systems • Incorporating AI-driven monitoring solutions to check on AI models/algorithms

In context to Information Security and lines of defense, USAGM's approach will need to be coordinated effort across all LODs ton include CIO Council on Data Management

Examples of ADS/AI Risk	Lines of Defense	Examples of Governance and Control Activities
• Increased risks of both benign and malicious cyber intrusions / breaches. • Significant risk of disruption to the company's operations from unintended machine-made decisions or actions. • Lack of accountability for outcomes due to inadequate control and responsibility structure • Faulty financial projections or calculations that undermine the integrity of financial planning and reporting. • Competitive disadvantage resulting from bias replication and blind spots due to hidden assumptions and biases in data. • Heightened impact of threats related to safety, trust, and alignment with the ethics and values of the organization. • Compliance violations and reputational damage resulting from poorly designed or monitored AI. • Violating the safety, trust, fairness, or transparency expectations of the organization or its stakeholders	**1st LOD**	• Develop standards for AI development and "kill switch" mechanism • Leverage enterprise sandbox for AI to shape governance and controls • Data Curation for AI (volume, velocity, variety)
	2nd LOD	• Incorporate bias detection and monitoring • Use control networks to monitor/surveil outputs from AI solutions • Make risk management nimble and dynamic to adopt/deploy AI applications with business units
	3rd LOD	• Internal Audit using independent neural networks or comparable techniques to test AI solutions • Adopt Governance by Design philosophy through defined boundaries of transparency and accountability
	CIO Council (DATA MGT)	• Periodic review of the performance measures/scorecard and key decisions associated with AI models/algorithms • Review of design principles and guardrails associated with data sets, techniques, and use cases

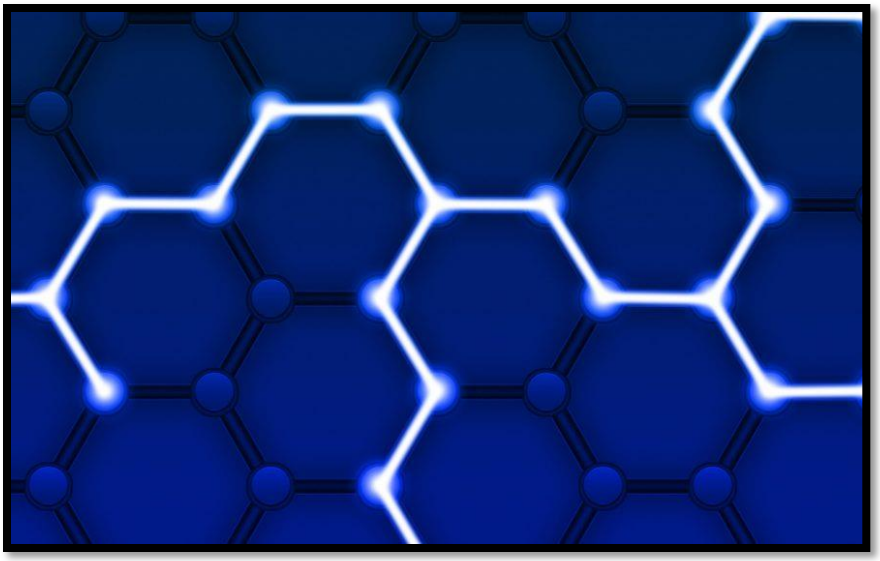


Figure 4 Ethic and Algorithms

John Hopkins COE—Ethics and Algorithms Toolkit

These tools offered by Johns Hopkins University new Center for Government Excellence helps governments build capacity for decision making that is rooted in evidence, transparent accountability, and citizen engagement. The interdisciplinary center is part of the university's 21st Century Cities Initiative, a cross-disciplinary research effort for urban study and change range in their complexity, technicality, and topic. In general, AI ethics includes topics of fairness, accountability, transparency, privacy, human rights, and security, among others. Regardless of topic, modality, or origin, most of the tools include additional links to other tools and papers and highlight that the tool is meant to aid and promote investigation and discussion, not necessarily to be a standalone solution. Below are the various tools in approximate order of technicality reviewed.

There are five documents which comprise the toolkit:

1) **Overview and Introduction:** The overview section of the toolkit is comprised of level-setting background information that will be useful when traversing subsequent sections of the toolkit. We have outlined a few real-life scenarios where the toolkit might be applied, provided definitions, and more. For example, while we briefly touched upon machine learning in the previous module, the toolkit overview helps you understand more about the various types which exist, such as supervised learning, unsupervised learning, and so on.
2) **Part 1: Assess Algorithm Risk:** In Part 1 of the toolkit, there are six major steps (or questions) to help you and your stakeholders characterize an algorithm. Many of these steps have multiple components, but also include clear instructions on how to summarize those stages in order to complete the step.
3) **Worksheet for Part 1:** Since this document can be difficult to navigate, we have developed a **worksheet for Part 1**, designed to help you track your responses to the individual steps and how they are combined into overall risk values. It's worth noting that although answering a series of questions seems simple, you will almost certainly need additional people to help - whether they are stakeholders, data analysts, information technology professionals, or representatives from a

vendor that you are working with. Don't expect to complete this part of the toolkit in just a few hours. Some of the steps will evoke considerable discussion.

4) **Part 2: Manage Algorithm Risk:** Although it's helpful to know how concerned you should be about various aspects of your algorithm, that's only half the battle. Although there may be a few cases where the risks are too severe to proceed, there are often ways to mitigate them. Using Part 2 of the toolkit, you identify specific techniques to help address the considerations you identified in Part 1. The results of Part 2 will be highly customized and specific to the factors you evaluated in part 1. Some of the recommendations can introduce significant burdens that are more appropriately addressed within large-scale programs, such as those that support the social safety net. It is not unusual to need executive and political support to be successful.
5) **Appendices:** Although this isn't specifically required reading in order to use the toolkit, the appendices provide plenty of additional context and depth. The first appendix contains a list of in-depth questions to help you understand your data in more detail. The second provides additional background on bias and how easily it can arise.

Figure 5 ADS/AI Lines of Governance

KPMG ADS/AI Audit Framework

KPMG has developed a risk and control framework which looks at 17 categories for managing risks and controls for AI solutions. We identified 78 risks in total, and 106 controls. The important areas to look at include things like strategy, governance, human resource management, security management, and IT operations.

For instance, how are AI initiatives aligned to enterprise strategy and how is innovation driven? Who in the organization will be responsible for the use of AI and any mistakes it makes? How will you protect against new AI threats, and how will you manage the AI inventory?

Organizations and their business units must approach AI with a focus on specific areas. Do we know what the risks are, the controls we need and how we would audit them? Is the audit function influencing the strategy of 3 lines of defense, and can it clearly articulate its own strategy? AI is no longer a theoretical possibility; it's here. It will continue to evolve, presenting us with great opportunities, but also a whole new set of risks to consider. Now is the time for internal auditors to play a leading role, get fully involved, and help their businesses get it right from the outset.

- Artificial Intelligence Governance - key features: Designs and sets up criteria for building and continuous monitoring and control of AI solutions and their performance, without impeding innovation and flexibility.
- Artificial Intelligence Assessment - key features: Conducts diagnostic reviews of AI solutions, and risk assessments of control environments to determine organizational readiness for effective AI control. Provides methods and tools to evaluate business-critical algorithms, puts testing controls in place, and

oversees design, implementation and operation of AI programs to help address AI's inherent challenges: integrity, explainability, fairness and agility.[3]

The KPMG AI Internal Controls and Risks Management Framework is structured on the following 17 components as illustrated below:

Figure 6 AI Program Implementation - Internal Controls & Risk Management

The risk and control framework is designed to help those tasked with the safe delivery of ADS/AI. KPMG-UK division developed this framework specific to ADS/AI as a guide for professionals to use when confronted with the increasing use of ADS/AI in organizations across different levels of maturity. However, the guide might also be helpful for AI practitioners. KPMG staff have categorized risks into seventeen areas as set out in the diagram above and detailed further on the following page. Note that the framework represents an early attempt to provide a holistic approach to managing the risks around the use of ADS/AI, providing guidance to the audit and compliance community, and will continue to be refined over time. This approach can be applied internally to develop service level agreements (SLAs), validate Vendors offering ADS/AI services and capabilities as-well-as to Cloud Service Providers (CSPs) offering AI—SaaS.

3

Conclusion and Recommendations:

Working with AI algorithms and technologies will allow our organization to become much more complex (to the point that it exceeds the capacity to effectively manage). The nature of this increased complexity is also self-perpetuating and although it might appear as an oversimplification, it could well introduce "technical debt"[4]. By embedding controls in a system to mitigate technical debt after its implementation is typically far more costly than designing in the right controls at the start. This assessment methodology will help our effort in advance of FY 2020 budget in AI capabilities by providing us the opportunity to address risk and implement effective internal controls. Now is an optimal time to consider taking a positive and dynamic approach to building in internal controls to address potential risk and mitigation strategies.

The use of such advanced technologies will become material for many organizations, possibly sooner than anyone expects. When the time arrives, it will not be possible to get the right internal controls in place overnight and have the capability to manage the risks effectively, or to provide assurance. Hence it is key for CIO IT Governance, effective Risk Management and Compliance practices and capabilities to be develop alongside the evolution of the usage of such technologies moving forward.

Adopting and advancing AI require an organization and the people who work in it to embrace a more scientific mind-set. This means being comfortable with a trial and error journey to the final product, accepting risks and tests that fail; and continuously testing the feasibility of the product by introducing external shocks or data and observing outcomes. Essentially, it means creating a "sandbox" (a controlled, isolated environment representative of the business environment) across the organization. This mental shift is not just solely for Heads of business units or functions, but is relevant to all areas of the organization, including the Board and other functions such as enterprise risk and compliance, HR and IT. It is particularly important to involve all three lines of defense (business lines, risk/compliance and internal audit). "*As the guardians of compliance and controls oversight, full participation in the sandbox would allow them to understand some of the critical technical aspects, and help shape, from the start, the appropriate AI governance and risk management policies*".

The Office of the Chief Information Officer (OCIO) has developed this tool specific to the development and/or procurement of AI technologies as a guide for internal professionals to use when confronted with the increasing use of AI in USAGM and across different levels of maturity as our use expands.

Three Recommendations:

- **KPMG ADS/AI Assessment and Audit Capability:**
  - **1)** USAGM has an active contract with KPMG, it would be worth reaching out to them to see if they could assist on an as needed basis to:
    - Help with crafting ADS/AI Service Level Agreements
    - Assess ADS/AI capabilities being used on-site
    - Assess Cloud Services Self Assessments and any 3rd party certifications

[4] Technical debt (also known as design debt or code debt) is a concept in software development that reflects the implied cost of additional rework caused by choosing an easy solution now instead of using a better approach that would take longer.

- **Staff Development:**

2. If USAGM has an IT technician with Software Assessment skills, would it be worth getting them the necessary training to:

- Help with crafting ADS/AI Service Level Agreements
- Assess ADS/AI capabilities being used on-site
- Assess Cloud Services Self Assessments and any 3rd party certifications

3. Despite the importance of ADS/AI capabilities use agency procurement, software development, and implementation of algorithmic systems these activities occur with minimal to no transparency, agency or public notice, community of interest input, CIO, Enterprise Application Division or Information Security oversight, or policies in place for accountability measures. Procurement officers and agency staff often lack *technical expertise* to evaluate algorithmic systems, their capabilities, and potential risk and cascading consequences to the agency ecosystem(s).

- Recommend an agency wide training program that includes the agency's strategic vision for the use of ADS/AI capabilities, what each type of ADS/AI capabilities are available, what ADS and AI are, ethic issues, insider threat issues, transparency to our stakeholders and global community of interest of our use, and topics on fairness, internal/external accountability, privacy, human rights—right to be forgotten, and information and data security, among other related topics.