White Paper

Generative AI: Promise and Peril

ATARC AI & Data Policy Working Group

October 2024

ATARC

Advanced Technology Academic Research Center

ATARC would like to take this opportunity to recognize the following AI and Data Policy Working Group members for their contributions:

Anthony Boese, *Working Group Government Chair, U.S. Department of Veterans Affairs (VA)*

Ken Farber, *Working Group Industry Chair, TekSynap*

Tanya Kuza, *U.S. Department of Veterans Affairs (VA)*

Ken Wilkins, *National Institutes of Health (NIH)*

Sandy Barsky, *Oracle*

Brian Seborg, *University of Maryland Baltimore County Emeritus*

David Randle, *National Alcohol Beverage Control Association (NABCA)*

Suman Shukla, *Library of Congress (LOC)*

Dan Haney, *Secure By Design*

Youssef Takhssaiti, *Aqua Security*

Marc Abrams, *Harmonia*

Prathibha Muraleedhara, *Stanley Black & Decker*

Craig Nickel, *Alethia Labs*

Table of Contents

Introduction

The current state of Generative AI's[1] (GenAI) capabilities to make often human-like text, image, audio, and video content has reached impressive levels of sophistication driven by advancements in machine learning, deep learning, and neural networks. Thanks to this boon of recent developments, GenAI is now poised to impact every aspect of federal agencies' operations, from providing "chat bot" services for citizens accessing agency web sites, to informing the determination of citizens' benefits, to supporting the ongoing analysis of agency data and influencing operational decision making, and beyond. Given this, deliberation and care must be taken when planning for the use of GenAI and crafting the policies that will govern it. This paper aims to situate itself as an asset for those deliberations.

The below discussion begins with a minimal but necessary level-setting review of what AI and GenAI are and entail. Following that review is a short account of some ways in which elements of the Federal Government are currently using and planning to use GenAI. The discussion then shifts from the 'is' to the 'ought' and looks at several opportunities for the Federal Government to expand its use of GenAI. Next, the discussion delves deeper into what it would entail for Government to move in the directions suggested including advice on how to accommodate the innovation required to get the most out of GenAI, a review of some of the risks involved in using GenAI that are particularly relevant to Government, and some ideas for how to manage and mitigate those risks.

Ultimately, while the authors of this paper and the Advanced Technology Academic Research Center (ATARC) do not advocate for any policy or political position, we do suggest that Government undertake the efforts necessary to safely, securely, and justly leverage more of what GenAI can offer it and the American people.

Brief Overview of Artificial Intelligence (AI) and Generative AI (GenAI)

Artificial Intelligence (AI)

Artificial intelligence (AI) has grown in pervasiveness and popular attention over that past several years to a point that a general sense of what AI is and can do is approaching "common knowledge" status. This common understanding sees AI as a model or set of models operating on a computer or system of computers which is capable of performing tasks that typically require human-level intelligence. These AI systems use various approaches to

[1] Defined in EO 14110, Section 2, subsection p: "the class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content."

recognize patterns, make predictions based on large amounts of data, and emulate abstract concepts and patterns, making them effective for tasks like image recognition and natural language processing.

In the federal context, what AI 'is' is defined more specifically in two documents, with each definition still seeing use and citation across the federal government:

1) The should-be operational definition within the executive branch from EO 14110 "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" citing 15 U.S.C. 9401(3), which considers AI:

   "...a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action..."

2) The prior definition from the John S. McCain National Defense Authorization Act for Fiscal Year 2019, which considers AI:
   a) *"Any artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets, and/or*
   b) *An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action, and/or*
   c) *An artificial system designed to think or act like a human, including cognitive architectures and neural networks, and/or*
   d) *A set of techniques, including machine learning, that is designed to approximate a cognitive task, and/or*
   e) *An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision making, and acting."*

In this paper we will use AI in a way intended to be in keeping with these definitions as well as the common impression of AI; we do not apply a novel understanding of the term.

____

As per 3GPP TS 38.821, “**A non-terrestrial network refers to a network, or segment of networks using RF resources on board a satellite** (or Uncrewed Aerial System [UAS] platform)”. There is NTN and 3GPP 5G NTN. Protocols of NTN do not always comply with 3GPP.

Generative Artificial Intelligence (GenAI)

GenAI is one specific application of artificial intelligence which specializes in creating “new” content, including bodies of text, images, code, sounds, or similar material. Traditional AI does not “create” things except in the most pedantic sense that it can “create” information, mathematical conclusions, or other analytical products.

GenAI models are advanced AI systems that utilize extensive training datasets, neural networks, deep learning architecture, and user prompts to produce diverse outputs, including images, text-to-image translations, synthesized speech and audio, original video content, and even synthetic data. By 2024 the ubiquity of GenAI and public awareness of its existence was as wide as awareness of any other sort of AI and wider than some. GenAI’s pace towards broad public awareness hit an inflection point in 2017 with the seminal paper “Attention is All You Need[2],” which introduced the type of deep learning architecture necessary for GenAI called a “transformer”, and then another when ChatGPT opened for public use in 2022 How GenAI operates is a deeply technical process, a full discussion of which would be far beyond the scope of this policy- and governance-focused paper. Nevertheless, a very basic understanding of how a GenAI model operates will be useful for better engaging with the related topics presented here.

Unlike a human who will read a sentence in chunks paying more attention to certain words and phrases than others, a machine reads in order, one word at a time, with an equal attention to each word in turn. Similarly, while a human generally ideates things to express in words as whole concepts and can fill gaps or add content at a conceptual level, a machine instead relies on an “embedding model” which is a translation of large volumes of internet or other content into a statistical representation the relationships among the parsed chunks of words (referred to as a Large Language Model, or LLM). These models work not on whole words but on word fragments called “tokens”, representing them in a vector space with hundreds or thousands of dimensions. The statistics in these models represent probabilistic relationships between the tokens and are used to predict the next most likely token to appear in the string of tokens generated by the GenAI model. For example, an intelligent chat client uses embedding models as the initial step in generating and predicting an appropriate string of tokens in response to the user’s chat inputs. The LLM uses a moving “context window” of tokens

2 https://arxiv.org/abs/1706.03762

representing the current question asked by a chat user along with a certain number of prior tokens from the previous question and answer exchanges in the chat history.

Other versions of GenAI models and model sets include but are not limited to:

- Generative adversarial networks (GANs): These models are composed of a pair of neural networks wherein one produces data closely mimicking reality while the other assesses its authenticity. These networks are widely used for synthetic data generation, creating realistic images, enhancing photo resolution, and generating art.
- Variational Autoencoders (VAEs): VAEs function by compressing data and then restoring it to its initial state. They excel in producing new data that resembles the training data.
- Transformer Models: Transformer-based models, equipped with extensive neural networks and a transformer architecture, excel in identifying and memorizing patterns and relationships within sequential data. These models are renowned for their ability to grasp context in text, enabling them to produce coherent and contextually appropriate responses to prompts.
- Autoregressive Models: These models predict future data points by learning the dependencies between the sequences in the data. They are used in generating sequential data like text or music.
- Diffusion Models: These models begin with a noise distribution and meticulously transform it into a sample via a reverse diffusion process. They excel in producing high-quality images, renowned for generating detailed and coherent visual content.

Examples of How GenAI is Being Used by Federal Agencies

Responses to the rise of GenAI by agencies are varied, but generally range from cautious adoption in select non-sensitive use cases to complete bans on its use for the time being with no agencies asserting a permanent ban on the use of GenAI nor any using it freely. Moreover, focusing on the use cases where GenAI has been adopted, one can see a wide variety of relatively low-risk use cases that seem typical among entities beginning to embrace this new technology[3].

One common use for GenAI that Government is also leveraging is using GenAI's ability to synthesize open-source information and generate coherent text to aid in the creation of first drafts of documents that humans would subsequently review and develop. Similar is the related use case of using GenAI to improve the quality of human generated text by proof-

[3] Agencies are required to report AI use cases to a central repository (hosted at https://ai.gov/ai-use-cases/), and to also publicly report their AI use cases (e.g., https://www.dhs.gov/data/AI_inventory, AI Use Cases Inventory | HHS.gov, etc.)

reading the text for tone, style, semantics, punctuation, spelling, etc. A novel example of a document generation and proofing use case within Government comes from Federal Emergency Management Agency (FEMA) in the Department of Homeland Security (DHS), which is using GenAI to help State, Local, Tribal, and Territorial governments craft plans that identify risks and mitigation strategies as well as generate draft plan elements from publicly-available and well-researched sources, which can then be reviewed and customized to fit their needs.

Another fairly common use for GenAI being leveraged by the government is for generating and reviewing code. Similar to the document generation case, GenAI is used by some federal agencies like DHS to support coders by either generating portions of code, dynamically creating usable code through plain-language prompts submitted by the coder or acting as a “copilot” to the coder by actively suggesting code as the coder is interacting with an integrated development environment. Here again, it is expected that humans will ultimately review and approve the resultant code including testing it for functionality and purpose.

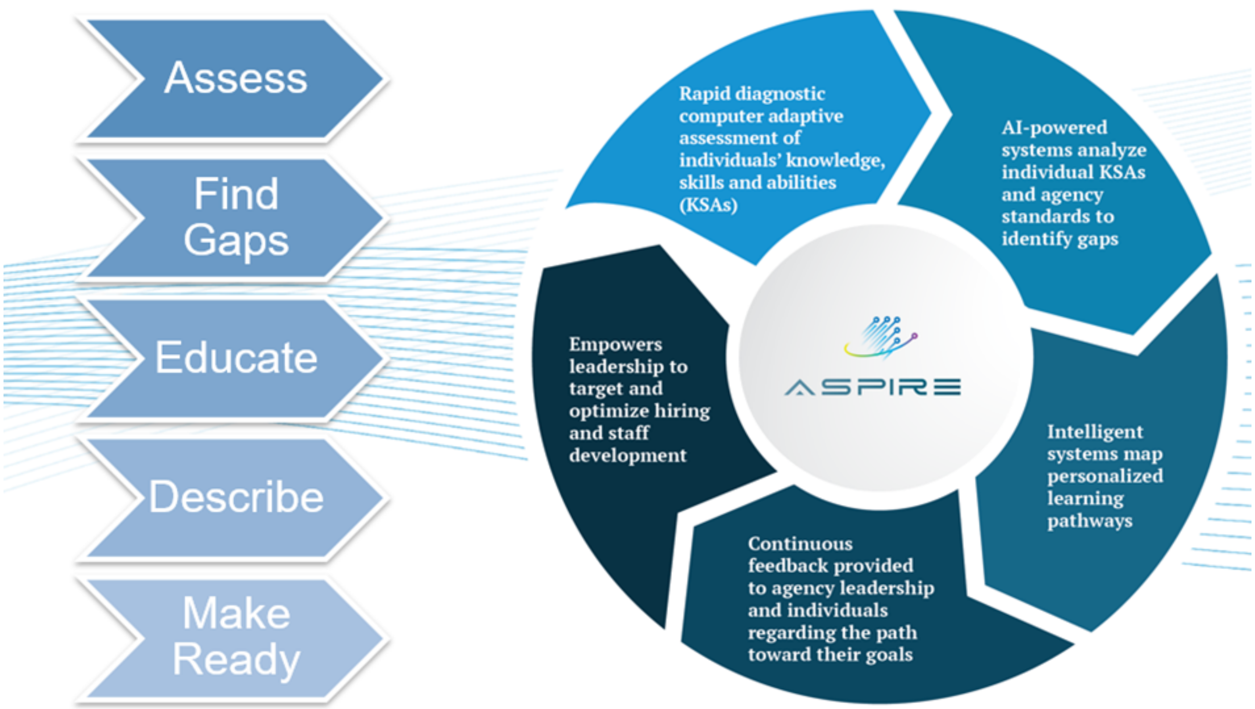
A third is the use of LLMs to summarize text and performed named entity recognition for key concepts, terms, figures, and authors. One example of this from within government comes out of the National Aeronautics and Space Administration (NASA) which finds that this use can support helping researchers identify relevant journal articles, on particular topics, and help provide summaries of those articles so that the researcher can significantly reduce their time to identify pertinent articles for further scrutiny. Similar technology is also being piloted at the Veteran Health Administration (VHA) of the Department of Veterans Affairs (VA) to review, summarize, and extract key clinical details from the notes and medical records that come into the Department after a Veteran patient is seen by an external provider, which the VA calls “Community Care Records.”

Finally, several agencies are using GenAI is differing ways to better their engagement with the public and other persons with whom they interact. For example, the Department of Energy (DOE) is exploring using GenAI to improve user experience and connectivity by producing an adaptive user interface that helps its clients more quickly connect to information that is pertinent to them. For another, several agencies use sentiment analysis (i.e., its ability to analyze words, phrases, and context, to determine the tone of text) to more efficiently and effectively take in, analyze, and respond to feedback and opinion from the public and from personnel.

Overall, the government is only beginning to use GenAI and seems to be taking a pragmatic approach by leveraging this technology for generally simple and low risk use cases. Nevertheless, this does not mean that AI and GenAI are not nor cannot be used to great positive effect by Government.

To illustrate, consider the interagency All Services Personnel and Institutional Readiness Engine (ASPIRE) from the VA as an example of a high-demand and high-value government service and product which uses GenAI and has been developed within the bounds of government regulations.

Example Spotlight: ASPIRE



ASPIRE is a VA-led FORUM Innovation Award winning, content agnostic, personnel assessment and upskilling platform designed with trustworthiness, accessibility, and equity in mind. It builds on a hybrid of successful Department of Defense (DoD) and private sector technologies to provide workforce development services. ASPIRE has been developed collaboratively by a diverse team from various backgrounds and expertise from across multiple agency, university, non-profit, and public sector partners and is designed with a focus on inclusivity and efficiency. The system addresses all users blindly at first, ensuring fair and equitable opportunities for development, and cost-effectively delivers that development guidance. ASPIRE is also beginning to work with partner universities and will be prioritizing partnership with schools in underserved geographies, community colleges, and minority-serving institutions like HBCUs, Latinx institutions, and Native American institutions.

ASPIRE uses computer adaptive assessments and intensive automated analytics to narrow down on the very specific gaps a person might have relative to the Key Skill Area (KSA) requirements of their specific role and agency or other benchmarking standard, and then automatically generates a personalized learning pathway for that person comprised of microlessons presented via multiple media to best meet varying individual learning needs. Content for these pathways comes from DoD, VA, NASA, and private sector partner resources, with more content generated by a unique cycle among a set of adversarial GenAIs, a subject matter expert, and a psychometrician, all in collaboration with an instructional designer. The process starts with one GenAI model creating the thing needed, for example an assessment item (e.g., a multiple-choice question or Parsons problem). Then, a second model evaluates the quality of that item and a third evaluates how well it fits among related items to fill the need at hand. Once the evaluator models are 'satisfied' the resulting item is reviewed by an SME for content accuracy and a psychometrician for evaluative validity and merit.

ASPIRE tackles common challenges faced by government agencies, including recruiting, motivating, and retaining their workforce, enabling that workforce to engage with emerging and rapidly developing technologies like AI, navigating evolving professional education and certification requirements, and responding to the demands of the quickly expanding technology and workforce policy landscapes. Additionally, its methods of assessing and educating are well fit to classroom and community education applications; it has been developed to follow current best practices and emerging research on teaching/learning, retention, and engagement. Here again ASPIRE leverages GenAI, this time to create avatars and scripts -sometimes live- such that users can have information presented in the language and level they need and by an instructor or conversation partner appearing how-with a user feels most comfortable or engaged.

Currently, ASPIRE is focused on AI, Data Science, and similar topic areas, but its content-agnostic nature positions it as easily expandable into other subject areas including some soft skills. It is designed to be adaptable to the unique needs of each context, application/use, and user, while maintaining a high level of consistency and coordination across varying educational and professional knowledge attainment standards, and leverages generative and non-GenAI to do so.

Opportunities for Further Use of GenAI by Federal Agencies and Their Values and Limits

Clearly, the Federal Government knows that GenAI exists, and its Executive Branch Agencies are in general starting to leverage this tool in at least select, low-risk cases. In so doing Government does better than it would were it to avoid GenAI altogether, but it is nonetheless falling behind relative to the U.S. private sector and to several other countries. Fortunately, it can make great gains in closing those gaps by seizing upon ready and near-future opportunities, including but not limited to those discussed in this section.

Gaining Efficiency and Speed on Even More (Redundant) Tasks

As seen above, GenAIs are currently deployed across several industries and agencies to do simple, repetitive tasks that do have characteristics which would make a traditional AI ill-fit for the task, such as the need for high interpersonal expertise and/or the highest levels of information security. This is the clearest and most widely accepted category of use cases for GenAI offering the benefits discussed below with generally very low risk, opportunities that should continue to be seized by those doing so and pursued by those who are not.

AI's comparative advantage to human personnel in the arena of simple, repetitive tasks stems mostly from its relative speed, inerrancy, and tirelessness. A GenAI model can analyze, generate, and summarize data or text significantly more rapidly than a human counterpart,[1] and can perform that function around the clock with no need for rest. To quantify: if we estimate that AI is a modest 5% faster than humans and can work a conservative three times longer than a person per day and without days off, then the AI has 466% of the work capacity of an average human worker.

This additional work capacity not only creates an opportunity for AI to generate more output, but also creates opportunity to increase product reviews and quality assurance instead or additionally. This ability for AI to, for example, generate a model, check over the model, re-tune the model, and then re-check the model all in the time it would take a human to simply generate the model, means that AI-generated models, and other products, could be more robust and reliable than human generated products within the same product delivery timeframe. Moreover, this increased capacity can still be accessed without taking onboard most AI-related risks by opting for human-AI teaming, especially where the AI is generating templates, first drafts, or other tools that the human can use to enhance the human's workflow. This teaming can also improve accuracy and consistency and shorten time to decision.

Regardless of how one gets to it, this work capacity differential entails two further comparative advantages to AI: fiscal cost savings and opportunity cost saving. The fiscal cost savings comes from AI generally requiring less funding to deploy on a per-task basis both because it takes less time, or empowers a human to take less time, on a task and is cheaper per hour for that work. Fiscal savings also comes over time thanks to AI's generally flat continuing cost as opposed to human personnel's year-over-year rising costs and the relative ease of scaling a given model to levels far beyond what one human could maintain, thus allowing single models to offset multiple humans' worth of costs.

The opportunity cost savings arises because of recovering said humans' work time. By deploying AI and freeing up humans' time, they can perform more meaningful, interesting work that requires creativity and human insight that AI cannot achieve. This may lead to the germination of innovation, the correction of bad habits developed while workloads were high, greater likelihood that issues will be noticed and time will be taken to correct them, and other similar beneficial activities that can accompany a workforce with enough available worktime and moderated pressure.

However, although laden with benefits and appealing for many use cases, not every simple and repetitive task is a good use case for AI. For example, tasks that require high interpersonal expertise and emotional intelligence would not be an appropriate place to deploy AI, though unfortunately some have tried. These situations might include medical intake, personalized customer service, and advanced trouble-shooting service desks. Also, ill-fit are those instances where the action is simple and repetitive but context varies drastically from instance-to-instance. Given how AI are trained and the limits of their perception and understanding, the risk is high that an AI will fail to perform or will perform incorrected and potentially even dangerously. Finally, function aside, GenAI tends to create issues for information security given how complex and opaque their sources and reporting are, and the way in which they tend to regurgitate information given to them in prompts. Given this, unless one has private, isolated generative models and tools, it is best to not use GenAI on work involving protected information such as PHI, PII, and classified information.

Improving Research

GenAI pushes the boundaries on the value that AI can add to research efforts far beyond what more traditional AI offers. For example, GenAI enables the more rapid development of synthetic data, a source of data that is growing more popular and useful as data and privacy security standards get higher and databases get harder to reliably protect especially in contexts where cyberattacks are common. Taking this to an extreme, GenAI could make possible the creation of exact enough digital twins for the twins to be useful for hypothesis and method testing. In fact, digital twins are already becoming a well-established option for research subjects in many contexts where the subject is not a complex living organism, and even that barrier is being overcome currently.

Also useful for research is GenAI's increased ability to screen large libraries of information and then synthesize commonalities and identify outliers from which it can generate new recommendations for research. For example, GenAI's are currently used to screen large compound libraries to seek out novel compounding opportunities to support drug, materials, and fuel development, and to do similar work on different source materials for other industries.

GenAI can also be used to increase research capacity among those aspects of research that are sometimes considered research support. Examples here could include accelerating candidate selection for drug trials, identify and assess practices and tools used by other studies to optimize project design, and assisting heavily in drafting literature reviews, grant applications, reports, and other writing-heavy research-related activities.

Finally, GenAI models will soon be able to take on neural style transferring. Here the AI would be mimicking the distinctive reasoning and knowledge base of a given researcher thus making that researcher available for consultation at any location and time with an internet connection, allowing researchers past and present to be more places at one time supporting research well beyond what their natural limitations would allow. This would be a huge asset for the speed and quality of research, and a massive step towards the democratization of access to science.

Performing Better Outreach and Communication with Citizens and Personnel

Every agency engages in outreach and external communications, whether to the public, to the interagency, or to the Legislature. For some this is a major part of their activities while for others this is a tiny and oft-overlooked part of their operations. Regardless of its importance and scale, GenAI can be an asset. As mentioned above, GenAI can help by rapidly creating drafts and/or options for communications materials and graphics, translating materials, and making materials more accessible.

Language translation is another potentially significant benefit to agency outreach and external communications. Domestically directed programs would do well to translate outreach and information materials into the several languages commonly spoken by those living-in and visiting the United States. As new policies and technologies increase the number, diversity, and notice and reporting requirements of government programs and activities, the need for document generation will only grow. Moreover, for each document made there are a handful more translations that need to be made, making the translation workload generally easier and more rote work, but also more voluminous– a perfect place to deploy a generative language model. Translation services are also useful for international exchange, Americans overseas, departments with large overseas footprints, with large proportions of local staff (e.g., USAID), and by the DoD during engagements and occupations abroad.

Accessibility is another potentially critical benefit of GenAI. The best idea and the most spectacular outreach materials will still fail if those materials are not accessible. Common examples of access supporting considerations include accommodations for disability, e.g. the deaf or hard of hearing, blind or poor-sighted, color-blind, and difficulties with manual dexterity. Material access is another concern, such as whether a person has the time, means, and wayfinding knowledge to access the materials (for instance, accommodating those with no or little internet access, those who might be home-bound, citizens overseas, etc.). We mentioned translation among languages above, but an access concern often overlooked is translation among communication levels. A message that is communicated in a way inaccessible to most, or in a way that will seems overly simple to most, etc. is unlikely to be taken well and seriously by a majority of its intended audience. Materials that are reliably mid-level, something most people can access, would be fairly effective, but best would be to have all materials available at different communication levels. This is especially the case for policies and procedures being communicated internally to agency personnel where it is often the case that different people will simply need to know different details about given topics.

Among the many GenAI technologies that would be useful for outreach and communications, AI avatars stand out as the most impressive and most poised to present massive value. Providing AI avatars to deliver messaging or even to interact with target audience allows for greater engagement with and understanding of messaging. With GenAI these avatars can do more and better by being able to be customized to a given sub-audience or audience member, allowing the messenger to look, sound, and act in a way that would most appeal to them and best enable their engagement and understanding.

Using GenAI as a Policy and Governance Tool

Government and its governance are replete with scores upon scores of overlapping laws, executive actions, agency policies, agency sub-component policies, professional regulations and standards, standard operating procedures, versions of each, and sometimes more. It is time and expertise intensive to be able to understand compliance requirements, find those rules and regulations in their current form, understand them, and referee conflicts and inconsistencies among them. Complicating these challenges is the general inconsistency of terminology, semantics, data, and information labeling among these governing documents.

GenAI presents an opportunity that would not have ever prior been practicable: to conduct full reviews of policy to provide exhaustive answer to queries from users about the rules to which they are beholden in a given role or for a given activity. Additionally, while doing so, GenAI can develop guides to cross walk and coordinate among laws and policies, which is valuable in any case but especially helpful with interagency, public-private, and international collaborations. Moreover, GenAI could also be drafting updated, better fleshed out, and/or cured policy as it goes along. One might even imagine a future wherein all policy, legislation, guidance, and regulatory language is automatically generated to be consistent with the pre-existing corpus of related language and documentation.

Having clear, consistent, and up to date polices across government would be an invaluable asset, and then also having a GenAI tool to better navigate the policy landscape would be a true game changer.

Building (Possibly) Better Arbiters

Though likely still years away, there is some interest in focusing GenAI development, especially in the public sector, on creating artificial decision makers. The simpler version of this effort might be a machine that can review policies, their historic application (precedence), and applications and activities under that policy to ensure consistent and rational applications of policy. For example an AI that reviews leave requests, expense reports, police tickets and ticket forgiveness, and other narrow and contained universes of decision making. In a more robust version, one that may require a "General AI", we can imagine AI judges, AI loan officers, AI immigration personnel, etc. In any version, the ability of AI to consider myriad counterfactuals rapidly and tirelessly could in the future make AI decision making more reliable and better tested than human decision makers. Current challenges related to AI explainability, issues related to legal rights concerns (such as being judged by a "jury of your peers"), and general public distrust of AI systems all present potential risks of such AI systems, and are discussed below.

Accommodating GenAI Innovation and Integration

Successful integration of new technologies requires agencies to make room in their operational capacity and in their workflow to accommodate personnel learning curves, as well as operating the new technology in parallel innovation. Making this room can be complicated and sometimes costly, if GenAI is being deployed to a use case and context well fit for it, then the gains over time could, given the AI advantages discussed earlier in this paper, return this investment several fold. While the specifics of what is done for this change management exercise will vary from instance to instance, there are a few things nearly every firm or agency will need to consider.

An effective first step is establishing a risk management framework (RMF). An RMF outlines the standards, policies, practices, and tools that an organization will use to identify and then prevent, mitigate, or combat, potentials risks that will accompany the deployment of a given technology. Each technology type will need its own RMF, so having an IT RMF or Data RMF is not sufficient to cover AI, and an AI RMF needs to include the right materials to enough depth to cover issues specific to GenAI. Additionally, once established, this RMF will need to be reevaluated and updated regularly, especially while AI technologies are still developing and changing rapidly. For a good examples to use as guides when making and updating an RMF consult the NIST AI RMF, which is the most widely recognized and used AI RMF, the Department of Energy AI Risk Management Playbook, or, on the private sector side, the IBM AI RMF.

The next need is to make actual space, especially compute, storage, and network capacity. The compute space need varies greatly depending on the type and size of GenAI model being generated. These models can get very large, sometimes surpassing 1TB of RAM, and the storage space for the model and data can measure in the petabytes (1PB = 1000 TB), though most will require less. Moreover, this virtual space rests, at least in part, on physical chips and drives, which will need to be mounted, powered, and cooled. Space of all sorts will be needed, and planning for that early is better than being reactive, which can cause workflow interruptions, slowed development speed, and potentially product delivery delays.

Workforce development is another need, educating current and would-be staff on GenAI and engaging them into the innovation process. No matter how good your RMF and how well you've done selecting your GenAI tool(s) and making space for it and the people who will maintain or use it, if your personnel are not sufficiently well trained on how to operate the tools, then there will be risk, backlogs, and significant reduction of the benefits the GenAI would otherwise have provided. To assess readiness and remediate any gaps therein, it would be best to build or buy an assessment and workforce development platform designed specifically for this purpose, or, failing that, at least be sure to provision personnel will high-quality and timely educational materials. ASPIRE is one such platform and is readily accessible by Government. Outside Government, for high quality materials with any associated tools, consider offerings from universities that excel in the AI topic areas such as MIT, Carnegie Mellon, and Stanford which are often made available to government personnel at no cost.

Finally, carefully screened AI subject matter experts can fill gaps too far removed from current personnels' knowledge, skills, and abilities for their upskilling into the roles to be tenable. After current personnel have been evaluated and upskilled such that any willing and able current personnel are engaged in GenAI work, the organization will likely have some GenAI-centered roles remaining and some other roles to backfill. It is important to make sure anyone applying, especially for their GenAI skills and expertise, is properly vetted before hiring them is considered. Formal credentials in AI, especially in GenAI, are new and unproven, with little consistency in offerings and little clarity in what a given credential holder learned and is able to apply in practice.

GenAI Risks of Particular Concern for Federal Agencies

GenAI has risks related to its use like any other technology, though with GenAI there is a real possibility that several of those risks are amplified due mainly to the nature and opacity of how GenAI models handle and use data. While the risks are varied and numerous, there are some that are particularly relevant to the Federal Government.

GenAI has high information-flow-based risks. Unlike other prompt-taking systems, GenAI models accept prompts, share them back and store them for use by the model (and those with access to the model) to better train the model. This process makes these prompts and the content contained within them available to those same people as well as their supervisors and investors, and potentially to anyone else who queries the system in the future. While the risks posed by prompt interaction with a GenAI model are relevant to any would-be user, many US citizens entrust the Government with sensitive, personal information that is not meant for the public (e.g., draft materials, classified information, or personally identifiable information, patient personal identifying information and/or personal health information). Given this, the likelihood of a leak of protected data and the impact level of a potential leak are both much higher for the Government than for other entities.

The Federal Government has a high bar of accountability for being able to show and explain how decisions are made, and to ensure that that decision making process is as reliable, unbiased, and incorruptible as possible. This high bar develops from facts of the matter about the government, for example that it is the enactor of the public's will and steward of the public's funds, and from the matters it tends to affect, such as the American people's lives, wellbeing, and livelihoods. This all runs contrary to accepting several of the risks present when using GenAI, or even any AI. One such risk is the limited traceability and irreproducibility of GenAI outcomes, which could lead to harmful or even illegal decision-making. Another is that GenAI models are prone to giving dated information and now-defunct conclusions derived from outdated training data. Because LLMs achieve their predictive abilities, in part, from ingesting very large data sets, these models will always exhibit the influence of the older data in their results. This is because models are trained on data up to a date, at best the date that the model's training stopped but more likely a date soon before the training began. This means the model will less data that are more recent, and a larger percent of data that are older. Some of the older data may be more relevant to a given prompt, especially in the Government context, given the inconsistent frequency with which laws, policies, and regulations change.

Another well-known risk is that AI is not neutral: AI-based choices are susceptible to inaccuracies, discriminatory outcomes, and embedded or inserted bias. If the data fed into these systems represents biased decision making, the output of these models will accurately recreate those biases, perpetuating historic harms. It is also misaligned with how AI performance and development have been benchmarked to date. The performance focus for these tools has been to generate "credible", or "similar" output, rather than to generate output that is in any sense "accurate", "comprehensive", or "timely", with the latter three being much more important and relevant to Government than the former two.

In addition to the potential benefits of synthetic data discussed in an earlier section above, the AI-driven creation and use of synthetic data also introduces a critical risk. As part of the AI system development lifecycle, generating synthetic data through algorithms may drive research into conceptual tunnel vision, and researchers may not be aware as it happens. This is an area of active research[4] which indicates that the ML models have limited ability to inject "unknown unknown" variations and characteristics into synthetic data sets. As these less robust data sets are used to train new data models, this recurring process leads to a compounding effect as ML model analysis drives less robust data sets which drive incrementally lower quality analysis, which drives even less robust data sets. The result may be an ever narrowing data set richness and reduced quality of ML model performance, while researchers who are focused entirely on their synthetic data sets remain unaware of the degradation.

Another concept discussed earlier in this paper – digital neural twins of researchers – may present the benefits described, but those benefits come with potential risks as well. Unless a strong, effective regulatory framework is put in place, the benefits of those digital neural twins may accrue not to researchers but to AI system owners. In addition to those researchers risking the loss of income and intellectual property, society as a whole would be at risk of the AI system owners executing a vast "land grab" of intellectual property. If the AI system owners are able to acquire copyright or patents based on the digital neural twins' research, they would

4 https://www.forbes.com/councils/forbestechcouncil/2023/11/20/the-pros-and-cons-of-using-synthetic-data-for-training-ai/

potentially be able to explore, discover, and establish IP ownership over scientific domains and creative spaces at machine speed.

Current funding lines and strategic priorities across the interagency and at the Presidential level are not aligned to cultivate a government workforce ready to leverage AI and GenAI tools, even if those tools themselves are sufficiently safe and secure. Government would be open to massive risks born of untrained users, limited access to system maintainers, and insider threats because there is little-to-no focus on the appropriate recruitment, retention, and -most importantly- upskilling required to have a workforce able to properly and safely use new and potentially high-risk tools like GenAI. No matter how safe a tool nor how well considered and written the guidance and rules for its use, if the user is not duly educated and trained, then the tool is as risky as if there were none.

The Federal Government is massive with over four million employees in the Executive Branch alone[5]. Wide deployment of GenAI models and tools across Government would create a significant amount of computer use in the creation, development, use, and maintenance of the GenAIs and storage of all the affiliated data. This increased computation and storage use would certainly come with increased costs to the taxpayer and negative environmental impact. While broader Government use of GenAI models would necessarily result in increased environmental impact, the net impact is difficult to estimate, given that the GenAI models would be displacing human agents (and thereby avoiding those agents' related environmental impacts).

Finally, the Federal Government and its servers and system are also a massive target for adversarial attacks, and GenAI databases and models would be prime targets given the information and insights they could contain in their design, what they "know", and the critical functions they may serve. The risk of attack is higher for the government than for any other entity, which may translate to an increase in risk to the models, their data, and all government operations that rely on them.

Effectively Managing GenAI Risks

Emphasize and Prioritize Transparency, Explainability, and Interpretability

It is hard to overstate the risk mitigation enabled by mandating that all AI models be transparent (meaning that users can examine the AI's data, logical flow, and analytic structures to see how it does what it does), explainable (meaning that the AI can provide a clear and cogent account of its output), and interpretable (meaning that the AI's explanation can be understood by a human; ideally there will be different levels of explanation, appropriate for

[5] https://www.whitehouse.gov/about-the-white-house/our-government/the-executive-branch/

users of varying AI expertise). Therefore, it is critical that any AI, and especially those used by public sector entities, have these attributes.

Moreover, it is useful and valuable to have transparency into explainable and interpretable model structures and behaviors, but this is not the only relevant sort. Also important for risk mitigate is interaction transparency, which deals with the communication and interactions between users and AI systems. Additionally important, especially for Government, is social transparency, which addresses AI deployment's ethical and societal implications, including potential biases, fairness, and privacy concerns.

Soft Start GenAI Efforts

Soft starting GenAI projects offers several key benefits. It allows organizations to gradually ramp up their GenAI initiatives, reducing the risk of costly failures and building trust in the technology. By starting small with controlled experiments in an isolated AI “sandbox” environment, teams can validate assumptions, test hypotheses, and iterate rapidly before scaling up. This approach also enables a more comprehensive evaluation of AI solutions, assessing not just technical performance but also factors like explainability, fairness, and alignment with operational objectives.

Although not tailored to GenAI as discussed above, the NIST AI Risk Management Framework (AI RMF) is nevertheless a valuable resource and guide on risk identification and mitigation, and it can support soft start efforts by providing a structured methodology to benchmark AI systems against known standards and quantify their impact. The AI RMF comprises four core functions: Govern, Map, Measure, and Manage. These functions guide organizations in managing AI risks throughout the AI lifecycle, ensuring that AI systems are valid and reliable, safe, secure, accountable, transparent, explainable, and fair. Utilizing these standards, decision-makers gain the insights needed to confidently greenlight deployments. Ultimately, soft starting accelerates the path to production-ready AI while mitigating downside risk.

Take Time to Build Trust

Building trust in AI solutions requires a multifaceted approach that addresses key risks and prioritizes transparency, robustness, and human oversight. Effective data governance is critical, as is having clean, high quality, and low bias training data as that data directly impacts the trustworthiness of any AI trained on that data. Given this, organizations should implement rigorous data quality initiatives, regularly testing and validating datasets to identify potential biases or inconsistencies that could lead to flawed or even harmful AI decisions thereby ensuring that the AI does not erode trust through error.

Additionally, diverse, multidisciplinary teams should be involved in the design, development, and deployment of AI systems to ensure a range of perspectives and experiences shape the technology. For instance, establishing dedicated roles and advisory boards focused on AI ethics can help guide responsible usage and proactively mitigate risks.

Keeping AIs under close oversight and regular assessment is also important. To enable this given the time investment required, automated processes can play a crucial role in building trust by providing consistent and repeatable assessments of AI systems. These automated evaluations can conduct regular audits and evaluations to ensure that AI solutions meet established standards and function as intended. This process helps identify potential flaws and ensures accountability. Additionally, employing automated processes for evaluating AI solutions ensures repeatable and unbiased validation, thereby enhancing trust in the technology. That said, it is also important to take and consider feedback from users and other relevant AI actors to enhance system performance and trustworthiness over time.

Moreover, building off of the first subsection, AI systems must provide a degree of transparency and explainability to foster trust. Techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) can help unpack the "black box" of complex models, shedding light on how specific features influence predictions. By enabling humans to audit and interpret AI decisions, organizations can more readily identify potential flaws and maintain accountability.

Ultimately, building trust in AI requires ongoing commitment, proactive risk management, and a people-centric approach that prioritizes transparency, fairness, and continuous improvement. Adopting frameworks like an AI RMF and leveraging other tools and strategies discussed in this paper can help organizations navigate the complexities of GenAI risk management and foster the responsible development and deployment of GenAI technologies.

Use Owned Environments and Models When Possible

Many of the risks associated with GenAI are related to how foundational models have been pre-trained as well as access to the models and the data used to train, evaluate, and test them. One effective way to mitigate these risks is to develop models from scratch within an agency-controlled environment, and to deploy them within agency-controlled infrastructure. This provides the agency with the greatest confidence that all of the risk management controls that the agency has decided are appropriate for the GenAI system are accurately and consistently applied. There are a few downsides to this approach, however, including cost, access to technological advances, and staffing. Establishing and maintaining a security-controlled environment is not inexpensive. The LLMs driving today’s advanced GenAI systems, such as ChatGPT, have received billions of dollars in investment over the past decade.

Conduct SBOM/SYSBOM Analysis

As a software system, any GenAI system brings with it risks associated with software procurement, development, and deployment. The federal government has recently required[6] all federal agencies to apply the NIST RMF, which includes a requirement to use a "software bill of materials" (SBOM) in all software procurements. An SBOM is a nested inventory of the pieces of code that make up a piece of software, where they live, who has access to them, who funded their development, and more. EO 14028 of 2021 spells out an SBOM as a "formal record containing the details and supply chain relationships of various components used in building software." System bills of materials are the same, but for systems instead of software.

SBOMs can also be indicative of a developer or suppliers' application of secure software development practices across the SDLC. Below is an example of how an SBOM may be assembled across the SDLC.

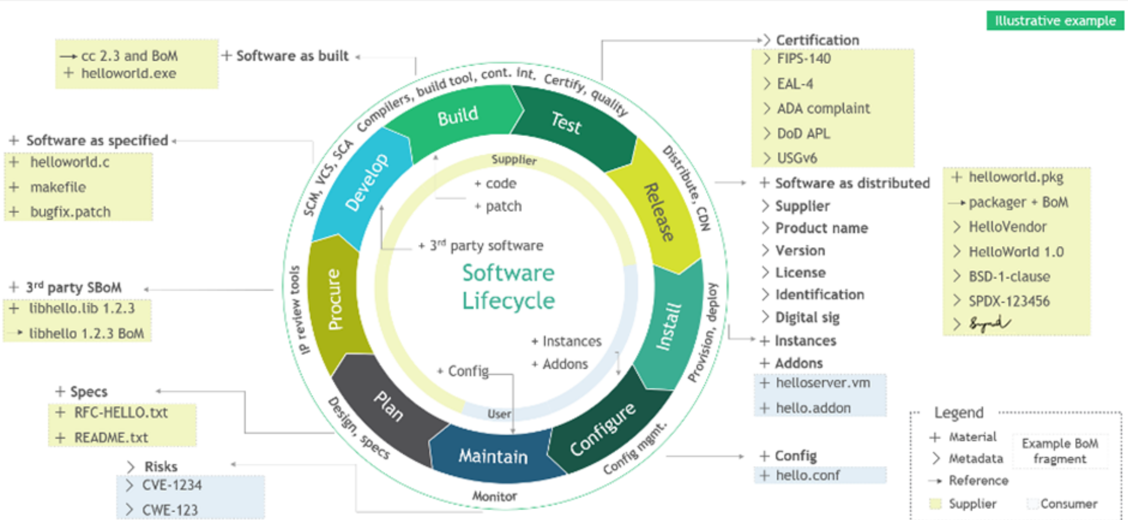


Image source: https://www.nist.gov/itl/executive-order-14028-improving-nations-cybersecurity/software-security-supply-chains-software-1

Regardless of the specifics of its definition or context, an SBOM must have certain minimum elements to meet federal standards, minimums which also provide a good starting point for private entities to begin to build their own SBOM/SYSBOM standards and practices. These minimums are:

[6] President Biden's Executive Order 14208, "Improving the Nation's Cybersecurity"

Minimum Elements	
Data Fields	Document baseline information about each component that should be tracked: Supplier, Component Name, Version of the Component, Other Unique Identifiers, Dependency Relationship, Author of SBOM Data, and Timestamp.
Automation Support	Support automation, including via automatic generation and machine-readability to allow for scaling across the software ecosystem. Data formats used to generate and consume SBOMs include SPDX, CycloneDX, and SWID tags.
Practices and Processes	Define the operations of SBOM requests, generation and use including: Frequency, Depth, Known Unknowns, Distribution and Delivery, Access Control, and Accommodation of Mistakes.

Source: *The Minimum Elements For a Software Bill of Materials by National Telecommunications and Information Administration*

That said, these are just minimums. Depending on what information on the software is available, the level of value and important of the use case, the security needed for the use case, suspicions about the software, future laws, and/or client demands might demand much more. Indeed, deep SBOM analyses in use today take steps as deep as analyzing binary composition and ablating chips to check for errant components to truly assess a software or system.

Summary and Recommendations

AI systems are currently deployed across several industries and from local to federal government. To date, Federal use of GenAI has been largely made up of tools assisting humans with low-risk, simple tasks among those agencies that use GenAI at all. While the government is better off doing these limited GenAI trials than not engaging with the technology at all, there are opportunities available to the government to get more out of GenAI with limited additional effort and no uncontrollable risks. These opportunities include using GenAI to gain efficiency and speed on additional redundant tasks beyond those for which Government currently uses GenAI, to improve research, to perform better outreach and communication with citizens and personnel, to coordinate and make more user-friendly Federal policy and governments, and -once the technology is available- to build better decision-making agents for the public good.

In order to be positioned to successfully use and gain value from GenAI for these and other lines of effort, there are certain steps Government should take to ensure they have an environment set up to house innovation and to ensure that risks are known, seen, and then avoided or mitigated. To set up an innovation-friendly environment, agencies should establish a risk management framework, make actual space (especially compute and storage space), educate current and would-be staff on GenAI and engage them into the innovation process, and hire carefully screened AI subject matter experts to fill skills gaps remaining after upskilling current personnel where possible. For risk mitigation, agencies should emphasize and prioritize transparency, explainability, and interpretability, soft start GenAI efforts, take time to build trust, and conduct SBOM/SYSBOM analyses. Additionally, no AI system should be solely responsible for any decisions that will directly impact a group's or individual's life or livelihood. Decisions of this magnitude have too much risk for high-impact and often irreversible damages and, for the foreseeable future, should be the responsibility of human agents.