HARVARD Kennedy School
MOSSAVAR-RAHMANI CENTER
for Business and Government

AI for the People: The Use of AI to Improve Government Performance

Mark Fagan
Harvard Kennedy School

2023

M-RCBG Faculty Working Paper Series | 2023-01
Mossavar-Rahmani Center for Business & Government
Weil Hall | Harvard Kennedy School | www.mrcbg.org

AI for the People: The Use of AI to Improve Government Performance

Mark Fagan[1]

AI is Everywhere

7:00 AM, the alarm goes off. 8:00 AM, you have a cup of coffee, scan the news, and check your email on your phone. During that single hour you have interacted with artificial intelligence (AI) numerous times. The coffee beans were harvested based on an AI algorithm. The news feed…curated by AI. The ads that came with the news…AI. The facial recognition to open your phone…AI. And more. AI is everywhere in your personal life. It is also increasingly used by your government.

From the RMV to the IRS to the TSA to the FDA, government agencies seek to provide quality services to its constituents, efficiently and with equity. Analytical tools have been introduced over the past 50 years to facilitate achieving this objective. These tools support an array of government functions from distributing public assistance to public safety and security to financial policy. Based on the success of deploying analytical tools and the rapid development of AI technology to complement these tools, agencies are turning to AI to accelerate creating value for the public.

AI Basics

AI is defined by the Oxford English dictionary as "the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages."[2] The operative word is intelligence, "the mental quality that consists of the abilities to learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate one's environment."[3]

There are four characteristics of AI. First, AI has the capacity to make decisions or at a minimum support your decision making. Second, AI the decisions require a combination of attributes of human intelligence from perception to problem solving to reasoning to learning language.[4] Third, AI systems combine data sources and take action based on the analysis. This contrasts with pre-programmed responses. [5] Fourth, the decision making provides feedback to the system for continual improvement.

---

[1] Mark Fagan is a Lecturer in Public Policy at Harvard Kennedy School. Research for the paper was provided by Emily Ratte and Nanditha Menon, Master students at Harvard Kennedy School.
[2] Oxford English Dictionary
[3] https://www.britannica.com/science/human-intelligence-psychology
[4] https://www.britannica.com/science/human-intelligence-psychology (P2 of 26)
[5] https://www.brookings.edu/research/what-is-artificial-intelligence/

What constitutes AI has evolved. Alan Turing, credited with the concept of thinking machines in the 1950s, established an AI threshold requirement that the computer to solve puzzles in a manner similar to humans.[6] Another view, articulated by John McCarthy, set the test as "getting a computer to do things which, when done by people, are said to involve intelligence."[7] Of course involving intelligence is subject to debate.

Since Turing imagined AI, it has become reality by his definition. The early applications were playing games, checkers specifically. Success here led to an interest in seeing if a computer could beat a world chess champion. The theoretical answer was yes. The computer could evaluate all permutations of its own and the opponent's moves and devise a plan to win. For years the practical answer was no since there was insufficient processing capacity to test all the alternatives. By the 1990s the combination of greater computing capacity and increasingly efficient use of that capacity, enabled IBM's Deep Blue computer, in 1997, to beat chess Grand Master Kasparov in a dramatic showdown.

Yet another view of what constitutes AI comes from Darrell West of the Brooking Institute. He suggests AI requires three ingredients: intentionality, intelligence, and adaptability.[8] Intentionality refers to the intent of the human developing the AI to have the computer make decisions in the manner of a human. The AI algorithms that operate autonomous vehicles are designed to replicate and replace a human driver. West describes intelligence as using machine learning and data to make intelligent decisions. The autonomous vehicle must find and follow the best route to its destination, while ensuring safety for its occupants and those in its path. At the core the computer is determining how to incorporate multiple factors and make tradeoffs. Adaptability refers to the ability of the computer to learn to adjust to new and ever-changing conditions. The autonomous vehicle is constantly gathering new information about its operating environment, and adjusting its performance based on instantaneous changes in conditions.

AI Use Cases

The evolution of AI has not been constrained by ideas for applications; rather, the limitations have been in data availability and computer processing speeds. In the past decade advances in data gathering and computing capacity have enabled a dizzying array of applications from hiring employees to setting bail in criminal cases to recognizing faces to writing reports to driving cars.

The broad range of AI applications are categorized into three groups based on level of human thought or "sophistication." Artificial intelligence applications are the least sophisticated. They seek to make intelligent programs and machines through writing static code. AI chess and voice-controlled applications (e.g. Siri) and robots for household chores fall into this category.

---

[6] For a comprehensive description of the Turing test see: https://www.britannica.com/science/human-intelligence-psychology (P10 of 26)
[7] https://www.britannica.com/science/human-intelligence-psychology (P2 of 26)
[8] https://www.britannica.com/science/human-intelligence-psychology (P2 of 26)

Machine learning begins with written code and then self-improves with each decision and piece of new information. Software that anticipates your email writing and fills in the next word or phrase is an example. Deep learning or neural networks are the most sophisticated applications. Here the system is programmed to replicate the thinking humans do in their brains accepting in and digesting large amounts of unfiltered and unstructured data. You do this while driving instinctively; autonomous vehicles do it via computer-based deep learning.

Today, AI has the potential to impact every aspect of a government organizations as they strive to deliver greater public value. Consider the US Department of transportation (DOT). Its top priority is a safe transportation system. Predicative analytics can be used to improve safety by identifying infrastructure at risk of failure early enough for a planned response. Safety can also be improved by the adoption of autonomous vehicles. 46,000 people die each year in car crashes; 90+% are the result of human error. If even a portion of the human error fatalities can be eliminated though autonomous driving, the public value gain is large. DOT also seeks to reduce climate change and increases resilience. AI can improve the identification of infrastructure at risk from climate change and prioritize resilience investments.

AI Use Cases for the Government

The AI use cases in the public sector address four core functions of government: (1) providing safety and security; (2) delivering public services; (3) collecting revenue; and (4) being effective and efficient. Examples are found around the globe. AI for safety and security are in evidence in airports in many countries. Narita Airport in Japan uses advanced robots to enhance security though detection of anomalies including people, baggage, and equipment. It also provides a video history of activity in the airport. The airport is also using facial recognition software to facilitate passenger movement throughout the airport while maintaining security protocols. Face Express identifies passengers via a photo image at their first touchpoint at the airport enabling them to pass through the rest of the airport without interruption. The program also reduces staff workload.[9] Other security applications, though more controversial, include identifying cyber threats, predictive policing, and terrorist threat detection.

AI is being used to improve the delivery of many governmental services from healthcare to criminal justice to weather. Chatbots, computer systems that simulate human conversation, are a high-leverage tool to provide quality services efficiently and with equity. The Rwandan government worked with Babylon Health to create chatbots capable of assisting the triage process for patients calling the hospital. Upon hearing the callers' symptoms, the triage tool would provide recommendations for accessing care. In theory, this would help address minor medical concerns that would otherwise require waiting for a doctor's appointment, and prioritizing those needing immediate care. As this was a pilot program, the AI only provided suggestions to nurses, who conveyed information back to patients; Future iterations could

[9] https://www.naa.jp/en/airnarita/automation.html

remove the call center nurses altogether, enabling access to health care at a vast scale, and more efficient use of staff resources.[10]

AI is facilitating tax collection. The OECD in 2019 reported that more than 40 tax authorities are using or plan to use AI.[11] For example, the Spanish government has teamed up with IBM's Watson for the use of this system to address questions about value added taxes. Since introducing Watson the number of email inquiries to the tax authority have declined by 80 percent as questions were addressed by Watson.[12] Moreover, AI is being used to detect payment anomalies and fraud. Several countries use AI to predict bad debts and prioritize collections. The US Internal Revenue Service is introducing an AI chat bot that is able to help those behind on tax payments to set up a payment plan.[13]

AI is also driving greater government effectiveness and efficiency, enabling better decisions and delivering services at lower costs. Attracting, hiring and retaining employees is a costly activity for government agencies. AI is improving the entire process. AI facilitates hiring by using algorithms to sift through large numbers of applications and select profiles best matched for the position at hand period. AI supports the hiring decision by analyzing potential for skills that the employee might not yet possess. In theory this also introduces less bias into the hiring process compared to traditional procedures, although many recent cases have found that bias embedded into AI systems through historical discrimination still leads to disproportionately low hiring of women and racial, socioeconomic, and other minorities.[14] Retention is improved by AI systems that provide human resources teams with information on employees at risk of leaving, and providing recommendations to ensure retention, from skills or leadership training to higher wages. Another application of AI technology for efficiency and effectiveness are programs that use chat bots to respond to constituent questions. An example is Singapore's Ask Jamie a virtual assistant helping businesses and citizens interact with 70 government agencies.[15]

In the US AI is used in support of public safety from eliminating road crashes to predicting criminal activity to evaluating a crime scene to identifying to obtaining the acoustic signature of guns . (The risks of these applications are detailed later in the article.) AI can improve road safety by analyzing vehicle crash data (e.g. vehicle type, weather, traffic levels, time of day, etc.) to predict unsafe conditions and provide recommendations for improving road safety. AI is also used to improve safety by predicting criminal activity. Algorithms are developed based on a wide array of historical data to predict where and when crimes might take place. This approach is proactive rather than reactive and leverages "data, intelligence and all the technology at our

[10] https://www.weforum.org/reports/chatbots-reset-framework-rwanda-artificial-intelligence-ai-triage-pilot
[11] https://read.oecd-ilibrary.org/taxation/tax-administration-2019_74d162b6-en
[12] https://www.ciat.org/the-use-of-artificial-intelligence-by-tax-administrations-a-matter-of-principles/?lang=en
[13] https://federalnewsnetwork.com/artificial-intelligence/2022/06/irs-expands-ai-powered-bots-to-set-up-payment-plans-with-taxpayers-over-the-phone/
[14] https://www.science.org/content/article/even-artificial-intelligence-can-acquire-biases-against-race-and-gender
[15] https://www.tech.gov.sg/products-and-services/ask-jamie/

disposal."[16] Ai is also helping detectives evaluate crime scene. The goal is to provide a comprehensive view of the relationship between objects, people, buildings, and weapons, at a crime scene using multiple images and observations to surface evidence or explanations that might otherwise be missed by individual law enforcement officials.

AI also can augment acoustic and visual identifiers. Advanced acoustics are used to discover the number and type of guns at a crime scene using scene geometry, audio recordings and the audio signature of different weapons. AI is also being used to improve forensic DNA testing. Poor quality DNA limits the ability of traditional analysis to tie individuals to crimes. Here a hybrid approach using traditional DNA analysis in combination with AI shows promise in linking DNA profiles to individuals.[17]

Supporting the criminal justice system is another application of AI. In the late 2010s, AI algorithms were developed to assist judges in setting bail for people awaiting trial. The decision tree for permitting bail is a function of the flight risk of the individual. If the risk is low a minimal bail is set. The higher the risk the large the bail payment. If the risk is too great bail is refused and the person awaits trial in jail. Those who do not have the ability to pay the bail are also incarcerated. A concern with this process is the inconsistent nature of the decision making. A study of 100,000 judicial decisions found some judges released upward of 90 percent of the defendants while the number was only 50 percent for other judges.[18] Moreover there is evidence that judges decisions are subject to bias and random considerations such as the time of day they are making the decision. Many judges rely on the Arnold Foundation's Public Safety Assessment (PSA), a nine factor scoring intended to measure the likelihood of a defendant appearing for their court date.

The PSA approach has been criticized for "a one size fits all" approach. To address this AI can be used to complement PSA by looking at a comprehensive set of data of thousands of pretrial detention rulings and focus on the specific defendant and environment. Quality decisions from these algorithms is predicated on unbiased data. A system developed at Stanford Computational Policy Lab addresses this concern by making certain "the predicted risk scores are the same regardless of race."[19] AI systems are in theory less biased, but may result in similar types of errors as the systems are trained on historical data on risk level, much of which is impacted by human bias.

Learning from the AI Leaders

AI is being adopted by governments around the globe. Where are the best geographies to learn from? Oxford Insights 2022 Government AI Readiness ranks the US, Singapore, UK, Finland and

[16] https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/
[17] https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs
[18] https://engineering.stanford.edu/magazine/article/can-ai-help-judges-make-bail-system-fairer-and-safer
[19] https://engineering.stanford.edu/magazine/article/can-ai-help-judges-make-bail-system-fairer-and-safer

Canada as the most prepared to adopt AI.[20] The rankings are based on three essential ingredients for developing and using AI in government. The first ingredient government readiness. This includes is having a vision and strategy for AI including clear governance and ethical frameworks as well as the digital expertise to implement the strategy. Agility is also required to enable government to adjust rapidly as technology and use cases evolve.

The second requirement is a robust technology sector that provides human capital, innovation and maturity to offer governments use case applications that create public value. AI research and development plus entrepreneurship are necessary for a thriving AI ecosystem.

The third requirement is data and associated infrastructure. AI is driven by data. Unless data are comprehensive and representative as well as readily available, AI will not reach its potential.

The US scores very high on a strong private AI sector with extensive R&D investment, innovation and off the shelf products available for government use. It scores lower than others on the portion of the population that has strong digital expertise. Singapore's strength is its government attitude embracing innovation and public servant skills to adopt AI. Singapore's 2019 National AI Strategy provides a comprehensive blueprint for government problem solving through AI. The Readiness report also identifies Singapore's need for greater development of governance and ethics foundation.

AI – The Concerns

AI not without controversy. Headlines including "Amazon scraps secret AI recruiting tool that showed bias against women"[21] and "Disinformation Researcher Raise Alarms About AI Chatbots"[22] highlight the risks and concerns with AI. The risks associated with AI abound. PwC segments the risks into national level risks and application level risks.[23] At the national level the risks include societal risks such as disinformation, excessive surveillance and even autonomous weapons. Economic risk such as job impacts, inequality and expanding the wealth divide also sit at the national level. The overarching concern is that of ethics; are the specific uses of AI consistent with the values and norms of society?

The application level risks begin with basic security of the algorithm and preventing cyber attacks. There are also control risks preventing a rogue placement of an AI app in less secure systems. Moreover, as the algorithms self-learn and adapt there is less human control and oversights of how decision making takes place. This is the so-called black box problem: not even the original programmer knows the exact piece(s) of data the computer is basing its decision

[20] https://www.oxfordinsights.com/government-ai-readiness-index-2022
[21] Reuters, Jeffrey Dastin 2018
[22] New York Times, Hsu and Thompson, 2/8/23
[23] https://hbr.org/sponsored/2021/12/how-organizations-can-mitigate-the-risks-of-ai

on. The most direct risk at the application level is that the algorithm generates poor and/or biased decisions.

Another risk is "deep fakes" - modification of data and images that generate credible synthetic outcomes that did not happen. Think photoshop on steroids. Deep fakes abound on social media, where users, more or less harmlessly, make jokes or perform dances with the face of a celebrity or politician. But even in these cases, it is easy to see the potential danger. Audio technology was used to dupe a corporation into sending 200,000 British Pounds to a foreign bank account by impersonating the voice of its CEO.[24] The impacts of disinformation range from undermining trust in corporations and government to reshaping social norms.

AI is routinely used in the US to facilitate decision making in public safety and criminal cases as detailed above; however, these applications also have risks. There is strong evidence that predictive policing algorithms are biased. Many of the algorithms are based on arrest data. But the chances of being arrested if you are black are twice that of whites. Department of Justice data shows that "a black person is five times as likely to be stopped without just cause as a white person."[25] The data associated with arrests socioeconomics, education, zip codes are used to predict who will offend in the future. This information in turn is used to intensively police those environments leading to more arrests and a vicious cycle is perpetuated and strengthened by the algorithms. Ironically, race is not used in the algorithms, but the other attributes are surrogates for race.

There are two additional concerns. First, the data used to train the algorithm may not be comprehensive and sourced in the same geography, society, or culture it is used. Some algorithms used in the US have been trained on data from Europe. Second, conviction data are more appropriate for proactive policing than arrest data as conviction data represents crime not the suspicion of crime. Given the bias in arrests as described above, using data from arrests rather than convictions is problematic. Why focus on arrest data? It is often readily available for police departments.[26]

This is not just a US concern. In the UK the Centre for Data Ethics and Innovation "suggested that identifying certain areas as hot spots primes officers to expect trouble when no patrol, making them more likely to stop or arrest people because of prejudice rather than need."[27]

The bias concerns also extend to setting bail. The intended benefits of AI support in judicial bail decisions detailed above have been called into question. In Massachusetts a commission investigating the need for bail reform found "A systematic implementation of a risk assessment

[24] 2022: The Year of AI Hopes and Horrors, Cindy Gordon, Forbes 12/30/22
[25] https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/
[26] https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/
[27] https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/

tool in Massachusetts is not likely to lead to a drastic improvement in bail decision at this time. Judges appear to be making good determinations based on comparatively high appearance and low pretrial detention rates…The drawbacks of implementing a currently available risk assessment too would likely outweigh any incremental improvement in bail decisions."[28]

Many jurisdictions have opted for AI to inform judges. Here the concerns center on bias in the algorithms. More than two dozen academics including Berkman Klein Directors Martha Minow and Jonathan Zittrain wrote an open letter about the technical flaws of pretrial risk assessment tools. They argued: "Actuarial pretrial risk assessments suffer from serious technical flaws that undermine their accuracy, validity, and effectiveness. They do not accurately measure the risks that judges are required by law to consider. When predicting flight and danger, many tools use inexact and overly broad definitions of those risks…To generate predictions, risk assessments rely on deeply flawed data, such as historical records of arrests, charges, convictions, and sentences. This data is neither a reliable nor a neutral measure of underlying criminal activity." [29]

Similar concerns regarding need for and use of AI have also been raised in other use cases, as public officials weigh the benefits of greater efficiency with the potential costs of inequitable or opaque decision making.

The AI – Jobs Connection

AI is projected to deliver $13 trillion in global economic activity by 2030.[30] This magnitude of change will disrupt many sectors. Chatbots are eliminating customer service jobs. Robots are replacing manual labor. Automated vehicles threaten the jobs of drivers. Mckinsey & Company estimate that 70 percent of companies will adopt AI in their organization by 2030. Almost all organizations will see impacts on labor. This growth represents a dual edged sword. MIT's research effort on the impacts of AI on work of the future concludes that: "recent fears about AI leading to mass unemployment are unlikely to be realized. Instead, we believe that—like all previous labor-saving technologies—AI will enable new industries to emerge, creating more new jobs than are lost to the technology. But we see a significant need for governments and other parts of society to help smooth this transition, especially for the individuals whose old jobs are disrupted and who cannot easily find new ones."[31] The Biden Administration has argued for AI that augments rather than replaces workers.[32]

Policymakers must address fears of significant job displacement and address the those whose jobs will be lost. Several options can mitigate the negative impacts. The long-term approach is to orient education in schools and universities to prepare future workers to be AI enabled. The

[28] https://d279m997dpfwgl.cloudfront.net/wp/2020/01/0102_bail-reform-report.pdf
[29] https://cyber.harvard.edu/story/2019-07/technical-flaws-pretrial-risk-assessments-raise-grave-concerns
[30] https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy
[31] https://workofthefuture.mit.edu/research-post/artificial-intelligence-and-the-future-of-work/
[32] https://www.whitehouse.gov/wp-content/uploads/2022/12/TTC-EC-CEA-AI-Report-12052022-1.pdf

short-term solutions center on re-training which requires the combined efforts of employers, government, and worker organizations.

Addressing the Limitations of AI – Guiding Principles

The criticisms of AI are fact-based and material. However, AI is here and its reach will only increase in the coming years. Therefore, policy makers must walk the fine line of allowing AI applications that create private and public value while regulating the technology to avoid bias and discrimination. Creating value without bias and discrimination is refer to as "responsible AI."

The starting point for responsible AI is establishing guiding principles for the development and use of AI. The OECD established a set of principles for AI implementation. The OECD's value-based principles include fairness, transparency and explainability security and safety and accountability. Their recommendations "aim to foster innovation and trust in AI by promoting responsible stewardship of trustworthy AI while ensuring respect for human rights and democratic values."[33] Their recommendations are intended to work in concert with other regulations covering data privacy and digital security. The OECD also provides guidance for policy makers recommending they foster an AI ecosystem, build human capacity, address labor market disruptions and build trustworthiness at the international level.

The US National Artificial Intelligence Initiative Office suggests improving AI trustworthiness "requires a multifaceted approach, including R&D investments addressing key technical challenges, development of metrics, standards, and assessment tools to measure and evaluate AI trustworthiness, engagement in the development of AI technical standards, governance approaches for the use of AI in the public and private sectors…"[34] The Biden Administration has provided direction for federal agencies in achieving trustworthiness in the government's use of AI. This guidance recommends: (1) it should be performance driven where the benefits outweigh the costs; (2) the use cases should be directly tied to the training data and the data be reliable and unbiased; (3) the results should be understandable to subject matter experts, well-documented and traceable; and (4) the use should also be monitored over time. [35]

New Zealand provides a helpful model for addressing the challenge of building trust. The country embarked on an "AI national conversation" intended to increase understanding of AI in the public, encourage engagement in AI-related education and jobs, and involve citizens in conversations on the use of AI by the government.

[33] https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449
[34] https://www.ai.gov/strategic-pillars/advancing-trustworthy-ai/
[35] https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government

Operationalizing the Guiding Principles

AI Ethics Oversight boards or committees are one tool for building trustworthy AI applications in government and the private sectors. An AI ethics committee serves as the organization's watch dog to ensure that the inputs, programming and outputs of AI systems developed in the organization or purchased through third parties systematically and comprehensively meet the organization's core values. Staffing of the board should include subject matter and ethics experts, regulatory lawyers and those responsible for the organization's strategy. The team should also include AI bias scouts who can identify vulnerabilities early on in the development stage.[36] Some companies have also deployed AI "champions" to serve as contact points between the board and regular employees.[37] Decision making authority is key. If the board says an algorithm must be changed…it must be changed.

Northeastern University has created an AI ethics advisory board to support organizations that do not have the capacity to provide their own AI oversight.[38] The board has more than 40 members drawn from multiple disciplines and sectors. It is also intentionally diverse. The role of the board is strictly advisory: "The main goal of the board is to have the opportunity to [ask] the right questions and get the right answers. And then they're on their own."[39]

Another tool to operationalize the principles and recommendations described above is the AI audit. Audits provide an independent assessment of the AI algorithm. Does it do what it is intended to do? Are the data unbiased? Are the risks mitigated? AI auditor's affirmative answers to these questions are trust builders. The audit process is non-trivial. "AI systems are not simply a few lines of code, but complex sociotechnical systems consisting of a mixture of technical choices and social practices."[40] The challenge is sufficiently complex that Stanford University's Human-Centered Artificial Intelligence group ran a competition to find enhanced audit solutions.

The current state of the art focuses on evaluating the four stages of the AI lifecycle – design, development, deployment and monitoring. In design make sure the goals, context and assumptions are well defined. They should be pressure tested against the organization's values and norms. The development portion of the audit centers in the technical aspects of the model. This is also where the data a reviewed to ensure they are comprehensive, accurate, unbiased and relevant to the context. During deployment the audit evaluates if the intended goals are achieved and if anomalies occur, they are researched and resolved. After deployment regular audits function to continually monitor the AI decisions for risks and bias.[41]

[36] https://hbr.org/2022/07/why-you-need-an-ai-ethics-committee
[37] https://www.weforum.org/whitepapers/responsible-use-of-technology-the-microsoft-case-study/
[38] https://www.techtarget.com/searchenterpriseai/feature/New-AI-ethics-advisory-board-will-deal-with-challenges
[39] https://www.techtarget.com/searchenterpriseai/feature/New-AI-ethics-advisory-board-will-deal-with-challenges
[40] https://hai.stanford.edu/news/stanford-launches-ai-audit-challenge
[41] https://www.auditboard.com/blog/ai-auditing-frameworks/

The US Government Accountability Office issued audit guidance for federal agencies in 2021. The report provides a checklist and details for ensuring the integrity of governance, data, performance and monitoring. In the high risk data domain, the GAO recommends documentation of data sources, testing the reliability of the data, assessing the variables used in the model and assessing the use of augmented (computer generated) data. The recommendations for performance include well defined performance metrics, component and system level testing as well as whether the outputs are appropriate for the given context.[42]

The audit solution is great in concept but less robust in practice. The concern is that audits give the imprimatur of trustworthiness but with limited standards to guide the audit and auditors "audit-washing" could result. The German Marshall Fund describes the concern: "The risk is significant that inadequate audits will obscure problems with algorithmic systems and create a permission structure around poorly designed or implemented AI. A poorly designed or executed audit is at best meaningless and at worst even excuses harms that the audits claim to mitigate. Inadequate audits or those without clear standards provide false assurance of compliance with norms and laws, "audit washing" problematic or illegal practices."[43] As mentioned above, they are also immensely resource intensive and may be difficult for smaller, new or companies to conduct.

To address this issue, ISACA is a global community with a focus on increasing trustworthiness of technology in general and AI in particular offers an AI Fundamentals Certificate. It covers (1) AI principles, concepts and uses; (2) risks and ethical requirements; and (3) essential software and algorithms for AI applications and possibilities.[44] They also published a manual for auditing artificial intelligence.[45] This is a start but only a start. For audits to generate meaningful trust they must be standardized, certified and enforced. The financial auditing structure is a helpful analogy.

The bottom Line

AI is part of our world and will increase in importance in the years to come. AI has the potential to improve the efficiency and effectiveness of the government. Current and future opportunities abound. However, there are risks that need to be understood and mitigated. Frameworks and tools are being developed to maximize the net value of AI.

Tapping this value is the responsibility of each manager in each organization. The path forward consists of the following five actions:

1. Define AI principles
2. Identify AI use cases

[42] https://www.gao.gov/assets/gao-21-519sp.pdf
[43] https://www.gmfus.org/news/ai-audit-washing-and-accountability
[44] https://www.isaca.org/credentialing/artificial-intelligence-fundamentals-certificate
[45] https://store.isaca.org/s/store#/store/browse/detail/a2S4w000004KoGpEAK

3. Confirm value with cost benefit analysis
4. Plan for design, development, deployment, monitoring
5. Audit throughout the life cycle
6. Learn and identify the next round of use cases