From Ethics to Operations: Current Federal AI Policy

Advanced Technology Academic Research Center
October 4, 2021

Table of Contents

1 Executive Summary

There are currently dozens of separate AI ethics, policy, and technical working groups scattered among various federal departments and agencies, spanning the defense, civil, and legislative spheres. Each of these groups is pursuing important goals of defining policies related to AI within their specific charters. While a few overall governance structures for AI policy have begun, we are concerned that the resulting policies may be incomplete, inconsistent, or incompatible with each other.

Our approach to mitigating this risk is to provide this report, comprised of

- a framework for identifying all categories of policy related to AI,
- a review of current AI policy, legislative, and regulatory activities,
- an assessment of the current federal AI policy environment, and
- recommendations for using the framework to promote a comprehensive, consistent, and accurate federal AI policy environment

This report deliberately does not create, suggest, or recommend specific AI policy.

Instead, we provide a general framework and an assessment of the current state of federal government AI policy. The framework is a structure of AI policy categories and the mapping of those policy categories to the federal government organizations that have developed related policies. The policy structure we provide is an ontology, representing the logical relationships between various policy categories. By mapping the policy formulation activity of federal organizations to the ontology of policy categories, we provide insight into the current patterns of policy formulation. Our findings highlight areas of policy that may currently be under-addressed. This insight informs a set of policy development recommendations that conclude this report.

Our intent in providing this AI Policy Assessment (AIPA) is to support senior government leaders in:

a) ensuring that the resulting policies are aligned to the **values and priorities of the American people**, are mutually **consistent**, and collectively **complete**, and

b) ensuring that all AI policies have **accountable organizations** responsible for the policy formulation and enforcement

2 Evolving Needs for National AI Policies

The emergence of AI as a driving force in technology, economics, philosophy, and culture has not escaped the U.S. government's attention. Dramatic demonstrations of AI capabilities range from self-driving cars to autonomously coordinated drone flights to Alexa and systems that can assess MRI scans better than humans. The economic impact of AI can be quickly assessed by reviewing the market capitalization of relatively new technology giants Google, Facebook, and Amazon. In addition to these demonstrated impacts, the federal government has witnessed the

explosion of investment in AI (from just under $9B in 2015[1] to over $50B in 2020 and expected to grow to over $110B by 2024[2]), and in response, has released four Executive Orders, created dozens of commissions, studies, and policy papers, developed hundreds of AI tools, and purchased billions of dollars of AI technologies and services. Even with these billions of dollars invested over the past decade and an increasing impact on our culture and society, AI technologies are far from achieving their potential. AI can be defined in terms of the systems that demonstrate AI or in terms of an engineering discipline creates such systems. In both definitions, the systems developed mimic human perception, reasoning, creativity, and emotional behavior. Anticipating the impact that such systems would have on our way of life is challenging. One way to prepare for that future is to take the time we have to carefully and deliberately develop the policy environment within which the U.S. government will create, procure, and regulate these technologies.

2.1 Current U.S. Federal Government Roles & Policy

2.1.1 Legislative Action Defining Federal AI Policy

Two primary pieces of legislation document the current federal AI policy:

1. The National AI Initiative Act (AI-IA) of 2020 (Division E of the National Defense Authorization Act for 2021) directs the Executive branch to stand up a National AI Initiative Office (NAIIO), supported by a Select Committee on AI (SCAI), an Inter-Agency Working Group on AI (AI R&D IWG), and a National AI Advisory Committee (NAIAC). These committees are chartered to promote the effective and efficient research and development of AI systems throughout the federal government, support effective and secure data availability, assess issues related to the AI workforce, and enable broad collaboration among federal AI stakeholders. The AI-IA directs the National Institute for Standards and Technology (NIST) to support the development of technical standards, terminology, and shared understanding across the federal government, industry, and academia. The AI-IA directs the National Science Foundation (NSF) to convene and support collaboration among academic research institutions and to establish up to five multi-disciplinary AI research consortia. The final section of the AI-IA directs the Department of Energy (DOE) to promote AI system development through its own research as well as funding up to five AI research centers as well as providing access to computational facilities in support of AI research and development.
2. The AI in Government Act of 2020 (AIGA) (Division U, Title I of the Consolidated Appropriation Act of 2020) establishes an AI Center of Excellence (COE) within the General Services Administration (GSA) to promote efforts to pursue AI systems within the federal government by collaborating with federal agencies, collecting and disseminating best practices, generating policy statements and guidance for agencies developing and deploying AI systems, and advising OSTP on AI technical and policy matters. The AIGA also directs OPM to update the skillset requirements and

[1] https://cset.georgetown.edu/publication/tracking-ai-investment/
[2] https://www.idc.com/getdoc.jsp?containerId=prUS46794720

occupational series relevant for AI system research and development, and to develop two-year and five-year federal workforce staffing plans.

2.1.2 Whole of Government AI Policy

There are currently six agencies and independent organizations developing AI technologies or policies on behalf of the entire federal government.

1. The White House Office of Science and Technology Policy (OSTP) provides the Executive Office of the President with advice on the scientific, engineering, and technological aspects of the economy, national security, homeland security, and public health. OSTP has supported the release of several Executive Orders defining overall policy goals for AI in the federal government. Following passage of the National Artificial Intelligence Initiative Act (AI-IA), OSTP established the National Artificial Intelligence Initiative Office (NAIIO), which has overall responsibility for coordinating a "whole of government" approach to developing, using, and regulating AI. The NAIIO includes a number of select committees focused on ML, research and development, law enforcement, and resource development and utilization. The NAIIO also includes the National Artificial Intelligence Advisory Committee (NAIAC) which advises the President and OSTP on a broad range of federal AI issues, including the state of commercial advances in AI, impact on the US workforce, societal impacts of AI, and progress on implementing the AI-IA.
2. The General Services Administration (GSA) is a leader across the federal Government in bringing numerous cutting-edge technologies to Agencies, including policies and governance practices. GSA launched a government-wide Artificial Intelligence Community of Practice in 2019, which brings together federal employees who are active in, or interested in, AI policy technology, standards, and programs. GSA's Office of Government-wide Policy (OGP) has developed a new pilot using AI for Prediction of Regulatory Compliance, known as the Solicitation Review Tool (SRT). The General Services Administration's AI Center of Excellence provides services to agencies across the federal government to promote and improve the adoption of AI. While the services they provide are primarily technical, all of the GSA IT Modernization Centers of Excellence follow the Guide to AI Ethics and the Data Ethics Framework developed in the COE.
3. The National Security Commission on Artificial Intelligence (NSCAI)was chartered and funded by the 2019 national defense authorization and delivered its final report in March 2021, comprised of:

   Part I "Defending America in the AI Era" (Chapters 1-8), outlines what the United States must do to defend against the spectrum of AI-related threats from state and non-state actors and recommends how the U.S. government can responsibly use AI technologies to protect the American people and our interests.

   Part II "Winning the Technology Competition" (Chapters 9-16), outlines AI's role in a broader technology competition. Each chapter addresses a critical element of the

competition and recommends actions the government must take to promote AI innovation to improve national competitiveness and protect critical U.S. advantages.

4. The Department of Commerce (DOC) is addressing multiple aspects of AI policy, regulation, and implementation in its various Bureaus:

   The National Institute for Standards and Technology (NIST) is pursuing efforts to help establish standards for how to design and implement AI systems, including ethics and trustworthiness.
   The Federal Trade Commission (FTC) is chartered to both protect US consumers and promote competition in the US economy. FTC has been directly involved in exploring issues related to facial recognition technology, competition and consumer protection with respect to AI, and the application of consumer protection legislation in the context of AI products and services. FTC has issued guidance to business on how best to position, develop, and deliver AI products and services to avoid conflict with existing and potential future legislation.

   The International Trade Association (ITA) is chartered to implement various trade controls included in recent appropriations that address perceived risks of exporting or importing potentially dangerous AI technologies.
5. The Office of Management and Budget (OMB) has issued final guidance to federal agencies on when and how to regulate the private sector use of AI. This document presents a broad perspective on AI oversight and provides a set of guiding principles intended to navigate the complex and dynamic landscape of trade-offs required in developing AI regulations.
6. Congress includes both a Senate and House Artificial Intelligence Caucus and is generating legislative language through reports and individual bills. The perspectives informing much of this work include
   a. the expectation that AI can benefit every major sector of the national economy,
   b. that the technology also presents a number of risks
   c. to ensure AI is responsibly used, and "trustworthy"
   d. there is potential national security as well as economic risk with respect to competition over AI with China
   e. concern that the US may be falling behind in the development of AI, or may suffer institutional disadvantages relative to autocratic regimes that have fewer restrictions on data collection and exploitation
   f. there is likely a need for long-term investment in basic research and infrastructure to protect the economic and national security of the United States

2.1.3 Agency-Specific AI Policy

There are currently 10 agencies and organizations either developing AI policy or technologies that are focused on their own implementation rather than cross-government. This list does not include all of the AI implementations that are currently being developed across the federal government but includes those that are aimed at agency-wide impact.

1. The Joint Artificial Intelligence Center (JAIC) is the Department of Defense's (DoD) AI Center of Excellence and supports the transformation of U.S. Joint warfighting and departmental processes through the integration of AI and enables the empowerment and unification of bottom-up AI development by innovators across the Defense Department. The JAIC leads the assessment for Department AI Transformation, and leads strategic planning for the JAIC itself. The JAIC has developed a Strategy, Guiding Tenants, and is Evolving partnerships with industry, academia, allies, and partners and has developed the DOD AI Ethics framework.
2. The Office of the Director for National Intelligence (ODNI) has developed an "Artificial Intelligence Ethics Framework for the Intelligence Community", providing AI principles and a supporting framework that set a foundation for how and when members of the DOD and IC should use, develop, and procure AI applications. ODNI developed its six principles to be consistent with those of the Department of Defense.
3. The Veterans Administration has established the National AI Institute, chartered to develop policy, approaches, and standards for developing AI solutions for the VA, as well as to incentivize, organize, and lead innovation and prototyping exercises to promote AI systems development to further the VA mission.
4. The National Science Foundation (NSF) is required (by the 2021 NDAA) to submit a report on ethics statements. NSF's Directorates for Computer and Information Science and Engineering (CISE) and Social, Behavioral and Economic Sciences (SBE) together with the Partnership on AI (PAI) are jointly supporting Early-concept Grants for Exploratory Research (EAGERs) to understand the social challenges arising from AI technology and enable scientific contributions to overcome them. NSF's CISE directorate invites researchers to submit proposals to its core programs that contribute to discovery in research and practice related to fairness, ethics, accountability, and transparency in computer and information science and engineering, including AI. The NSF Program on Fairness in Artificial Intelligence in Collaboration with Amazon (FAI) provides a significant opportunity to transform research across all areas of science and engineering, including AI:
    a. Advancing Fairness in AI with Human-Algorithm Collaborations
    b. Addressing the 3D Challenges for Data-Driven Fairness: Deficiency, Dynamics, and Disagreement
    c. Towards Fairness in Deep Neural Networks with Learning Interpretation
    d. Towards a Computational Foundation for Fair Network Learning
    e. Fairness-Aware Algorithms for Network Analysis
    f. Identifying, Measuring, and Mitigating Fairness Issues in AI
    g. FairGame: An Audit-Driven Game Theoretic Framework for Development and Certification of Fair AI
    h. Building a Fair Recommender System for Foster Care Services within the Constraints of a Sociotechnical System

    i. Quantifying Direct and Indirect Consequences of Racial Disparities in Outcomes Following Cardiac Surgery
    j. Auditing and Ensuring Fairness in Hard-to-Identify Settings
5. The Department of Commerce is addressing multiple aspects of AI policy, regulation and implementation in its various Bureaus:
    a. The National Oceanic and Atmospheric Administration (NOAA) is standing up an AI Center, pursuant to the 2021 budget, to enable broad application of AI technologies across its mission areas – weather and climate, coastal fisheries management, oceanic research, planetary observation and data collection, and fundamental research. NOAA's AI Center is also chartered to improve the availability, quality, and quantity of its data stores to commercial, educational, and research partners.
    b. The United States Patent and Trademark Office (PTO) is pursuing AI projects to support the Informal Adjudication process of patent application processing
6. The Securities and Exchange Commission (SEC) is pursuing AI for Regulatory Enforcement. SEC has a suite of algorithmic tools to identify violators of securities laws. For example, to detect fraud in accounting and financial reporting, the agency developed the Corporate Issuer Risk Assessment, which has a dashboard of about 200 metrics that can find anomalies in the financial reporting of more than 7,000 corporate issuers of securities. An ML tool identifies filers who might be engaging in suspicious activities by using historical data to predict possible misconduct.
7. The Department of Energy (DOE) has initiated the Artificial Intelligence Technology Office (AITO) which has been collaborating with the JAIC and other agencies' development of ethical AI principles. The DOE intends to draft its own set of ethical AI principles that will regulate how it develops, deploys and shares the technology.
8. The National Aeronautics and Space Administration (NASA) has been an AI/ML user and innovator for decades, and in June of 2021 released a Framework for AI Ethics.
9. The Social Security Administration (SSA) is pursuing AI tools for Formal Adjudication.
10. The Department of Health and Human Services (HHS) is pursuing AI tools in various ways. The Food and Drug Administration (FDA) is pursuing AI tools to support Regulatory Analysis. An AI Pilot Project catalyzed the GSA to initiate an effort to help Federal agencies in implementing AI to perform reviews of regulations and the enhance rule-making process.
11. The United States Postal Service (USPS) has AI efforts to support Autonomous Vehicles for Mail Delivery.
12. The Defense Innovation Board (DIB) provides the Secretary of Defense, Deputy Secretary of Defense, and other senior leaders across the Department with independent advice and recommendations on innovative means to address future challenges. The DIB released an AI principles report and supporting document in 2019.

2.2 Advancing Technologies

Driven by commercial success and the potential impact these technologies may have across industries and society as a whole, billions of dollars are being invested in every stage of the AI "pipeline". All supporting components of a successful AI ecosystem, from fundamental research in mathematics and computer hardware, to commercial ventures addressing niche and consumer markets, to education, are receiving significant investment.

In many ways, developing an AI program is no different from developing any software application. The motivation is generally a problem to be solved or a question to answer. Requirements are collected and managed throughout the development process. Today's AI systems are almost exclusively ML-based, and therefore require large amounts of carefully curated data. The type and detailed structure of the model to be used is implemented in software, and the data is run through the model until the desired accuracy is achieved.

2.2.1 An Open-Source Culture

AI research and system development is characterized by significant openness and collaboration among researchers and commercial organizations. The major AI software firms have published open-source development platforms and libraries and continue to update and publish as they are refined. Model design, data set selection and curation, testing methods, and evaluation parameters are all commonly shared among AI developers. An example of this sharing is the Cross Industry Process for Data Mining (CRISP-DM)[3], an open standard, six phase process for identifying, collecting, and using data from multiple sources. This open source approach is common, up to the point where a company implements the tool in a product or service offering, at which time the information is kept proprietary.

2.2.2 MLOps and Automated Testing

DevOps is a relatively new approach to software development and management currently being adopted throughout industry and the federal government. The goal of DevOps is to increase the speed of software development and deployment, moving to a model of continuous software release. MLOps refers to the inclusion of techniques, methods, and tools to enable continuous ML system development and deployment. MLOps requires additional steps and resources, as compared to DevOps, because of the need to select, access, curate, and apply training data to the ML models after they are implemented in software. A number of commercial firms seek to address these requirements, by offering tools intended to automate the data cycle of ML system development.

Similar to automating the data cycle, MLOps seeks to automate the test and validation phase of ML system development. As might be expected, many approaches to applying ML tools to these tasks are being evaluated, including: scanning the software code line-by-line to identify grammar or naming errors; evaluating the logic to identify pointer exceptions, memory leaks, or other run-time errors; simulating different operating environments to validate that the model executes correctly in them; evaluating changes in training data sets to determine if re-training

[3] https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

will be beneficial; evaluating the visual representation of the graphical user interface to determine its quality of user experience and whether it complies with applicable regulations.

2.2.3 Technical Approaches to Ethical AI, Transparent, Fair, and Explainable AI

There is general consensus that our country's benefit from AI technologies will depend on our ability to create and deploy AI systems that are understood to be ethical and trustworthy. The challenges are in stablishing a consensus definition of what ethical and trustworthy AI is, how to identify it, and how to prove it in practice. One approach currently being pursued is to define concepts such as *ethical*, *transparent*, *fair*, and *explainable* algorithmically. Examples include the Quantitative Input Influence model[4], an approach to modeling human trust in AI systems[5], and multiple attempts to define fairness algorithmically[6].

Being able to prove an algorithm fair, trustworthy, explainable, and transparent would be beneficial for researchers and application developers but would not necessarily resolve the public perception issue for AI systems. There is good reason to believe that a negative public perception of AI is building[7]. The history of nuclear power technology in the United States may be instructive. Although demonstrably safer, more efficient, and initially (before the accident at Three Mile Island) more economical, nuclear power has suffered a significantly negative public perception for three decades. Only recently has public perception of nuclear power in the U.S. been roughly equal between supporters and those opposed, possibly driven by concerns over global climate change. The extent to which this negative perception has inhibited construction of new nuclear power plants is not quantitively known but may have played a part. AI technology may also be delayed, avoided, or resisted from a similar, wide-spread public concern over its safety, ethics, or fairness.

2.2.4 Explainable AI

This is the "black box problem" of AI is created by the standard approach to developing ML models, which results in a highly complex set of data matrices, with hundreds or thousands of connected nodes. Each node includes a "weight" or value. As the ML is "trained", these weights are adjusted to improve the model's performance. This dynamic, complex system is not designed to, and is not generally able to, provide an easily understood explanation of how it arrived at any given result. Significant research is underway, including through the DARPA XIA program[8], to develop AI systems that are explainable. Much of this effort is to understand what types of tradeoffs, in accuracy, complexity, and reliability, might be necessary to provide varying levels of explainability. Approaches include:

- constraining the structure of the ML model to correspond to human-relevant constructs (e.g., predetermining the components of the ML model for recognizing images of a cat by

[4] https://www.andrew.cmu.edu/user/danupam/datta-sen-zick-oakland16.pdf
[5] https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8332-draft.pdf
[6] https://www.annualreviews.org/doi/pdf/10.1146/annurev-statistics-042720-125902
[7] https://towardsdatascience.com/people-dont-trust-ai-we-need-to-change-that-d1de5a4a0021
[8] https://www.darpa.mil/program/explainable-artificial-intelligence

creating a set of models that correspond to ears, body shape, tails, fur patterns, and head/face features),
- applying statistical inference techniques, such as Generalized Additive Models, to the model output to determine the features that led to the model's output; essentially reverse engineering the model,
- applying statistical behavior modeling to the ML model, using techniques such as Layer-Wise Relevance Propagation, to interrogate the ML model's changes in state as it generates output

2.2.5 Human - AI Collaboration

A significant amount of effort is focused on determining and enabling the most effective ways for humans to use AI. Depending on the context, "effective" may be defined with more of a focus on the user experience, for instance, in entertainment settings such as video gaming or immersive or interactive media. Another context may prioritize successfully achieving a goal, such as winning a game or contest. The "centaur" model of human – AI collaboration places the human in the loop as the key decision maker while the AI provides continuous decision support in the form of status updates, data analysis, predictions, and strategy assessment.

2.2.6 Edge AI

As computing devices continue to increase in speed as they shrink in size, and as our devices become continuously connected over faster networks, the advantages of providing AI capabilities at the "edge" (including mobile phones and devices, control hardware, and network endpoints such as wireless Network Access Points) are being explored. Being in closer proximity to the user and being trained on localized data sets may enable faster and more accurate response from the AI system. Having the model, data sets, and response history stored locally may also provide information security, system resiliency, and user privacy benefits.

2.2.7 Artificial General Intelligence (AGI)

AGI is a system of machine-based intelligence that demonstrates human-level common sense, context awareness, judgment, and creativity. Some definitions of AGI include further capacities for empathy, creativity, and consciousness. While some technology leaders such as Elon Musk and Bill Gates have expressed their fears of AGI, most researchers believe it is at least decades away, and many believe it is not achievable[9]. Nevertheless, AGI is widely seen as the end goal of AI.

To a greater extent than most technologies, the broad societal understanding of AGI is heavily influenced by works of fiction. The Terminator movie series, the Matrix movie series, the movies I Robot and AI, and dozens of other films feature an AGI that takes over the world and decides that humans are a threat. Fear of AGI is generally driven by the perspective that, if an AI system were able to become self-aware, it might "decide" to protect itself from being turned off or deleted. If it were able to use its "understanding" of networks, databases, and

[9] https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/

infrastructure, then it might become self-replicating, and may be able to increase its intelligence with each generation. At computer speed, this cycle of replication + intelligence gain might result in greater than human intelligence very shortly after AGI is achieved. Achieving this level of AGI is also known as the "singularity" – the point at which machine-based intelligence surpasses human intelligence. The behavior, capabilities, and risks of such a super-intelligence are sufficiently uncertain to support a wide variety of doomsday scenarios.

2.3 Emerging Policy Issues

Even as the various agencies, boards, and committees address AI policy needs as described above, advances in AI technology and broad economic and socio-political trends are creating near-term policy issues that have not yet been addressed. The policy goals outlined in EOs are generally strategic and are therefore applicable to new technologies and trends.

2.3.1 Emerging Technical Policy Areas

1. Integrated circuit (IC) manufacturing: the IC industry has entered a phase of apparent instability, resulting from the global IC manufacturing capacity of advanced chips being concentrated in only three companies (Intel, Taiwan Semiconductor Manufacturing, and Samsung Electronics[10]). As this consolidation took place over the past five years, the COVID-19 pandemic created multiple bottlenecks in the global supply chain for mid-tier ICs that add automation and advanced features to many consumer products (including cars, appliances, and consumer electronics). The resulting unreliability of supply of ICs presents a potential national security risk as well as an ongoing threat to AI system research and development.
2. Affective Computing (AC): research in AC is focused on creating systems that integrate awareness of a user's emotional state in their interaction with the user. This is done by creating haptic, visual, and auditory sensors (e.g., measuring how quickly or forcefully the keys of the user's keyboard are struck, capturing and analyzing the dilation of the user's pupils or skin temperature), integrating that input to some model of the user's emotional state, and adjusting the system's visual, haptic, or auditory output in a manner to improve the user's experience or the system's performance. Development of this technology or deployment at scale presents new policy issues for the federal government, both in the privacy concerns regarding data collection, but also in the creation, development, recording, and utilization of data and models intended to analyze and adjust users' emotional states. Affective systems present a new, unique technical capability to manipulate and exploit citizens' emotions, and therefore present a novel threat to privacy.
3. Neural Links: Neural links are electrical interfaces created by direct physical connection of electrical sensors to the brain of the user. The goal of this research is to bypass extremity-based interfaces and directly "read" and "write" information to the brain.

[10] https://www.bloomberg.com/graphics/2021-chip-production-why-hard-to-make-semiconductors/

Current research and development has demonstrated the ability to "read" a primate's mental instructions to move a cursor on a screen[11]. While the research and development in neural links is very early stage, the technology presents policy challenges of privacy as well as ethical testing and validation. Additional issues may arise with commercialization as the risks of a user being hacked become more clear – would the individual being hacked be responsible for their actions? What kinds of information could a hacker access through a neural link? What would the ramifications of dormant neural link hacks have on criminal justice, if the possibility exists that any accused person may have been the victim of a neural hack? While these concerns may appear speculative, we may find ourselves facing them in the next few years.

2.3.2 Emerging Non-technical Policy Areas

In a similar way that the internet's societal impact is still evolving, AI's societal impact will likely not be clear for decades. The potential benefits have been discussed above, as have been a number of risks. In addition to these technically focused impacts, AI is likely to interact with other societal trends in ways that are difficult to predict, but likely consequential.

1. Criminal Justice Reform: AI has the potential to both support criminal justice reform and worsen the inequities and shortcomings of our current system. If applied to analyze and better understand patterns of behavior and the cause – and – effect of current systems, AI could provide insight to improve policies, laws, and practices. If AI is deployed to automate and expand the reach of existing laws or practices, any of their shortcomings or problematic impacts will be amplified. This issue is made critical by a consistent pattern that ML tools display: since they are trained on large existing data sets, they perform poorly when analyzing under-represented populations[12]. They also "learn" historic bias because they are trained to replicate historic decision-making, as represented in curated training data sets. These shortcomings of ML models are driven by the mathematics of optimizing for predictive accuracy across the entire training data set. They also create a societal risk if we create systems with built-in bias and use them to make judgments about millions of US citizens.
2. Environmental Impact: As ML models increase in size and complexity, they require longer and more intensive computation to train. This computation requires increasing amounts of electricity, with resultant emissions of greenhouse gases. Training the GPT3 natural language model, for instance, reportedly emitted 552 metric tons of CO_2[13]. As industry and government move forward generating increasing numbers of large, dense neural networks, greenhouse gas emissions may rise dramatically. Unless mitigation measures are identified and implemented, AI is poised to be a significant driver of

[11] https://www.theverge.com/2021/4/8/22374749/elon-musk-neuralink-monkey-pong-brain-interface
[12] O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Penguin Books.
[13] https://arxiv.org/ftp/arxiv/papers/2104/2104.10350.pdf

electricity consumption and worsening levels of greenhouse gases over the next few decades.

3. Workforce Impact: The net impact of AI on the number of employment opportunities is of critical importance. As AI systems "learn" more skills, they are expected to displace many physical laborers, and to gradually displace knowledge workers[14]. The skills needed for effectively using or collaborating with AI systems are not yet known. As AI systems become more empathetic and emotionally aware (as in the AC discussion above), will they displace humans in positions currently thought to be "AI proof"? The workforce impact that began with manufacturing automation is likely to increase in scale and scope repeatedly over the next few decades. We will need an effective set of policies to mitigate those effects and to take full advantage of the benefits AI will bring to the workplace.
4. Economic Impact: If the workforce impacts described above are realized, we may approach an economy with only a minimal amount of manual labor and a dramatically reduced volume of what is currently considered knowledge work. We have almost no experience with such an economy, and haven't developed, much less validated, economic models or systems that can effectively and fairly allocate resources, set prices, inform markets, coordinate supply chains, and reward innovation, performance, or creativity in an economy based on ubiquitous AI.

3 Current Federal AI Policy – an Assessment

3.1 Purpose and Value

This paper has been developed as a refence and guideline for senior leaders in the federal government who, as stewards of enormous quantities of data on millions of Americans and their national resources, have responsibilities to ensure they are used for the collective benefit of all. As discussed elsewhere, technology is available to derive important knowledge as never before – and there is an historic opportunity for the government to use advanced analytics including AI to improve efficiency and empower innovation. But there is also great potential for misuse and for the benefits to be delivered unequally. This assessment and framework provide a set of principles and policy processes that will ensure the government protects, shares, and presents its data and insights in an ethical, transparent, and responsible manner. Given the elevation of the Office of Science and Technology Policy directorship to a Cabinet position, the nation's Chief Science Officer will have a seat at the table in Domestic Policy Committee meetings. This assessment and framework are intended to help that office to leverage this unique opportunity in American history for the good of all.

The potential benefits of coordinated policy result, in a word, from trust. Coordinated policy creates transparency, consistency, and accountability. If federal agencies contribute through

[14] https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Future%20of%20Organizations/AI%20automation%20and%20the%20future%20of%20work%20Ten%20things%20to%20solve%20for/MGI-Briefing-Note-AI-automation-and-the-future-of-work_June2018.ashx

governance structures to approve and adopt a set of principles and policy processes as presented in this framework, they can hold one another accountable and share their data accordingly. Approving a policy framework leads to its general publication. When "we the people" know the privacy and standards and other ethical principles to which our government has committed, we can hold them accountable. Additionally, a widely adopted set of principles and processes creates stability for economic investment of industry in supporting government programs. This fosters innovation and enables the identification and then promulgation of the very best ideas to leverage government data for good.

The risks of inconsistency in data use, and lack of trust, are sadly illustrated by several recent examples. Misuse of big data and the insights it can reveal enables political entities to target the spread of inflammatory and even false information to gain political advantage, undermining America's democratic principles. Unequal access to benefits of big data insights furthers racial and wealth inequality. Unethical marketing practices interfere with the open competition vital to our economy and stifle innovation. And enemies of our country seeking to exacerbate discord can target destructive messages. The power of targeting messages and telling people what they want to hear has been clearly demonstrated.

Absent transparency and trust, federal agencies have been reluctant to share data with one another, and the population has been reluctant to provide personal information to the government. The harms of distrust and lack of sharing range from the terribly inefficient to the tragic. Incomplete sharing of health record information between the Department of Defense and the Department of Veterans has resulted in frustration and sometimes even denial of benefits to veterans for far too long, and electronic health record adoption has only partially alleviated the problem. Ineffective and incomplete sharing between federal intelligence agencies contributed to failure to prevent the September 11th attacks. The people's reluctance to trust the federal government to establish a single national identifier is one reason our health care system is one of the world's most costly.

The advent of AI significantly increases both the potential benefits and the risks of government's use of large volumes of data. This framework is therefore focused on AI policy and processes for ensuring that ethical principles are applied and sharing standards are adopted. The value of adoption of the framework includes reduced risks of incomplete or inconsistent AI policies across the various federal agencies, improved accountability for organizations with AI policy responsibility, and improved communication among AI stakeholders regarding when, where, and how AI policies are being defined and enforced.

3.2 Approach

Acknowledging our lack of charter to determine policy, our approach does not include specific policy recommendations. Instead, our approach addresses three issues that we believe are critical to getting AI policy right, regardless of any specific policy considerations:

1. Constitutionality: any set of AI policies should first and foremost be designed and implemented to support the values and goals expressed in our constitution, as amended

2. Completeness: the federal government's AI policies should address all areas of AI policy that are necessary to ensure that AI is developed for the benefit of the nation's citizens
3. Consistency: the federal government's AI policies should define the nation's AI policy goals, without confusion or conflict, when enforced as a whole, regardless of which agency or organization develops and implements them

3.3 Structure of the AIPA

This combination of goals led us to define a matrix with one axis that indicates which agency or organization has developed (or recommended should develop and implement) AI policy categories and a second axis comprised of the policy categories.

The matrix can indicate a variety of information by populating its cells. We have prepared one matrix that shows the current federal government AI policy environment. This was developed by researching publicly available records of those agencies and organizations.

The AIPA does not include information on AI initiatives, only efforts to establish AI policies. The OECD currently manages an "AI Policy Observatory"[15]that indicates the U.S. currently has 47 AI initiatives. Not all of these initiatives, however, are intended to address AI policy. Those that address policy questions are included in the AIPA.

[15] https://oecd.ai/dashboards/policy-initiatives?conceptUris=http:%2F%2Fkim.oecd.org%2FTaxonomy%2FGeographicalAreas%23UnitedStates

3.3.1 Organization Axis

Executive Branch Organizations	White House	White House
	OMB	Office of Management and Budget
	DOD	Department of Defense
	JAIC (DOD)	Joint Artificial Intelligence Center
	DARPA (DOD)	Defense Advanced Research Projects Agency
	NIST (DOC)	National Institute of Standards and Technology
	ODNI	Office of the Director for National Intelligence
	DHS	Department of Homeland Security
	DOT	Department of Transportation
	NASA	National Aeronautics and Space Administration
	FDA	Food and Drug Administration
	OPM	Office of Personnel Management
	GSA	General Services Administration
Congress	House SS&T Comm	Committee of the House
	Armed Services Comm	Joint Committee
	Judiciary Comm	Joint Committee
	Senate Commerce Comm	Senate Committee
Boards and Councils	Partnership on AI (NSF)	Program on Fairness in Artificial Intelligence in Collaboration with Amazon – National Science Foundation
	CIO Council	Federal Chief Information Officers' Council
	AWPAB (DOC)	American Workforce Policy Advisory Board – Department of Commerce
	NSTC DPC NEC	National Institute of Standards and Technology – Department of Commerce
	NSCAI	National Security Commission on Artificial Intelligence
Standards Organi-zations	IEEE	Institute of Electrical and Electronics Engineers
	ISO	International Organization for Standardization

Figure 1 Organizations Developing AI Policy

Our research into current active federal government organizations in AI policy indicates the following organizations have developed wide-ranging policy on AI. This list is not intended to be comprehensive of all federal government organizations but focused on those that have produced AI policy documents. The list does not indicate all organizations that have developed AI strategies, only those that have created AI policies. We also have excluded documents that have presented only broad intent or goals, such as "promoting U.S. competitiveness in AI" or "protecting citizen's privacy rights" without further definition or policy. Our research included all three branches of government, but the judicial branch, being comprised of the Supreme Court, is silent on policy. We have also included independent organizations as well as non-governmental organizations that may influence either the definition or implementation of AI policy.

3.3.2 Policy Category Axis

The Policy Category Axis is intended to be a comprehensive catalog of all policy categories relevant to the federal development, acquisition, and use of AI technology. Rather than an unstructured list of categories, we have provided an ontology of policy categories, illustrating how some policies are logically and causally grouped together, and how some are more closely related to others. Our intent in developing an ontology is to provide a structure that supports both goals of completeness and consistency. The structure does not indicate any priority of policies but does indicate logical categories or causal dependencies.

The policy ontology does not include policies related to investing in AI, promoting AI technical development, or similar industrial policy areas. We consider these policy areas to be of general interest, but not specific to AI. Similarly, the policy ontology does not include a category of the

type “AI should be developed for the beneficial of humanity” or “AI should be developed to further human rights” or similar broad constructions. This is because such a category would only represent a collection of the policy categories included in the ontology, and so would not provide additional information. One lesson learned by multiple contributing authors to this work is that much time and effort can be spent in discussing and debating the merits of specific relationships between elements of an ontology, without improving the usefulness or effectiveness of the resulting work. Figure 2 provides the policy ontology with definitions for each element.

			Policy Category	Definition
AI Policy	Ethics	Characteristics	Bias	The AI system demonstrates an unexpected or inaccurate skewing of results affecting one or more group of cases
			Transparency	Providing full documentation of all phases of AI system development, training, testing, use, and impact
			Explainability	Ability of the AI system to explain to a layperson how a result was determined
			Representativeness	The AI system addresses representative cases and has been trained on representative data sets of interest
			Autonomy	The AI system complies with U.S. law, regulations, and policy when operating without human input
		Outcomes	Fairness	The AI system’s design and use protect the rights of affected persons
			Accountability	The AI system is designed and used with documented and enforced roles and responsibilities
			Confidentiality	The AI system’s design and use protect users’ information from access, alteration, or destruction
			Impact	The AI system’s design and use protect the rights of affected groups
			Safety	The AI system’s design and use do not decrease the overall safety of affected persons
			Privacy	The AI system’s design and use support the privacy rights of affected persons
			Discrimination	The AI system’s design and use protect the rights of affected persons and groups
			Trustworthiness	The AI system’s design and use justify public trust in its use
			Security	The AI system’s design and use protect it and its data from access, alteration, or destruction
	Operations	Development	Methodology	Standards, practices, procedures, and tools used in developing, testing, and using AI systems
			Testing	The design, data sets, criteria, and groups involved with ensuring the AI system’s performance
			Diversity	Ensuring that the teams involved in the AI system lifecycle reflect those affected by the system’s use
			Data Management	Standards, practices, procedures, and tools used to manage training, testing, usage, and impact evaluation of the AI system
			UX Design	Designing the overall user interaction with the AI system to support the users’ needs and expectations
			Security	Protecting the AI system from unauthorized access, alteration, or destruction
			Performance	Ensuring that the AI system satisfies all stakeholder requirements
			Workforce	Ensuring that the federal workforce is prepared to effectively create, deploy, and use AI systems
			Risk Management	Fully assessing the risk types, potential harms, and risk management options for AI systems
		Procurement	Requirements	Clearly stating meaningful and appropriate requirements for AI systems including Characteristics, Outcomes, and Development elements
			Impact/Safety	Specifying details for the AI system’s Impact and Safety requirements, including quantifiable terms and measurement and evaluation methods
			Selection	Providing and applying clear selection metrics and standards for AI systems
			Specifications	Providing quantifiable specifications and/or ranges for requirements for AI systems
		Collaboration	Representation	Ensuring that all affected groups are represented in collaboration activities for AI systems
			Participation	Effective engagement with AI systems stakeholder groups and appropriate federal agencies and organizations

Figure 3: AI Policy Ontology

3.3.3 Current State of U.S. federal Government AI Policy

Through a combination of Executive Orders (EOs) and reports released by the National Science and Technology Council (NSTC), previous administrations have provided AI policy guidance, priorities, and goals. Figure 4 indicates the current state of federal AI policy formulation, with references for the supporting documentation.

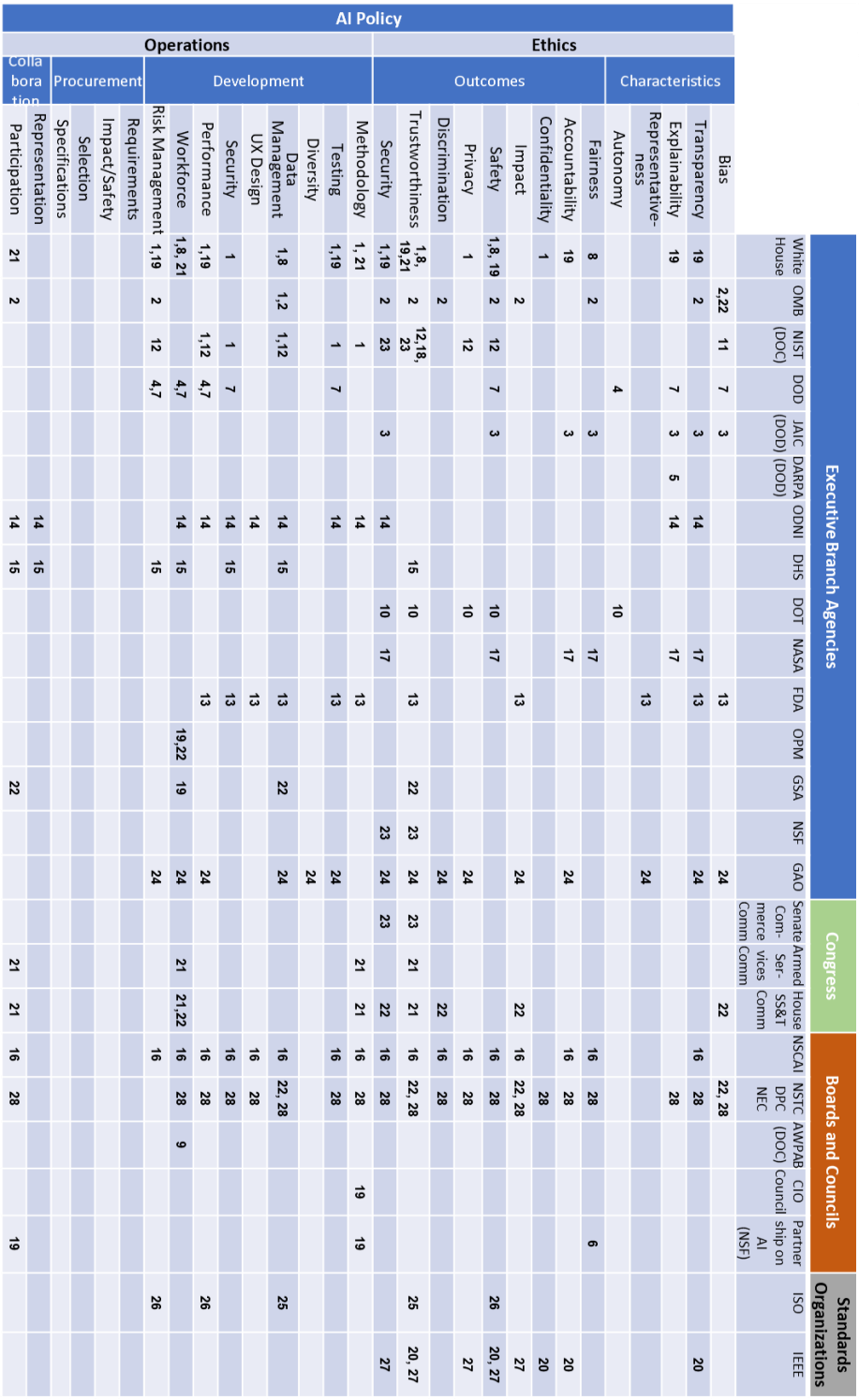

Figure 4 The Current State of U.S. Federal Government AI Policy

1. Executive Order 13859 of February 11, 2019
2. Guidance for Regulation of Artificial Intelligence Applications, M-21-06, November 17, 2020 https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf
3. DOD Ethical AI Principles https://www.ai.mil/docs/Ethical_Principles_for_Artificial_Intelligence.pdf
4. DOD Policy on Autonomous Weapons https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf
5. DARPA Explainable AI Project https://www.darpa.mil/program/explainable-artificial-intelligence
6. NSF Program on Fairness in Artificial Intelligence in Collaboration with Amazon https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505651
7. DOD AI Strategy https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF
8. OECD AI Principles https://www.oecd.org/going-digital/ai/principles/
9. American Workforce Policy Advisory Board https://www.commerce.gov/americanworker/american-workforce-policy-advisory-board
10. DOT AV 4.0 https://www.transportation.gov/sites/dot.gov/files/2020-02/EnsuringAmericanLeadershipAVTech4.pdf
11. Draft NIST Special Publication 1270 A Proposal for Identifying and Managing Bias in Artificial Intelligence https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270-draft.pdf
12. U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf
13. FDA Artificial Intelligence / Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan https://www.fda.gov/media/145022/download
14. ODNI Augmenting Intelligence Using Machines – AIM https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf
15. DHS AI Strategy https://www.dhs.gov/sites/default/files/publications/dhs_ai_strategy.pdf
16. Final Report of the National Security Commission on Artificial Intelligence https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf
17. Framework for the Ethical Use of AI https://ntrs.nasa.gov/api/citations/20210012886/downloads/NASA-TM-20210012886.pdf
18. Draft NISTIR 8332, "Trust and Artificial Intelligence" https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8332-draft.pdf
19. Executive Order 13960 of December 3, 2020
20. IEEE Ethically Aligned Design https://ethicsinaction.ieee.org/
21. National Artificial Intelligence Initiative Act of 2020 (Division E of the 2021 Defense Appropriations Act) https://www.congress.gov/bill/116th-congress/house-bill/6395?q=%7B%22search%22%3A%5B%22%5C%22Artificial+Intelligence+Initiative%5C%22%22%5D%7D&r=5&s=6
22. AI in Government Act of 2020 (Division U Title I of the Consolidated Appropriations Act of 2021 https://www.congress.gov/bill/116th-congress/house-bill/133/text
23. Identifying the Output of Generative Adversarial Networks Act 2020 https://www.congress.gov/116/plaws/publ258/PLAW-116publ258.pdf
24. Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities https://www.gao.gov/assets/gao-21-519sp.pdf
25. ISO Standards IEC:20546:2019, IEC TR 20547-1:2020, IEC TR 20547-2:2018, IEC TR 20547-3:2020, IEC TR 20547-5:2018, IEC TR 24028:2020, IEC TR 24029-1:2021 https://www.iso.org/committee/6794475/x/catalogue/p/1/u/0/w/0/d/0
26. IEEE P2802 - Standard for the Performance and Safety Evaluation of Artificial Intelligence Based Medical Device: Terminology https://standards.ieee.org/project/2802.html
27. IEEE 7010-2020 - IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being https://standards.ieee.org/standard/7010-2020.html
28. The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf

Figure 5 References – The Current State of US Federal Government AI Policy

3.3.3.1 Observations on the Current State of U.S. Federal AI Policy

Looking across the categories of ethics, the primary focus of current AI policy seems to be on Safety and Trust. Current policy also addresses Outcomes much more often than Characteristics and addresses Development to the near exclusion of Procurement and Collaboration.

3.3.4 Recommendation for Future AI Policy Development

Our review of the current documentation related to federal AI policy reveals a consistent pattern for how AI policy is formulated, disseminated, and developed. Executive Orders (EOs) are generally the impetus for new AI policy, describing the administration's AI perspectives, priorities, areas of focus, and overall goals. As directed in those EOs, other federal agencies and organizations then develop policies, analyses, and frameworks to support the EO's goals. Our recommendations reflect this pattern and consist of two suggested EOs, three specific set of policies to be developed, and one suggestion for improving communication and AI policy development.

3.3.4.1 Executive Order on Representation, Impact, and Discrimination

The three policy categories of Representativeness, Impact, and Discrimination are addressed only lightly by current federal AI policy. This is in stark contrast with much of the current discussion of the societal impact of AI. From the ongoing repercussions of Google's firing their two leaders of AI ethics[16], to the release of the film, "Coded Bias"[17], AI systems' potential to bring negative impacts on disadvantaged populations has become a theme of discussion outside of technology research and policy discussions.

In order to address these issues, we recommend an Executive Order be issued clearly conveying to the federal government, the public, and all vendors involved in AI development and deployment, that the AI systems of the federal government will be fair and just to all who use them.

3.3.4.2 Executive Order on Privacy

The two policy categories of Confidentiality and Privacy have been addressed in past EOs and subsequent agency and committee work products, however, as the technology advances, cybercrime proliferates, and commercial applications broaden, the potential harm to our citizens and nation increases. In addition, privacy and confidentiality can negatively impact citizens in two ways – the first in the collection and use of large amounts of data for training models (presenting privacy risk by creating large data sets that collectively can target individuals as well as creating targets for cybercrime) and the second in the use of AI systems in ways that directly impinge on privacy and confidentiality (such as facial recognition and affective interface systems).

[16] Accessed at https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/ on July 1, 2021
[17] https://www.nytimes.com/2020/11/11/movies/coded-bias-review.html

An Executive Order focused on ensuring the privacy and the confidentiality of all data and interactions with federal government AI systems is necessary to enable the trust needed for these systems to provide their potential benefits.

3.3.4.3 NIST to Establish AI Development Standards

In 2019, NIST issued a planning document outlining its approach to developing technical standards for AI development - *U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools*[18]. Our research indicates that policy developers have addressed the majority of technical standard policy areas, providing guidance in areas including Methodology, Testing, Data Management, and Performance. To date, very little policy attention has been paid to UX, and no direction has been given to ensuring Diversity in AI planning, development, and deployment teams.

We urge NIST to prioritize AI standard development among its activities, including creating and publishing a detailed roadmap for developing AI standards. By continuing the effort begun in its 2019 Plan, NIST can provide the technical leadership needed to ensure that AI standards are identified, collaboratively prioritized, coordinated across federal organizations, and developed with the appropriate input and rigor. Until the AI Standards Roadmap is published, agencies and industry will be operating in an environment of uncertainty and risk as they develop new data sets, tools, and platforms. NIST leadership in this area can significantly reduce this risk and enable efficient and effective innovation across government and industry.

3.3.4.4 OMB to Establish Procurement Standards

As federal agencies seek to procure increasing numbers of AI tools and solutions, we suggest that OMB prioritize the creation of clear policies on how AI requirements are developed, how AI tools and vendors are evaluated, and how AI tool development and testing is managed. AI systems present procurement processes with particularly problematic challenges. Commercial AI offerings are advancing rapidly, adding new capabilities, and expanding the use cases they can address. Some of these new capabilities (such as the AC systems described above) present unique policy concerns. The industry-standard reliance on open-source tools, data sources, and platforms complicates licensing and raises security concerns for many agencies. As procurement processes adapt to new methodologies such as Agile and DevOps, MLOps[19] adds more factors to consider, especially with respect to data security, availability, and testing management.

Until more clarity is available regarding the trade-offs required between AI system performance parameters such as accuracy, fairness, trustworthiness, and bias, agencies will be challenged to specify requirements among these competing metrics. The problem of AI systems' bias has led to a call for greater diversity in development, testing, and deployment teams. Documenting requirements addressing this need would be a new challenge for federal procurement

[18] https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf
[19] MLOps refers to the integration of ML data preparation, testing, tuning, and evaluation into the standard DevSecOps framework

processes. AI's "black box" problem is especially acute in the case of the federal government's procurement and use of AI systems. Current regulation found in the Federal Acquisition Regulations (FAR) provides strong intellectual property (IP) protection for vendors, covering their proprietary algorithms, as well as the data they use and the data they generate[20]. These protections may create additional uncertainty for AI system acquisition, as the FAR currently relies on a clear distinction between "software" and "data". When current ML systems are "trained", they generate new data (the refined weights of nodes) and this data is integrated into the new ML model. In effect, the AI system blends customer software with the new data generated. Federal agency procurement policies need to explicitly address how rights to that data and the resulting AI systems are to be distributed, and under what constraints and conditions.

3.3.4.5 NAIAC to Effectively Engage Stakeholders in Ongoing Dialog

Pursuant to the National Artificial Intelligence Initiative Act of 2020 (AI-IA), the administration has initiated the National Artificial Intelligence Advisory Committee (NAIAC), chartered to advise the administration on AI topics ranging from US competitiveness, workforce impacts, effectively implanting the AI-IA, and societal impacts of AI. To effectively address the policy challenges described in this report, we recommend that the NAIAC be fully implemented as soon as practicable, with a high priority on public engagement. Following the previous discussion regarding challenges related to developing trustworthy AI (Section 2.2.3 above), effective communication is critical to understanding and responding to all stakeholders' concerns regarding AI systems. The NAIAC's mandate to include technical, societal, and policy issues in its advisory role gives it a unique position to convene a broad array of stakeholders. Input from these multiple stakeholders will need to be captured and integrated in a transparent manner, so that technical leaders and policy makers are accurately informed as they move forward in developing the needed policies, regulations, and practices.

3.3.4.6 De-emphasize "Trust" as a Policy Goal – Focus on Trustworthiness

Trustworthy AI was established as a policy goal of the federal government by EO 13960 in December 2020[21], and has been cited by subsequent policies and legislation. As initially described, this goal placed the burden of effort on the part of AI system developers to ensure that the systems they developed are worthy of trust. We recommend that this perspective be maintained throughout all federal policy, regulation, and communication. This recommendation is intended to avoid any federal policy or regulation either placing a burden on users for them to either sincerely or perfunctorily "trust" a federal system they are using. We similarly recommend against any use of technology to perceive, model, or manipulate a user's trust in any federal system (see the discussion of Affective Computing in Section 2.3.1 above). Federal government policy makers and AI system developers should seek to avoid the kinds of policy and ethical risks inherent in placing the burden on citizens to trust federal systems.

---

[20] https://www.acquisition.gov/far/part-27#

[21] https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government

4 Next Steps

As depicted in Figure 6, a hierarchy can be imagined that illustrates policies informing laws, which are codified into regulations (as in the CFR or FAR), that are then enabled by standards (such as promulgated by NIST), that then inform methods (proven techniques, utilizing defined standards, that accomplish required tasks that fulfill regulations), that are then enacted through processes (step-by-step, detailed workflows, often documented as Standard Operating Procedures (SOPs)) at the agency level.

Our perspective is that AI policy goals are fulfilled only at the process level – the point where an AI system requirement is documented, AI system tests are defined, AI system performance is assessed, or a citizen's user experience of AI systems conform to methods, standards, regulations, and laws that accurately reflect the AI policy. Our hope is that the clarity, comprehensiveness, and traceability that the AIPA provides supports the transparent distillation of policy goals into the subsequent "layers" of the pyramid.

Figure 6 Policy to Process Pyramid

4.1 Socializing and Validating the AIPA

As the result of an ATARC working group, this report has received feedback from a cross-section of government and industry leaders. We recommend that additional feedback be solicited and be used to both refine the AIPA and to build consensus on its structure, purpose, and value. Of particular interest are the bodies that have recently been established within OSTP (e.g., the NAIIO, NSCAI, NAIAC, and the AI R&D IWG), GSA (AICOE), DOD (JAIC), and NSF to develop and enforce AI policy. A short period of collaboration with these groups may result in a tool that all would find useful and supportive of their mission.

4.2 Applying the AIPA: Three Use Cases

The AIPA is intended to support federal technology policy leaders and managers in defining, enforcing, and understanding AI policies that are relevant to their roles. If successful in providing this support, the AIPA will help build trust in the American people that its government is effectively and efficiently providing governance over, and extracting value from, AI technologies. The following use cases illustrate potential scenarios demonstrating the AIPA's value.

The use cases described below take advantage of the information contained in the AIPA, which is arrayed in both columns and rows. The rows of the framework provide information associated with a specific category of AI policy – what the policies are and how they are related

to each other. The columns of the framework provide information on the roles and current policies formulated by the organizations listed.

4.2.1 Use Case 1: Assessing the Current State of Implementing the NAII with the AIPA

One anticipated use case for the AIPA is in support of the NAIAC's charter to advise the President and the National Artificial Intelligence Initiative Office (NAIIO) on the management and progress made in implementing the AI-IA. If the AIPA's AI policy ontology is reviewed, revised, and accepted, it would be a useful tool to organize all of the current AI policies developed by various federal agencies into a single view. While we have endeavored to provide the "current view" in this report, there are certainly updates, corrections, and additions likely necessary. Once completed, and if regularly updated, the AIPA would provide an accurate, comprehensive, and visual representation of the federal government's AI policy environment. With such an artifact in hand. The NAIAC might then evaluate those areas of AI policy that are underdeveloped (by assessing policy "coverage" in the rows of the AIPA). Similarly, the NAIAC might review organizational roles and responsibilities to ensure that the appropriate groups are involved in defining and implementing AI policy (by assessing the policy role defined in each column of the AIPA). Following such an assessment, further direction might be given to various agencies to adjust the AI policy environment.

4.2.2 Use Case 2: Communicating AI Policy with the AIPA

Given the demonstrated interest that the federal government has in engaging a broad set of stakeholders in the national conversation around AI (e.g., five public workshops in 2015, publishing the findings and recommendations of the National Security Commission on AI, chartering the NAIAC to lead public engagement on AI issues) it seems likely that this communication will continue as new policies, initiatives, and regulations are developed. Given the broad range of policy areas relevant for AI, those stakeholder communications can be improved by framing policy considerations in the categories provided by the AIPA. The policies, organizations involved, and scope of implementation can all be displayed for the areas of AI policy that interest each stakeholder group. If a stakeholder engagement event is planned, for instance, the AIPA can be updated to show any data relevant to the stakeholders' policy interest, including relevant legislation at all phases of development, specific policy statements, related legislative actions, and commercial or technological advances affecting the policy. If used consistently to support stakeholder engagement, the AIPA will enable more effective, consistent, and meaningful communication.

4.2.3 Use Case 3: Providing Accountability for AI Policy with the AIPA

The AIPA documents all governmental and associated organizations' roles in defining and implementing AI policy for the federal government. As such, whenever societal or technological issues related to AI arise, whether related to new technical development or a new awareness of societal impact, the framework can be used to quickly provide an accurate picture of which organizations relevant to the new issue are currently involved, what the current policies are, and how they're being implemented. This information can help inform an appropriate

response, whether that response consists of new or refined policy, regulation, finding, or communication with stakeholders.

5 Appendix A – Acronym Table

Acronym	Represents
AC	Affective Computing
AGI	Artificial General Intelligence
AI	Artificial Intelligence
AI R&D IWG	Artificial Intelligence Research and Development Inter-Agency Working Group
AIGA	AI in Government Act of 2020
AI-IA	Artificial Intelligence Initiative Act of 2020
COE	Center of Excellence
COP	Community of Practice
DOC	Department of Commerce
DOD	Department of Defense
DOE	Department of Energy
DOJ	Department of Justice
EO	Executive Order
FAR	Federal Acquisition Regulation
IC	Integrated Circuit
JAIC	Joint Artificial Intelligence Center
NAIAC	National Artificial Intelligence Advisory Committee
NAII	National Artificial Intelligence Initiative
NAIIO	National Artificial Intelligence Initiative Office
NIST	National Institute for Standards and Technology
NOAA	National Oceanic and Atmospheric Administration
NSCAI	National Security Commission on Artificial Intelligence
NSF	National Science Foundation
NSTC	National Science and Technology Council
OSTP	Office of Science and Technology Policy
PTO	Patent and Trademark Office
SCAI	Select Committee on Artificial Intelligence
SOP	Standard Operating Procedure
UX	User Experience

6 Appendix B – Details of Current Federal Organizations' AI Activities

Guidance/Strategy is for entities that provide guidance, recommendations, and or strategy either government-wide or within their entity (Agency, Board, Group).
Implementation is for entities that are or have already implemented AI projects.
Government-wide is for entities that give Guidance/Strategy across most of the Government.

1. **White House Office of Science and Technology Policy (OSTP).** (Guidance/Strategy, Government-wide) Provides the President and others within the Executive Office of the President with advice on the scientific, engineering, and technological aspects of the economy, national security, homeland security, health, etc.
    a. Several Executive Orders about AI in the federal Government.
    b. Established the National Artificial Intelligence Initiative Office as part of the National Artificial Intelligence Initiative Act, which recently became law as part of the annual National Defense Authorization Act (NDAA).
        i. https://www.congress.gov/bill/116th-congress/house-bill/6395/text
    c. Creation of a select committee on the subject and codifying into law new AI research institutes.
2. **General Services Administration (GSA).** (Guidance/Strategy, Implementation, Government-wide) A leader across the federal Government in bringing numerous cutting-edge technologies to Agencies, including Policies and Governance.
    a. Launched government-wide Artificial Intelligence Community of Practice.
    b. Brought together federal employees who are active in, or interested in, AI policy technology, standards, and programs.
    c. GSA's Office of Government-wide Policy (OGP) has developed a new pilot using AI for Prediction of Regulatory Compliance, known as the Solicitation Review Tool (SRT).
    d. The General Services Administration's AI Center of Excellence (https://coe.gsa.gov/coe/artificial-intelligence.html#service-offerings) studies policy/AI ethics. While the services they provide are more technical, all of the GSA IT Modernization Centers of Excellence follow a Guide to AI Ethics (https://coe.gsa.gov/docs/CoE%20Guide%20to%20AI%20Ethics.pdf).
    e. Data Ethics Framework. 2020 Data Ethics Framework Draft
3. **National Security Commission on Artificial Intelligence.** (Guidance/Strategy, Government-wide) The Commission studies a multitude of issues, including ethical considerations. See their Charter for more information.
    https://www.nscai.gov/about/charter/
    AI Final Report Summary:
    - Part I, "Defending America in the AI Era" (Chapters 1-8), outlines what the United States

    must do to defend against the spectrum of AI-related threats from state and non-state

actors and recommends how the U.S. government can responsibly use AI technologies to protect the American people and our interests.

- Part II, "Winning the Technology Competition" (Chapters 9-16), outlines AI's role in a broader technology competition. Each chapter addresses a critical element of the competition and recommends actions the government must take to promote AI innovation to improve national competitiveness and protect critical U.S. advantages.

https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf

4. **Department of Commerce.** (Guidance/Strategy)
    a. add info on DoC from meeting on 29 April, being sent via email.....
    b. **United States Patent and Trademark Office (PTO).** (Guidance/Strategy, Implementation, Government-wide) AI work on Informal Adjudication.
    c. **NIST.** (Guidance/Strategy, Government-wide) The National Institute of Standards and Technology participates in interagency efforts to develop AI standards, including one on AI trustworthiness.
        i. Four Principles of Explainable Artificial 14 Intelligence
        ii. OMB Circular A-119, Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities
            1. ITI Response to NIST-2019-0001 on Artificial Intelligence Standards
            2. Principles for the Stewardship of AI Applications
5. **Office of Management and Budget (OMB).** (Guidance/Strategy, Government-wide) Issued final guidance to federal agencies on when and how to regulate the private sector use of AI. This document presents a broad perspective on AI oversight, offering a set of guiding principles, etc (from brookings.edu).
6. **Congress.** (Guidance/Strategy, Government-wide)
There is both a Senate and House Artificial Intelligence Caucus.

    Federal AI strategies are coming from the Hill through NDAA modified language, and individual bills.

    - AI can benefit multiple sectors from finance to national security, and the members of these caucuses permeate the relevant committees.
    - Wanting to utilize AI is not a new thing in Congress, but now they are finally getting enough background and advice from companies/organizations to start making meaningful strides towards creating useful legislation.

    There is a lot of debate on the Hill currently about AI. Members seem to want to ensure it is responsibly used, and "trustworthy".
    There are also undertones from certain members about losing in competition to China, and an understanding that we as a nation are somewhat farther behind in the development of AI than other international contenders.

    - **https://artificialintelligencecaucus-olson.house.gov/**

Legislative Framework In Work: *Securing American Leadership in Science and Technology Act*

A strategy to ensure American competitiveness
The U.S. is facing two fundamental challenges to our competitiveness and growth as a nation:
First, foreign countries, especially China, are threatening to outpace us in the science and technology that has paid dividends to our country's economy and national security for decades.
Second, we must respond to a changing climate and develop next-generation technologies to understand it, address it, and mitigate it.
The Securing American Leadership in Science and Technology Act creates a long-term strategy for investment in basic research and infrastructure to protect the economic and national security of the United States.

7. **Joint Artificial Intelligence Center (JAIC).** (Guidance/Strategy) The Department of Defense's (DoD) Artificial Intelligence (AI) Center of Excellence. Supports the transformation of U.S. Joint warfighting and departmental processes through the integration of Artificial Intelligence and enables the empowerment and unification of bottom-up AI development by innovators across the Defense Department.
    a. The JAIC Strategy and Policy Directorate influences Department of Defense strategic policy, leads the assessment for Department AI Transformation, and leads strategic planning for the JAIC itself.
    b. Developed a Strategy, Guiding Tenants, and is Evolving partnerships with industry, academia, allies and partners. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF
    c. Leading in military AI ethics and safety.
8. **Securities and Exchange Commission (SEC).** (Guidance/Strategy, Implementation) AI work on Regulatory Enforcement. SEC has a suite of algorithmic tools to identify violators of securities laws. For example, to detect fraud in accounting and financial reporting, the agency developed the Corporate Issuer Risk Assessment, which has a dashboard of about 200 metrics that can find anomalies in the financial reporting of more than 7,000 corporate issuers of securities. An ML tool identifies filers who might be engaging in suspicious activities by using historical data to predict possible misconduct. And has several other tools in use and obviously sophisticated policies that guide them.
9. **Department of Energy (DOE).** (Guidance/Strategy, Implementation) AI & Technology Office AITO has been looking at separate ethical AI principles released by the Department of Defense and the intelligence community for inspiration, Cheryl Ingstad said during the Microsoft Federal Science & Research Summit
    a. The Department of Energy intends to draft its own set of ethical AI principles that will regulate how it develops, deploys and shares the technology, said the director of the Artificial Intelligence & Technology Office

10. **National Aeronautics and Space Administration (NASA).** (Guidance/Strategy, Implementation) AI/ML user and innovator for decades. AI work on various aspects including AI Ethics.
    a. https://jpl-nasa.libguides.com/blog/AI-Ethics-and-Responsibilities
11. **Social Security Administration (SSA).** (Guidance/Strategy, Implementation) AI work on Formal Adjudication.
12. **HHS.** (Guidance/Strategy, Implementation)
    a. **Food and Drug Administration (FDA).** AI work on Regulatory Analysis.
        i. https://www.fda.gov/media/145022/download
    b. AI Pilot Project catalyzed the GSA to initiate an effort to help federal agencies in implementing AI to perform reviews of regulations and the enhance rule-making process.
    c. Office of the National Coordinator for Health IT--researching. https://www.healthit.gov/
13. **United States Postal Service (USPS).** (Guidance/Strategy, Implementation) AI work on Autonomous Vehicles for Mail Delivery.
    a. https://www.uspsoig.gov/sites/default/files/document-library-files/2017/RARC-WP-18-001.pdf
14. **Defense Innovation Board (DIB) (Board reduced recently).** (Guidance/Strategy) The mission of the Defense Innovation Board is to provide the Secretary of Defense, Deputy Secretary of Defense, and other senior leaders across the Department with independent advice and recommendations on innovative means to address future challenges. The Defense Innovation Board provides specific recommendations but does not implement change itself. It identifies and works with "sponsors" inside DoD to take action, creating a sustainable foundation for successful ideas to take hold. (ref: About (defense.gov))
    a. The Defense Innovation Board released an AI principles report (https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF) and supporting document (https://media.defense.gov/2019/Oct/31/2002204459/-1/-1/0/DIB_AI_PRINCIPLES_SUPPORTING_DOCUMENT.PDF) in 2019 that is highly regarded with the DoD; however, the DIB is currently in transition.
    b. DIB statement on AI and contributing members: AI (defense.gov)
15. **The Office of the Director for National Intelligence.** (Guidance/Strategy)
    a. Principles and a supporting Framework — aim to set a foundation for how and when members of the IC should use, develop and procure AI applications.
    b. ODNI developed its six principles to be consistent with those of the Department of Defense
    c. Intelligence, Office of the Director of National, and Admin. "Artificial Intelligence Ethics Framework for the Intelligence Community." Accessed February 15, 2021.https://www.intelligence.gov/artificial-intelligence-ethics-framework-for-the-intelligence-community.

16. **National Science Foundation (NSF).** (Guidance/Strategy, Implementation) The FY21 NDAA includes several references to AI ethics. For example, the National Science Foundation is required to submit a report on ethics statements.
    a. NSF's Directorates for Computer and Information Science and Engineering (CISE) and Social, Behavioral and Economic Sciences (SBE) together with the Partnership on AI (PAI) are jointly supporting Early-concept Grants for Exploratory Research (EAGERs) to understand the social challenges arising from AI technology and enable scientific contributions to overcome them.
    b. NSF's CISE directorate invites researchers to submit proposals to its core programs that contribute to discovery in research and practice related to fairness, ethics, accountability, and transparency in computer and information science and engineering, including AI.
    c. NSF Program on Fairness in Artificial Intelligence in Collaboration with Amazon (FAI): This initiative provides a significant opportunity to transform research across all areas of science and engineering, including AI.
        i. Advancing Fairness in AI with Human-Algorithm Collaborations
        ii. Addressing the 3D Challenges for Data-Driven Fairness: Deficiency, Dynamics, and Disagreement
        iii. Towards Fairness in Deep Neural Networks with Learning Interpretation
        iv. Towards a Computational Foundation for Fair Network Learning
        v. Fairness-Aware Algorithms for Network Analysis
        vi. Identifying, Measuring, and Mitigating Fairness Issues in AI
        vii. FairGame: An Audit-Driven Game Theoretic Framework for Development and Certification of Fair AI
        viii. Building a Fair Recommender System for Foster Care Services within the Constraints of a Sociotechnical System
        ix. Quantifying Direct and Indirect Consequences of Racial Disparities in Outcomes Following Cardiac Surgery
        x. Auditing and Ensuring Fairness in Hard-to-Identify Settings

7 Appendix C – Contributors

The following individuals contributed significantly to the research, writing, editing, and final preparation of this report. This core team was supported for months by the members of ATARC's AI Data Policy Working Group, whose support provided a wide range of opinions, perspectives, and feedback. We hope the result is a broader, more generally useful and accessible report.

Ken Farber
Industry Co-Chair
TekSynap
ken.farber@teksynap.com

Kristine Lam
Government Co-Chair
GSA
kristine.lam@gsa.gov

Name	Affiliation
Ellery Taylor	GSA
Erica Briscoe	CalypsoAI
Erica Dretzka	DHA
John Scott	VA
John Sprague	NASA
Ken Wilkins	NIH
Kyoung-cheol (Casey) Kim	University of Georgia
Richard Eng	MITRE
Steven Lee	Rotunda Solutions