Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025 ASEE
Paper ID #46238

Engineering U.S. Responsible AI Policy, A Survey, 2020-2025

Daniene Byrne Ph.D., Stony Brook University

I study policymaking for emergent technologies as a design process with social justice impacts. As a SUNY PRODiG+ Fellow in Stony Brook's College of Engineering and Applied Sciences, in the Department of Technology, AI and Society, I am interested in the controversies, consequences and ongoing development of Responsible AI policies for youth-related technologies in media and education. My social science research, connects policy, STS, science communication, and media studies - all relevant to understanding technological policy development, stakeholder voices and the intertwined cultural, social, and political impacts. My dissertation focused on policy design processes for automated driving systems (ADS).

Engineering U. S. Responsible AI Policy, A Survey, 2020-2025

Abstract

The increase in public access to large-scale AI and the enormous variety of current and potential applications has created widespread excitement and sparked concern over unknown and unintended consequences. While AIs rapidly advance into useful tools across broad applications, we do not yet understand AIs' potential harms, social impacts, and outcomes. The public is increasingly using free AI platforms that produce text, images, and media based on prompts, known as Generative AI (GenAI). At the same time, researchers in industry, government, and academia recognize a need for responsible governance of AIs. They question how to regulate powerful AIs being developed at the frontier of computing. Engineers play an important, informative role in this process, offering valuable technical and design knowledge to policymakers, including concerns about risks and ethical applications. This summary identified research papers, governance documents, and industry approaches to responsible AI policy design within the U.S. It provides an overview of the voices at the heart of designing AI policy and demonstrates the challenge of responsibly regulating emergent AI technology. Findings support coursework related to engineering ethics and societal impacts, engineering policy communication, and design projects focused on GenAI. Documents are presented chronologically and interwoven with government initiatives to demonstrate the impact of Executive Orders on shaping AIs' outcomes. Findings will enhance future engineers' expertise in the realities, challenges, and impacts of developing and responsibly governing AIs.

Introduction

The National Academies of Science and Engineering pointed out "Computing research has an obligation to support human flourishing, thriving societies, and a healthy planet [1]". This obligation is a matter of taking responsibility and embedding responsible practices and policies in AI design, execution, and makeup, a priority that builds trust and ensures safety.

Artificial Intelligence, derived from research on deep neural networks has been described as algorithms that allow computers to recognize and learn from patterns in data, and to simulate human intelligence, bypassing the need to provide step-by-step instructions. The U.S. Government's definition of Artificial Intelligence is

> "Artificial Intelligence The term "artificial intelligence" means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to (A) perceive real and virtual environments, (B) abstract such perceptions into models through analysis in an automated manner; and (C) use model inference to formulate options for information or action[2]".

Earlier forms of AI include traditional approaches to automating and optimizing tasks, based on coding commands and predictive AI, which generates forecasts based on historical patterns and

is used in medicine, finance, and weather forecasting. These are referred to as “weak” because they are task-specific, in contrast to “strong” Generative AI (GenAI)[3]. NIST, the National Institute of Standards and Technology describes GenAI as, “The class of AI models that emulate the structure and characteristics of input data to generate derived synthetic content. This can include images, videos, audio, text, and other digital content[4]”. GenAI uses unsupervised statistical processes of data analysis that rely on machine learning, deep learning, and neural networks to process massive datasets. It includes weighting data to capture relevant patterns [5], [6]. Large language models (LLMs) are a form of GenAI, that uses natural language processing (NLP) to generate text used in chatbots and personal assistants, in response to prompts based on word order probabilities. Beyond GenAI, a “stronger” form of AI, known as Artificial Superintelligence or ASI, is under development and expected to “surpass a human’s intelligence and ability[6]”. AIs have been described as being on the verge of changing “not just the field of computing but nearly every field of science and human endeavor[5]”. Some in the industry have framed them as the first steps toward Artificial General Intelligence (AGI), meaning systems that think more like humans in numerous ways. Like humans, AGI will have the ability to ‘think’ about many things across many domains, requiring different recall of datasets and intuition.

This literature survey describes how policies around responsible governance are taking shape as strong AI technologies emerge, and public interaction with them expands exponentially. In November of 2022, the first generative AI (GenAI) ChatGPT, created by OpenAI, was widely released to the public. Earlier versions had been in development and were tested and used for years but the public access and availability made this release significant. Soon other versions of GenAI that were being privately developed also became public. This has kicked off a race to develop more accurate and powerful GenAI products and to win over the public’s attention. Today (early 2025) GenAI tools are available to “anyone with a smartphone in their pocket” for free or at a very low cost and are being increasingly adopted in businesses[5]. Their responsible creation, application, and governance of their usage are still a matter of debate. AIs have been introduced as indispensable tools for increasing productivity while industry governing boards work to understand their governance, risks, and benefits[7].

Developing useful high-quality tools that become the go-to place for consumer use is a race to improve processing, accumulate data, and encourage regular dependent usage. As with cell phone attachment, this paradigm-changing technology is on the verge of becoming embedded in the daily lives of most technology users spurring a race to impress consumers and build loyalty. Major AI companies and their AIs, such as Alphabet (Gemini), Meta (llama), Anthropic (Claude), Microsoft (Copilot, Sydney, and Bing), X (Grok), and OpenAI (ChatGPT) are focused on developing faster, smarter systems that effectively respond to varied prompts across numerous fields and establish user loyalty. Generative AI model creation is moving quickly,

open-source code is available, and new players such as a Chinese LLM, Deep Seek. In early 2025 Deep Seek shocked the markets with a relatively inexpensive GenAI, requiring less processing, changing the game in AI creation. For those tracking AI developments leaderboard websites such as LLM Stats.com and Hugging Face.co, allow up-to-date comparisons of generative AI.

Responsible AI implies that society can depend upon digital tools that are trustworthy and unbiased, operate transparently, protect human privacy, and in their operation, improve lives for humanity. Details about the meaning of these terms and how to responsibly govern AI vary between industry, government, and academia. Developing Responsible AI responsibly implies addressing issues of ethics, bias, harm, and untruths while keeping alert to future dangers in new unforeseen applications and misuse and social impacts. Responsible AI can also include ensuring American industry leadership prevails and censorship is reduced or eliminated. AIs are engineered tools and therefore there are levels of responsibility for the maker - designer, user, or regulator of a tool. Responsible creation, application, and governance of AI usage is still a matter of research and debate. This paper addresses the shifting policy landscape on AI responsibility across the last three years, focusing on GenAI and Artificial General Intelligence (AGI).

Today's engineering students are preparing for jobs at the front lines in our AI future. Their careers will span across industries that regularly use AI and some may be directly involved in its development and application. To fully prepare future engineers, engineering education must encourage students' reflection on the interplay of engineered systems as life-changing technologies, with societal impacts and intersections with governance of the economy and human wellbeing.

Since 2014 the literature on Responsible AI policy has increased exponentially in parallel with its expanded usage. This survey identifies current literature up to late 2024, on Responsible AI Policy and Design across these domains of government, education, and media and describes it in the context of changing policy approaches. Sources support discussions about opportunities, challenges, controversies, and future directions.

The goals of this paper are:

- Provide an overview of AI policy dialogue from 2022 -2025 for engineering students.
- Identify and share Responsible AI perspectives from Government, Academia, and Industry.
- Reveal the interplay between industry, government policymakers, and academic scholars on responsible AI policy design.

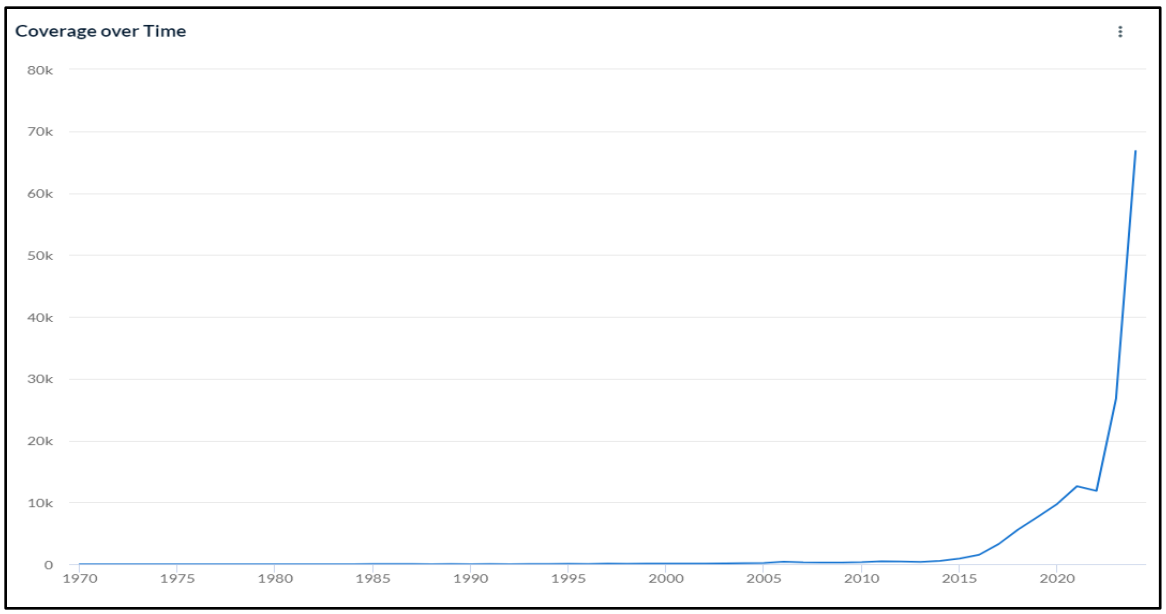

Figure 1. Responsible AI policy 1970- November 2024. EBSCOhost databases search.

Due to time and space limitations, this summary could not address:

- AI usage in specific disciplines, such as Healthcare, Medicine, Real Estate, or Education
- Books on the topic of AI
- Complete analysis of AI Industry Responsible AI policies
- Technical aspects of AI
- Responsible AI policy literature pre-2022
- Ongoing shifts in the executive branch and their administrative impacts post-April 2025

Background:

Like all technologies, AI are tools can be beneficial or harmful and, therefore, must be designed, used, and governed responsibly[8]. The public's previous experience, and subsequent social impacts, of the rollout and adoption of social media have increased awareness of the need for a cautionary, responsible approach to digital technologies[9]. Initial approaches to the internet and social media apps were largely naïve, emphasizing the benefits of shared information and access without fully considering potential dangers. Within the last decade, as smartphones, social media apps, and front-facing cameras became the norm, we have seen negative impacts, including new threats to childhood experiences and shifts in childhood socializing[9]. In addition, it is widely acknowledged that social media platforms have shifted public discourse, increased public expressions of hatred, recruited a new networked generation of cybercriminals and spread large

conspiracy theories, confusion of facts, and distrust of science and the mainstream media[10]–[12].

In 2025, AI technologies, such as content-recommending algorithms and chatbots, are changing the ways we interact with data and process information. As industry doubles down on the information that can be collected, measured, and processed, AI technologies shape our communications and interactions with one another, our privacy, how we learn and work, and how we socialize. Future AI capabilities will extend far beyond those being developed today as researchers strive for the grail of AGI. Increasingly, LLMs on our phones will be embedded in our lives as useful assistants that may also potentially serve as agents of confusion. Working from data collected online, they will have an increasing ability for content generation, deepfake creation, and circulation and amplification of false and dangerous content with no relationship to ground truth.

The need for responsible future-centered AI governance policies is a developing, complex, consequential, multistakeholder design challenge. Future engineers need to be equipped not only with engineering knowledge but also with social awareness and critical abilities to evaluate and question AI and to imagine the benefits of responsible systems. They also need an understanding of governance processes and the ability to share their understanding with policymakers.

Research Questions:

What do literature, governance documents, and surveys reveal are major concerns about

- Responsible AI Policy Design within the U.S.?
- How are concerns about responsible AI surfaced, discussed, and implemented in the U.S.?
- How are shifts in U.S. executive power connected with outcomes for Responsible AI Policy?

Methods

A stepwise literature selection method was inspired by a PRISMA[13] literature review to systematically identify consequential literature on Responsible AI Policy, in November 2024. Documents were downloaded from academic databases Web of Science and EBSCO Host. U.S. Responsible AI Governance documents were pulled from webpages such as the Whitehouse.gov and Whitehouse.gov archives, the Federal Register, the Congressional Record, and government Agency webpages. AI Industry perspectives on the policy process were found on industry web pages. A snowball approach drew in relevant academic papers, webpages, and policy documents rolling out as late as February of 2025. This was necessary to capture the rapidly changing AI

policies today and to show the scholarly voices in the light of industry and government policy processes. The Roper Center for Survey Research was accessed for public opinion data. A 4-part framework identified sources that laid the groundwork for this research.

1. Timely - Legal or Government documents related to US Policy needs on Responsible AI for public consumption (from 2020 forward) were identified. These included Congressional hearings, US Federal government AI policy websites such as NIST, and Executive orders.

2. Accessible Peer-reviewed scholarly papers were identified in literature searches on Web of Science and in EBSCO host for Responsible AND AI AND Policy AND Design. Papers related to the concerns and consequences, promises, and challenges focused on GenAI were selected.

3. Industry documents were found through corporate website searches for policies specifically focused on US Governance of AI (as opposed to global or other national). I looked for Industry-related recent (2023-2024) policies or statements on Responsible AI Practices and looked at the ITI Global position paper[14].

4. Background documentation foundational to understanding current policy trends, such as GenAI industry lobbyist documents and policy statements, were included. Some relevant upcoming conference papers and survey research were also cited[15].

Research began in the Fall of 2024 with The Congressional Research Service Report, R-47373 Science and Technology report to the 118th Congress issued 10-15-2024[16] which shares the historical background of needed AI policy and current public laws around AI already on the books: PL 116-283, PL 116-260, PL 117-167, PL 117-207, and PL 117-263.

Next came a review of President Biden's 2023 Executive Order 14110 to help identify the scope of AI policy challenges and opportunities[17]. Then, Agency documents from NIST.gov and the AI.gov websites were reviewed. Corporate Documents were added based on mentions of corporate policy leadership in policy papers.

The Web of Science (WOS) and EBSCOhost were searched in late November 2024 for documents using the search terms "Responsible AND AI AND Policy AND Design". Available documents were loaded into MAXQDA (qualitative analysis software), and using a systematic, exploratory method, first reading every abstract. Selected papers addressed AI governance challenges, policy, theory definitions, ethics, and broad literature reviews. One hundred eighty-eight initial articles, books, conference proceedings, and papers were identified from WOS, and twenty-eight were identified from EBSCOhost. References were narrowed down based on being published in a peer-reviewed journal and if the content fit the topic of Responsible GenAI based on their title and abstract. Those that were accessible for download were then uploaded into

MAXQDA software for closer reading and preliminary coding[18]. This process identified nine relevant academic papers focused on themes involved in Responsible AI Policy Design processes for AGI, discussed here. No LLMs or AIs were used in the process of identifying or summarizing papers. Articles related to specific applications of AIs and LLMs, such as for health and medicine, biological research, or education, were not included as they are beyond the scope of this work.

Findings and Discussion

There are many aspects of Responsible AI governance, and different groups, inventions, and government moves seek to address them. The rapidly changing policy environment and pace of change in the AI industry and their increased political activities were not anticipated when this study began. Developments led me to approach events and arguments for responsible governance chronologically, describing the forms of responsibility and groups addressing them discussed at the time. The shifting stakeholder priorities become apparent as different groups exhibit power over the creation of AI and the policy process. Throughout the government, those in leadership positions on the National Science Foundation (NSF) and the White House Office of Science and Technology Policy (OSTP) overlap in areas of decision-making. Describing influences on responsible AI policy as they occurred sequentially documents how the policymakers arrived at our current approach to responsible AI governance. For future engineers, this provides historic insight into many levels of responsibility to consider in the design, application, and public experience of AI. It also places consequential and evolving engineering decision-making in the context of the U.S. science and technology policy system.

Support for Research

In 2020, Congress passed the National Artificial Initiative Act, promoting American AI leadership, and President Trump signed it into law. Soon after, on January 12, 2021, President Trump created the National Artificial Intelligence Initiative Office as part of the OSTP with the support of the NSF to develop a shared research infrastructure for AI[19]. A new division, the National Artificial Intelligence Research Resource Task Force (NAIRR), would work within NSF and coordinate cross-agency and industry support for US AI technologies. NAIRR was a three-year pilot project and is an exemplar of the government policy of taking responsibility in coordinating an AI initiative[20][21]. In NAIRR, academic researchers, and industry all worked steadily on AI advancement.

In 2021, the Information Technology Council (ITI), which is the lobbying arm of the AI development industry, presented five policy points to promote and support AI: investment in R&D, facilitating trust (users trusting companies), using AI for cybersecurity, global interoperability, and AI engagement. In addition, they suggested an approach to regulations

aligned around common parameters that ensure regulation is risk-based and context-specific and includes immediate harm responses[14].

In August 2022, under President Biden, the Chips in Science Act became law. The purpose was to increase US chip production and funding in support of AI through investment in infrastructure and jobs in AI development and research. Another example of US Leadership taking responsibility for developing the outer limits of growing AI-based technology.

Respect for Human Rights

In October of 2022, the White House OSTP produced a Blueprint for an AI Bill of Human Rights. This 75-page document laid the groundwork for expectations from all AI systems, notably calling for 1) Safe, effective systems, 2) Algorithmic Discrimination Protection, 3) Data Privacy, 4) Notice and Explanation, and 5)Human Alternatives, Consideration, and Feedback[22].

AI Capabilities and Public Protection from Harm

Not long after, in November of 2022, OpenAI publicly released the first version of a publicly accessible LLM, Chat GPT. They assumed that public use of ChatGPT would help with testing. At the same time several other companies were also working toward these capabilities but were not certain they were ready for release[23]. The race to develop and refine AI that had been happening behind closed doors was now open to the public. The release of ChatGPT marked a distinct shift in thought about the potential of AI in the hands of the public and an urgency for regulation.

In November 2022, Brookings (a left-leaning think tank) published a Global AI Research Agenda, later cited in the U.S. Global AI Research Agenda released 3-14-24. At the same time, a U.S. National AI Strategic Plan was under development.

Agency Approaches to Responsible AI

The Office of Science and Technology Policy and NSF were addressing the meaning and form that responsible AI research should take across government agencies. This NAIRR initiative, which began in the Trump administration, released the results of its pilot in January of 2023. Determining NSF would house the interagency AI research group see NAIRR.nsf.gov[21]. The structure included three advisory boards one covering science, one technology, and one ethics.

> “An Ethics Advisory Board to advise the Operating Entity on issues of ethics, fairness, bias, accessibility, and AI risks and blind spots. The Ethics Advisory Board’s intended roles are to (1) evaluate the ethical use of AI, computational, and data resources by

> NAIRR awardees as well as issues related to scientific integrity, and help the Operating Entity ensure that privacy, civil rights, and civil liberties are not violated; (2) evaluate and advise on the fairness and appropriateness of data and training delivered by the NAIRR; (3) provide guidance on approaches to understanding issues of ethics, bias, and fairness and on NAIRR ethics policies and practices; and (4) handle concerns and/or complaints brought to the Operating Entity's attention or by the User Committee. The Ethics Advisory Board should provide periodic insight and feedback on a broad range of policy issues, guidelines, and practices, including in areas such as privacy, civil rights, and civil liberties. The Ethics Advisory Board should be selected to include 22 experts in privacy, civil rights, civil liberties, and ethics as well as to represent user groups, scientific societies, advocacy and civil society groups, and government[20]."

It would phase in government research support of AI across numerous AI industry leaders.

Their report's conclusion states:

> "The NAIRR can help create opportunities for progress across all scientific fields and disciplines, including in critical areas such as AI auditing, testing, and evaluation; trustworthy AI; bias mitigation; and AI safety. Increased access and diversity of perspectives would, in turn, lead to new ideas that would not otherwise materialize and set the conditions for developing AI systems that are inclusive by design[20]".

Policy Dialogue on Responsible AI

In October 2022, leading computer science researchers who were members of the Association of Computing Machinery (ACM), Technology Policy Council, released a statement of nine Principles for Responsible Algorithmic Systems: legitimacy and competency, minimizing harm, security and privacy, transparency, interpretability and explainability, maintainability, contestability and auditability, accountability and responsibility, and limiting environmental impacts[24]. They noted these "are meant to be inspirational in launching discussions, initiating research, and developing governance methods to bring benefits to a wide range of users while promoting reliability, safety, and responsibility. In the end, it is the specific context that defines the correct design and use of an algorithmic system in collaboration with representatives of all impacted stakeholders[24]".

In 2022, speaking to AGI Erik Brynjolfsson pointed out that "an excessive focus on developing human-like artificial intelligence can lead us into a trap" and argued that responsible AI creation meant setting the goals of augmenting rather than creating human intelligence[25]. By December 2023, Daniel Schiff published a detailed review of the U.S. AI policy agenda from 2016 to 2020. Using the Multiple Streams Framework, which targets policy, politics, and problem streams, he

assessed 63 policy documents and determined that U.S. policy documents paid minimal attention to ethical and social concerns around AI, relative to the focus on geopolitical and economic issues[26].

By March 2023, "The Growing Influence of Industry in AI Research" appeared in the Journal Science as a Policy Forum. Ahmed, Wahed, and Thompson described the trends in AI, which included the growing influence of Industry on AI research[27]. As quickly as new AI technologies were developed, the industry was turning them over into products along with influencing and originating new AI research. In 2021, the US invested $1.2 billion in AI research as compared to $340 billion invested in AI by private industry [27]. They pointed out that increasingly, top computer science researchers were leaving academia, allured by Big Tech's investment and the extremely large AI systems they develop. Ahmed, et.al. proposed the government's goal should be to ensure the presence of sufficient capabilities to help audit or monitor industry models or to produce alternative models designed with the public interest in mind." This would ensure academics are capable of "shaping the frontier of modern AI and benchmarking what Responsible AI should look like[27]".

The U.S. National AI Strategic Plan was updated in May 2023 and put research needs at the forefront of safe, secure, and trustworthy AI. It placed the burden on the government to ensure the public good. It included strategies to address societal implications, ensure safety by setting standards and benchmarks, and emphasize:

> "The federal government plays a critical role in this effort, including through smart investments in research and development (R&D) that promote responsible innovation and advance solutions to the challenges that other sectors will not address on their own. This includes R&D to leverage AI to tackle large societal challenges and develop new approaches to mitigate AI risks. The federal government must place people and communities at the center by investing in responsible R&D that serves the public good, protects people's rights and safety, and advances democratic values[28]".

Executive Order Includes AI Responsibility

In October of 2023, President Biden set out Executive Order 14110 Safe, Secure, Trustworthy Development and Use of AI, which laid out the eight Executive Principles and spurred 100 initiatives, overseen by NIST (the National Institute for that were met between 2023 and 2024 [17], [29]. The Order identified the following paraphrased challenges of working with AI and held every Executive agency to uphold the necessary Principles of:

1. Safe, secure, robust, reliable, repeatable standardized evaluations of AI
    - To mitigate risk and address issues of security and complexity.

    - For example, watermarking and authenticity (knowing an author)
2. Promoting Responsible innovation, collaboration, and competition (through government funds)
3. Supporting American Workers
4. Advancing equity and civil rights and ensuring AI is accountable to protect against bias, discrimination, and abuse.
5. Protecting the interests of Americans who interact with or purchase AI or AI-enabled products
6. Protecting privacy and civil liberties concerning AI
7. Training a government workforce for skilled, responsible use of AI
8. Leading globally in AI progress

This spurred widespread agency participation and reporting led by NIST, and the development of an Artificial Intelligence Risk management framework and developed a congressionally funded division for risk assessment to ensure safe AI use across agencies[29], [30]. Agency responses to the order were efficient and within a year all agencies had complied with initial guidelines. In addition, NIST created an AI study section ARIA (Assessing the Risks and Impacts of AI).

AIRA later issued an early report of their testing procedures and has begun responsible AI risk assessments on Large Language Models (LLMs) [30].

Executive Order 14110 also led to the creation of the AI.gov web resource (no longer accessible in Feb 2025), which included an AI Talent Surge initiative, led by Biden's AI and Tech Talent Taskforce, to recruit AI professionals for jobs across the U.S. government. Biden announced the call in October 2023 and through Tech Talent Fellowships and Direct Agency hirings, and the response was summarized in an April 2024 Report:

> "The response from the public has been fantastic. In the month after EO 14110 was issued, tech talent programs hiring as part of the AI Talent Surge saw an average 288% increase in AI applications compared to previous periods; some tech talent programs saw up to 600% - 2000% increases in AI applications. Moreover, public interest in Federal AI roles remains high. From January through March 2024, applications for AI and AI-enabling roles have doubled as compared to similar periods in 2022 and 2023. The message is clear: the public is ready and motivated to join the Federal Government to work on AI priorities[31]".

The initiative also facilitated direct hiring and established a training program within DOE and DOD to train new workers. The report ended with ten Recommendations for Federal Government to take responsibility for ensuring the U.S. was an AI leader by further increasing AI capacity[31].

Responsibility at the AI Frontier

The term “Frontier AI” is becoming regularly used to refer to the cutting edge of known and unknown artificial intelligence capabilities. The term “Frontier AI” harkens back to Vannevar Bush’s initial promotion for the Government to fund basic scientific research: “Science the Endless Frontier”, and like the famous appeal, gave the impression of unbounded optimism and potential for AI while minimizing ethical impacts and harms[32]. Frontier AI implies a continuing “next” latest unfolding and ongoing invented technological frontier where we might create, as Thomas Kuhn would describe them, paradigm-shifting technologies.

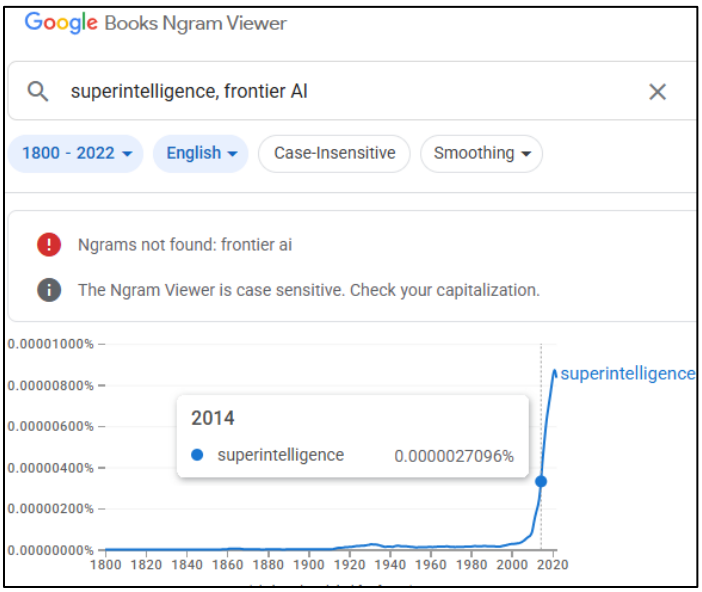

Figure 2. Google Books N-Gram on the term Super Intelligence and Frontier AI.

Superintelligence, Paths, Dangers and Strategies, a 2014 book by Philosopher Nick Bostrom, referred to the creation of AI that exhibits advanced speed and a broad range of intelligence and capabilities beyond those of humans[33]. This is also referred to as Artificial General Intelligence (AGI). When ChatGPT was released in 2022, although it was a somewhat awkward LLM, it sparked the public imagination for a much more powerful kind of AI technology. Whether referred to as Frontier AI, Superintelligence, or AGI, the term implies superhuman technologies and has a hype factor generating great worry from industry and technology developers. AI of the future offers both enormous potential and extreme risks that necessitate strong governance protections[34].

In February 2024, academic developers at Berkeley AI Research, BAIR spoke to new AI, which moved past single LLMs, recognizing “state-of-the-art AI results are increasingly obtained by compound systems with multiple components, not just monolithic models[35]”. Discussions of risks and responsibilities associated with this frontier technology remain to be seen.

By March 2024, G. Helfrich argued we should reject the term “Frontier AI” - as a dangerous glorification that contributes to hype without critically assessing the range of technological harms[36]. Helfrich pointed out that more harmful and immediate risks of LLM deserve immediate attention, including issues of “severe, pervasive, social, psychological and environmental harms that large scale generative machine learning (used in areas such as social media) are already perpetuating[36]. Her points resonated with Haidt’s recognition of the damage social media and constant cell phone use and apps are having on youth[9]. In the case of social media, companies are protected from public lawsuits by Article 230, which states they are not liable for content posted by their users. This has caused grave danger (increasing suicide, violence, abuse, and exploitation) and has been especially dangerous for the vulnerable, including the elderly, stateless, and youth. It was addressed by the Supreme Court in 2023 and remains unchanged [37].

A second AI Policy Forum article appeared as the May 2024 cover of Science, titled “Managing Extreme AI Risks among Rapid Progress[38].” This consensus paper, by twenty-five prominent authors across numerous disciplines, described the dangers of unregulated AI growing increasingly intelligent as “… systems that can autonomously act and pursue goals”. The authors recommended a combination of active governance and research and development in areas of oversight and honesty, robustness, interpretability and transparency, inclusive development, understanding emergent challenges, evaluating dangerous capabilities and AI alignment, risk assessment, and resilience[38]. They suggested governance solutions should include “enforcing standards and preventing recklessness and misuse”, with specific policy suggestions that build on current regional and voluntary guidelines such as:1) proactive risk reduction through mandatory - increasingly rigorous - risk assessments that target developers as responsible; and 2) standards for reducing harms through mitigative strategies in approaching autonomous AI; for example, creating policies that are triggered as AI reaches milestones. They recommended institutions 1) protect and promote low-risk work and research; while 2) focusing risk oversight on the “few, most powerful systems -trained on billion-dollar-supercomputers – which will have the most hazardous and unpredictable capabilities[38]”.

They also supported governance changes to ensure regulators can keep up. These involve 1) mandating whistleblower protections, 2) incident reporting, 3) registration of key information and datasets, and 4) monitoring model development and usage. They recommended allowing external audits at all times, including on-site monitoring for nefarious activities and emergent dangers such as self-replication, large-scale persuasion, breaking into other systems, and hampering autonomous weapons and pathogen development.[38]

The authors emphasized that AI cannot be considered safe and that “developers of Frontier AI should carry the burden of proof to demonstrate risks are acceptable[38].” They push

governments to “set risk thresholds, codify best practices, employ experts and third-party auditors[38]”. Liability frameworks and consequential evaluations were suggested as a means to incentivize safe AI and prevent harm. In addition, as the government builds, the authors recommend IF-Then commitments, describing preventative actions they will take if technologies pass red-line capabilities[38].

Senate Judiciary Hearing: Insider perspectives on AI policy Development

The bi-partisan Senate Judiciary Hearing, Oversight of AI: Insider’s Perspectives, 09-17-2024, led by Senators Richard Blumenthal and Josh Hawley, addressed a full range of technological harms and the lack of industry concern with social issues[39]. These included calls for transparency and understanding models, privacy, understanding the data fed into AI, its origins, methods for cleaning or tagging it, and the potential for bias. Four AI developers, Mitchell , Harris, Saunders, and Toner, provided testimony on the potential harms of AI and recommended policy solutions [23], [39]–[42]. These documents serve as valuable resources for policymakers and educators delving into policy processes.

Mitchell’s testimony provided detailed potential stakeholder groups, policy gaps, and solutions, diagramming methods to ensure responsibility[43], along with descriptions of terms and common misconceptions[41]. Mitchell co-led the ethical AI team at Google and now works at Huggingface.com. She co-authored the critical landmark paper: *On the Dangers of Stochastic Parrots* which recognized the need for critical evaluation of AI as it impacts lives, the recognition that LLMs are statistically randomly processed language repeaters, “stochastic Parrots”, generating language based on probabilities and not intelligent thought.

Testimony by David Evans Harris, Senior Policy Advisor, California Initiative for Technology and Democracy, Chancellor’s Public Scholar, UC Berkeley, San Francisco, CA, provided evidence supporting three major claims:

> “First, voluntary self-regulation does not work; Second, the solutions for AI safety and fairness exist in the framework and bills proposed by the members of the committee; and Third, not all the horses have left the barn. There is still time[40]”.

Saunders, a Former Member of Technical Staff at OpenAI in San Francisco, CA., emphasized the need for making insider communication on AI concerns and responsibilities safe and easy. He described why he left OpenAI:

> “Current AI systems are trained by human supervisors giving them a reward when they appear to be doing the right thing. We will need new approaches when handling systems that can find novel ways to manipulate their supervisors or hide misbehavior until

> deployed. The Superalignment team at OpenAI was tasked with developing these approaches, but ultimately, we had to figure it out as we went along, a terrifying prospect when catastrophic harm is possible. Today, that team no longer exists; its leaders and many key researchers resigned after struggling to get the resources they needed to be successful[23]".

Toner is the Director of Strategy and Foundational Research Grants Center for Security and Emerging Technology, at Georgetown University, Washington, DC. Her testimony began with concerns about the speed and lack of oversight or public protection from AI. Toner emphasized that the science of measuring and managing AI risks and progress is immature, developers are under enormous pressure to achieve launch dates and raise funds, and systems built and deployed now are affecting millions of lives even though we don't understand the science of their harms. Regulation is complex as AI has great potential for good. She then recommended transparency requirements for third-party audits, whistle-blower protections, resourcing NIST and other protective agencies supporting increased hiring of AI experts for government, and ensuring governance includes liability for AI harms.

The forward momentum in AI research and development, along with the promise of life-saving advances and general frontier optimism combined with the push of international competition and the promise of future profits, may be part of the explanation. It may also be that there are conflicts of interest between responsibilities in policy priorities, namely, supporting industry to ensure global leadership versus public safety and safety from emergent unknown harms. Some scholars have pointed out that developers want AI policy to focus on Artificial General Intelligence and away from the more immediate harms of bias, loss of privacy, misrepresentation of fact, and social damage[25], [41].

[2] Dario Amodei, Anthropic CEO: *"My chance that something goes, you know, really quite catastrophically wrong on the scale of human civilization might be somewhere between 10-25%."*
Sam Altman, OpenAI CEO: *"The bad case [...] is, like, lights out for all of us."*
Geoffrey Hinton, Turing Award winner: *"I think we've got a better than even chance of surviving it. But it's not like there's only a 1% chance of [superintelligence] taking over. It's much more than that."*
Ilya Sutskever, OpenAI co-founder: *"The future is going to be good for the AIs regardless. It would be nice if it would be good for humans as well."*

Figure 3. Quotations from AGI Developers in Toner's report at the Oversight of AI, Insider's Perspectives Hearing 9017-2024 [41].

Perhaps the most pressing and incomprehensible question is why any leader would pursue a technology that they believe has a 10-25% risk of a catastrophic impact on civilization.

Around the time of the 2024 Presidential Election, Stanford's Institute for Human AI released the Digitalist Papers[44] This group of 12 papers speaks to the responsibilities of technologists in ensuring new forms of AI support Democracy and merit further exploration.

Assessing Risks and Impacts of AI, ARIA, the new government agency set up under NIST, was an outcome of a Biden executive order and was created to assess the dangers of generative AI [45]. The December 2024 ARIA Evaluation Design Document presents motivations for existing and the approaches to evaluating LLMs. The idea is to create a model test platform that makers of LLM can use and imitate for in-house testing.

> "In contrast to current approaches that rely on probabilities and predictions, ARIA will enable direct observation of AI system behaviors and potential impacts on users. ARIA pairs people with AI applications in scenario-based interactions designed around specific AI risks and studies the results. Applications are submitted to NIST from around the globe and are evaluated on the basis of whether risks materialized in the scenarios and the magnitude and degree of resulting impacts. Participating teams will learn whether their applications can maintain functionality across the varying contexts of the test environment… Evaluation of AI applications starts in ARIA's three-level testbed, in which each level uses a different testing approach to explore potential risks and impacts: 1. Model testing: confirm claimed capabilities 2. Red teaming: stress test and attempt to induce risks 3. Field testing: examine positive and negative impacts that may arise under regular use[46]".

Scholarly work on AI Responsibility and Policy

The role of big tech in the AI policy process is a key theme for scholarly research in policymaking. The Multiple Streams Framework(MSF) was developed by Kingdon in 1984[47], was applied to a 2024 investigation of AI policymaking[48]. Cairney and Jones, 2016, describe three independent streams as a problem stream, a policy solutions stream, and a political stream, and the framework allows each one to be explored across dimensions of actors and their power and influence[26]. Kingdon theorized that the streams generally function independently, but opportunities arise upon their intersections. The recent paper "How and Why is the Power of Big Tech Increasing in the Policy Process? The Case of Generative AI" argues that reimagining MSF is necessary. It expanded the three streams to include a fourth Technology Stream (with two branches): Innovation-Centric and Big-tech Centric[48].

2024 policy research findings, such as "The Governance Fix? Power and Politics in controversies about governing generative AI[49] share concerns that the focus of the government and developers on the major existential risks of superintelligence (models able to outsmart humans)

overlooks other important considerations. For instance, the model's purpose and the role it will have in society as a technological assistant (versus as a replacement). It points out the largely limited roles for the public voices in policy decisions. Describing it as a "paradox of generative AI governance" where a highly salient "widely accessible technology" is narrowly governed[49].

Feminist scholars Drage, McKinsey, and Brown took a practical, direct qualitative approach to understanding responsibility. Instead of addressing policymaking, they went straight into the industry. Their study of responsibility in the development and deployment of AI obtained access to "AI practitioners and tech workers at a single multinational AI technology company[50]." They interviewed employees of all levels of access and skills from across the company. They asked questions such as "What is responsibility?" and "Who (here) is responsible (for different aspects of work done and product developed)?" This approach identified that within the company, there was no overall shared vision or process for responsibility in making and distributing AI. They pointed out the need for defining responsibility in the context of specific jobs and a need for direct chains of responsibility and blame[50].

Ethical issues are central to philosophical and interdisciplinary debates on how to build AI. Yet, no go-to best practice or approach exists. There is little agreement on applying philosophical concepts of morals to machines. Several scholars have pointed out the difficulties in translating philosophical ethical approaches into operational AI outcomes[26], [51], [52] "Mapping the Landscape of Ethical Considerations in Explainable AI Research, "scrutinized the relationship between explainable AI (XAI), a trend in developing AI meant to embed procedures for ensuring ethics. AI applications are ethically motivated. They found that "while many papers acknowledge an engagement of ethics, there is often a lack of deep engagement with theories and frameworks[52]". A finding that reinforced Schiff's work from 2023 [26].

References to ethics occur regularly in industry policy initiatives. An interdisciplinary approach aimed to extract lessons learned from previous technology governance, namely the consequences of the lack of oversight in previous dangerous algorithms of (social media-based) technologies.

Technology Scholars, AI4People, a group of concerned scientists, describe ways to structure AI to best serve humanity, foster human care, expand opportunities, and minimize risks. They develop an ethical framework for AI based on four principles from bioethics: Beneficence, Non-malfeasance, Autonomy, and Justice, and to these, they add Explicability[53]. They are intent on reframing technical goals away from the race to GenAI and toward enriching humanity.

Commentary by Mona Sloan explores the "Controversies, contradiction and "participation" in AI" addressing the role of participants as users and content providers, but the lack of their inclusion in decision-making[54]. Sloan views AI as a largely bureaucratic project that needs to

be challenged. She recommends the way to do so in three steps: recognizing that AI narratives should be addressed to help understand how they mirror how we order and organize society; acknowledging how participation can disrupt AI's narrative "quasi-magical" status and that we need to reframe participation, so it is unscripted, unpredictable, and beyond bureaucratic control. The themes from this research support the scholarship from Kim, Zhu, and Eldardiry, "Toward a Policy Approach to Normative Artificial Intelligence Governance: Implications for AI Ethics Education[51]." They argue for the necessity of "Policy Orientated AI Ethics" engaging students in policy discussions and teaching them how to use resources. The two approaches they offer are "Integrating the Policy Dimension into AI Systems Design and Teaching Policy Processes for AI Governance[51]".

AI Industry Responsibilities

For AI Developers, responsibility takes many forms, such as protecting developers, users, and those impacted and producing high-quality products. Google's 2024 policy initiatives speak to supporting opportunity, responsibility, and security[55]. The Anthropic Responsible Scaling Policy 2024 update describes internal controls, demonstrating responsible AI use through standards[56]. Meanwhile, in late 2024, supporters of the increased use of data in governance spoke to a need for efficient sharing across government organizations [57].

Responsible Policy and Politics

The politics of Responsible Policy have never been more apparent than between December 2024 and February 2025. In December 2024, a new advisor, the Whitehouse, the 'AI Crypto Czar', David Sacks (PayPal founder), was hired by incoming President Trump and a few weeks later would be key in President Trump's AI policy. On January 7, 2025, Meta dropped Facebook's content moderation policy, and Elon Musk introduced to X platform users the first LLM app trained on their social media data, Grok.

On January 13, Open AI's Vice President of Global Affairs, Chris Lehane, added the OpenAI o1 "Open AI Economic Blueprint[58]". Titled in response to the earlier Whitehouse Blueprint for an AI Bill of Rights[22]. Lehane frames the development of AI Technologies described as "frontier models" as the most state-of-the-art large language models that lead on capability benchmarks, key in a "race that Americans can and must win[58]." The document kicks off a ChatGPT US tour and fundraising effort to encourage investors to contribute to the company and its push to install large data centers across the U.S. The Industry is proceeding as it has with other digital technologies. First, it generates in-house rules and then turns to the government for support.

On January 14th, 2025, in a response to his Executive Order 14141, Biden stated "AI will have profound implications for national security and enormous potential to improve Americans' lives

if harnessed responsibly, from helping cure disease to keeping communities safe by mitigating the effects of climate change." He continued, "However, we cannot take our lead for granted... We will not let America be outbuilt when it comes to the technology that will define the future, nor should we sacrifice critical environmental standards and our shared efforts to protect clean air and clean water[59]". Executive Order 14141 was conceived with the Department of Commerce to protect AI chip manufacturing and limit access to chips to U.S. allies[60].

A New Chief Executive

On January 20, 2025, the same day President Trump was inaugurated, DeepSeek-R1, a Chinese LLM, debuted. This surprising AI operates at a high level but was trained at a fraction of the cost of leading AIs. It caused a major market disturbance and the recognition that, AIs have the potential for increasingly ability with less expense [61]. That same day, Biden's Executive Order 14110, on the Safe, Secure, Trustworthy Development and Use of AI was revoked.

A New Executive Order Redefines Responsibility for AI

On January 23, 2025, President Trump issued Executive Order 14179 Removing Barriers to American Leadership in Artificial Intelligence.

> Section 1. *Purpose.* The United States has long been at the forefront of artificial intelligence (AI) innovation, driven by the strength of our free markets, world-class research institutions, and entrepreneurial spirit. To maintain this leadership, we must develop AI systems that are free from ideological bias or engineered social agendas. With the right Government policies, we can solidify our position as the global leader in AI and secure a brighter future for all Americans. This order revokes certain existing AI policies and directives that act as barriers to American AI innovation, clearing a path for the United States to act decisively to retain global leadership in artificial intelligence[62].

Section 5. revokes President Biden's Executive Order 14110 and calls for a thorough investigation of all initiatives brought about by it. The order charged " The APST, the Special Advisor for AI and Crypto (David Sacks a hedge fund investor, formerly of PayPal), and the Assistant to the President for National Security Affairs, APNSA, Michael Waltz, (a former Army Special Forces Officer), to identify and revoke all actions associated with Executive Order 14110 of October 30, 2023 (Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence). It remains to be seen if all previous work within NIST, ARIA, and NAIRR will be approached, revoked, or further enhanced by the new administrators[62].

On Feb 6, 2025, a call for comments from the National Science Foundation changes was posted as a request for comment NS_FRDOC_0001-3479, "Request for Information on the

Development of an Artificial Intelligence (AI) Action Plan". This notice for comments, required before making changes to AI actions by three agencies, the NSF, Networking and Information Technology Research and Development (NIRTD), and National Coordination Office (NCO)is open until March 15, 2025, as of 2-20-2025 comments posted were not viewable[63].

National Institutes of Standards and Technology on AI-specific Risks

As of February 2025, the NIST - National Institute of Standards and Technology's webpage Responsible and Trustworthy AI Resource Center listed AI risks that differ from traditional computing risks in 14 ways:

> "Compared to traditional software, AI-specific risks that are new or increased include the following:
> - The data used for building an AI system may not be a true or appropriate representation of the context or intended use of the AI system, and the ground truth may either not exist or not be available. Additionally, harmful bias and other data quality issues can affect AI system trustworthiness, which could lead to negative impacts.
> - AI system dependency and reliance on data for training tasks, combined with increased volume and complexity typically associated with such data.
> - Intentional or unintentional changes during training may fundamentally alter AI system performance.
> - Datasets used to train AI systems may become detached from their original and intended context or may become stale or outdated relative to deployment context.
> - AI system scale and complexity (many systems contain billions or even trillions of decision points) housed within more traditional software applications.
> - Use of pre-trained models that can advance research and improve performance can also increase levels of statistical uncertainty and cause issues with bias management, scientific validity, and reproducibility.
> - Higher degree of difficulty in predicting failure modes for emergent properties of large-scale pre-trained models.
> - Privacy risk due to enhanced data aggregation capability for AI systems.
> - AI systems may require more frequent maintenance and triggers for conducting corrective maintenance due to data, model, or concept drift.
> - Increased opacity and concerns about reproducibility.
> - Underdeveloped software testing standards and inability to document AI-based practices to the standard expected of traditionally engineered software for all but the simplest of cases.

- Difficulty in performing regular AI-based software testing, or determining what to test, since AI systems are not subject to the same controls as traditional code development.
- Computational costs for developing AI systems and their impact on the environment and planet.
- Inability to predict or detect the side effects of AI-based systems beyond statistical measures[64]".

NIST explained that "existing frameworks and guidance are unable to:

- adequately manage the problem of harmful bias in AI systems;
- confront the challenging risks related to generative AI;
- comprehensively address security concerns related to evasion, model extraction, membership inference, availability, or other machine learning attacks;
- account for the complex attack surface of AI systems or other security abuses enabled by AI systems; and
- consider risks associated with third-party AI technologies, transfer learning, and off-label use where AI systems may be trained for decision-making outside an organization's security controls or trained in one domain and then "fine-tuned" for another[64]".

On February 6, 2025, the National Science Foundation, NSF posted a notice requiring Information on the Development of an Artificial Intelligence (AI) Action Plan[65]. The ACM responded on March 15, 2025, with nine recommendations including these headlines: 4. AI Governance is the responsibility of the Implementing and Developing Entities; Policymakers Should Provide Incentives and Initiate Processes for Voluntary Best Practices; 5. Science-to-Policy programs and Bug-Bounty Programs should enable AI accountability; and the call to 6. Address Challenges to AI Products & Solutions through Public-Private Initiatives.

The ACM's fourth recommendation addressed AI Governance and Responsibility which supported a soft-law approach, similar to that used for regulating autonomous vehicles.

4. AI Governance is the Responsibility of the Implementing and Developing Entities; Policymakers Should Provide Incentives and Initiate Processes for Voluntary Best Practices

The rapid and unpredictable nature of adoption of broadly available AI applications makes governance of the use and deployment of AI by an organization a critical oversight function. Governance of AI should address questions on the responsible party for adopting and using AI in public, institutional, and private settings. AI governance includes an analysis of the AI systems used, the data involved, the algorithmic functions, and use cases to establish policies, processes, and controls to ensure safety, fairness, and compliance.[11]

The consequences of use must rest with the individual and implementing entity. Governance should address questions on the responsible party for adopting and using AI in public, institutional, and private settings. The reliability of the systems serving as AI may vary significantly without quality control models for applications designed to accomplish critical tasks or provide services/benefits. Vendors should be held to specifications that are enforced through contracts.

Figure 4. Detail of the ACM Recommendation 4. On Responsible AI [65].

In the area of Education and AI, responsibility came up as they recommended the Action plan prioritize: “Structured, intentional AI education for all students, evolving from early exposure to AI tools and learning how to use them responsibly and ethically, to understanding algorithmic design, and ultimately to the ability to develop AI solutions and perform research to advance the field [66]”.

Recommendations 7 through 12 all related to AI as related to Education and the Workforce:

7. AI Action Plan should Include AI Education, Workforce Development, and Research.
8. AI Education is Essential for AI Leadership and should Leverage CS2023, [67]
9. The U.S. Should Structure AI Education for Global Competitiveness.
10. Workforce Development and Reskilling Must be a Priority.
11. AI Education and Research is Necessary to Sustain America’s AI Leadership.
12. Global Competitiveness and AI Education (should be prioritized)[66]”.

On April 3, 2025, the DOE announced an initiative to partner with industry in creating 16 new AI data centers across the U.S. Secretary of Energy, Chris Wright stated “The global race for AI dominance is the next Manhattan project, and with President Trump’s leadership and the innovation of our National Labs, the United States can and will win.” “With today’s action, the Department of Energy is taking important steps to leverage our domestic resources to power the AI revolution, while continuing to deliver affordable, reliable, and secure energy to the American people[68]”.

On April 3, Mark Przybocki Chief Information Access Division NIST, Information Technology Laboratory, described the ARIA pilot exercise, "will adjust some of our milestones for the current year, our ARIA "North Star" remains the same – to advance the measurement science and assessment capabilities for AI technology, with a focus on both positive and negative outcomes associated with AI. An updated report will be forthcoming in the Summer of 2025 [69].

On April 7th, 2025, the White House Office of Management and Budget and the Assistant to the President for Science and Technology released two memoranda, M55-21 and M25-22 related to the Responsible Use of AI in the federal government. It "revised policies to facilitate responsible AI adoption to improve public services, marking a "fundamental shift" from the prior Administration; changes included introducing forward-leaning, pro-innovation, and pro-competition mindset rather than pursuing the risk-averse approach, removing unnecessary restrictions. It intended to have government agencies embrace AI adoption, to become more "agile, cost-effective, and efficient." Expected outcomes included "improving lives of the American public while enhancing America's global dominance in AI innovation[70]". In a call to remove barriers to innovation, it redefined Agency Chief AI Officer roles to "serve as change agents and AI advocates, rather than overseeing layers of bureaucracy." They are to "promoting agency-wide AI innovation and adoption for lower-risk AI, mitigating risks for higher-impact AI, and advising on agency AI investments and spending. It also suggested the creation of a "high-impact AI" category to track AI use cases that require heightened due diligence because of potential impacts on the rights or safety of the American people[70].

Several of these suggestions were reflected in the April 23,2025 Executive Order: Advancing Artificial Intelligence Education for American Youth[71]. It refers to a nationwide challenge for reimaging AI in education, and states "Within 90 days of the date of this order, the Secretary of Education shall issue guidance regarding the use of formula and discretionary grant funds to improve education outcomes using AI, including but not limited to AI-based high-quality instructional resources; high-impact tutoring; and college and career pathway exploration, advising, and navigation[68]".

The changes in AI policy approaches between November 2024 and April 2025 have been dramatic, yet the basic need to ensure AI enhances humanity remains an urgent governance challenge. I April of 2025, Elon University's Imagining the Digital Future, asked numerous experts to imagine our AI future in year 2035[72]. Close to 200 experts responded, with the majority expecting considerable, deep meaningful change or dramatic fundamental change in the next ten years.

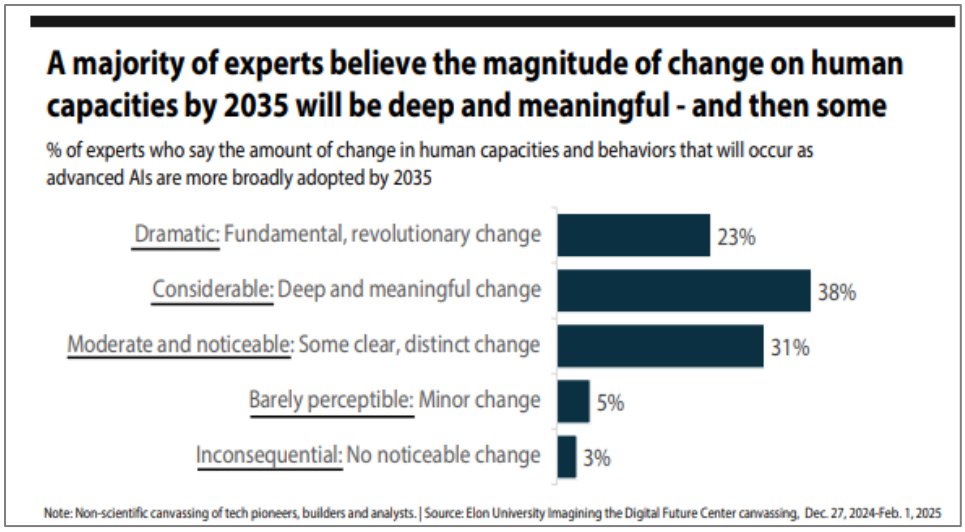

Figure 5. Non-Scientific Canvassing of Tech Pioneers, Feb 2025[71].

As GenAI tools become a regular part of daily life, we still lack the control necessary to ensure we are protecting people from AI-related harms. Numerous concerns remain across immediate issues from trustworthiness, transparency, and some negative impacts of use including addiction and depression; to cybercrime: to the many educational research challenges such as understanding their impact on education and the workforce, including skill perception, attention, persuasion, and learning; to more far-reaching concerns about disruptions and consequences of AGI.

Research Limitations

This research represents a limited snapshot of a changing process in an era (2022- 2025) of enormous policy upheaval. This paper began as a dive into the arguments around developing responsible AI, and a defined literature review led to the recognition that the voices of scholars are only a part of the policy debate and although necessary may have little impact on AI governance today. This work does not include the insights in several recent (2021-2025) conference papers and numerous books in the area of responsible AI, policy, and design[73]. An overview of conference papers and of papers on pre-publication sources such as those on arXiv will inform a future paper. This research excluded much of the literature by humanists, AI ethicists, legal scholars, and philosophers which were published in a book format and not in the targeted databases.

Due to time and scope, the documentation on Responsibility from GenAI companies was not comprehensive and will be addressed in a future project.

Conclusion

C.P. Snow spoke to two cultures, and the dangers of siloed knowledge[74]. Today, sixty-four years after his famous Rede Lecture, cultural, academic, and social silos persist. The dangers of misunderstanding across cultures of humanism and technology expand far beyond the academy. In the literature and documentation on today's technological development of AI different usages of terminology, especially the use of “trust” and “responsibility” in policy debates demonstrate how groups can talk past one another, or adopt the use of the same words, but use them to imply different governance meanings.

While academics focus on bias, ethics, and attention to the term responsibility within a tech firm; responsibility from an industry perspective means protecting established industries and helping them flourish, through increasing R&D, promoting trust, ethics, and collaboration. For President Trump and his administration, a responsible approach means decreasing government involvement in industry regulation and freeing the industry to be responsible on its own. This policy vision supports industry flourishing as a societal partner, creating the AI educators and learners will depend on in the future. It also means switching government services to become AI-enabled and dependent, and ensuring Americans as young as five regularly use AI within their educational experience. It means creating more public-private partnerships to support AI in education[71]. Today AI policies center on the support and growth of the American industry as the most responsible way to manage AI. They ensure progress in AI is not stymied by regulations. This shift in the role of government from a more hesitant and cautionary stance for protecting citizens to one that supports experiment and exploration with AI, exposing all learning citizens to new technologies, and encouraging them to trust the industry even as history has demonstrated trust does not necessarily represent trustworthiness [72].

The term Responsible AI has emerged as a rallying call for university-level interdisciplinary research focused on ensuring safety and risk mitigation – along with the need for educating students with skills to help them evaluate AI, prepare for future work with AI, and mitigate the negative impacts of AI and society. In early 2025, literature research on the ethical and responsible aspects of AI gives an incomplete view of the situation. Executive orders, policy memos, and industry and lobby groups papers provide a better assessment of the situation. U.S. AI policy can sometimes seem to largely be a matter of the dance between the Chief Executive and technology leaders.

Broadly speaking, for humanists and social scientists, building Responsible AI will require a better understanding of the capabilities and basic engineering of today’s technologies, along with evaluations of our engagement with them and our ways of understanding them (as well as their ways of interpreting us – warts and all). It will require understanding how we humans will have autonomy in our decisions to use them, and how we can understand vulnerability gaps[75]. It

will also require creating together with engineers the future goals for developing technologies that support and enhance humankind and are limited in their destructive and harmful capacities. It will also mean bridging the cultural divide and openly discussing all aspects of technology and how to think about responsibility across all aspects of technological impacts. For engineers, this means evaluating the capacities of our technologies and considering their future consequential and potentially revolutionary applications.

The recent shifts in AI policy and the meaning of Responsible AI are likely to continue. They mean that engineering education must go beyond technological know-how, to explore the context of engineering systems and their societal impacts[51]. This includes evaluating how cutting-edge engineering research is prioritized and funded, considering the needs and future outcomes, as well as understanding the political and social dynamics and outcomes that technologies play into the space for engineers' voices, the needs of industry, and the role of whistleblowers. It means designing future AI and AI programs and research with consideration of how they will morph as they interact within social communities, both in intended ways and unintended ways. It means imaginary design challenges that include understanding social needs, consequences, necessary protection, and systemic challenges with increasing unknowns.

STS scholars have historically framed challenges in the context of dynamic systems and controversy studies; Philosophers evaluate AI in terms of ethical contexts; social scientists look at the social and political impacts of technological changes, while developers and politicians frame them in the language of business as opportunities and challenges. It will be up to the future engineers to account for numerous stakeholder voices and to design a future that advances all of humanity; one that can listen to stakeholders and inform human-centered leadership. In February 2025, the state of federal governance for AI, and AI policies generally within the U.S. were in flux. In April of 2025, we are beginning to see the solidification of new, foundational national initiatives for AI in Education. This only reinforces the need for students to have a foundational understanding of the literature and AI policy generation up to this point, so they have a better way to evaluate and contribute to decisions and research in the future that are likely to impact their work and lives.

On January 15th, 2025, Issues in Science and Technology Magazine arrived with a cover story featuring Darío Gill, the new Chairperson of the National Science Board, which oversees NSF Research Funding, and current Senior Vice President and Director of Research at IBM. In an interview with Molly Galvin, he declared “Technology has been elevated to the same level of geopolitical importance as things like trade or military alliances. It’s actually the new currency of power[76]”. This was directly confirmed a week later by the recent direct involvement of leaders of the Tech industry’s Inauguration appearances and their increasing involvement in governance.

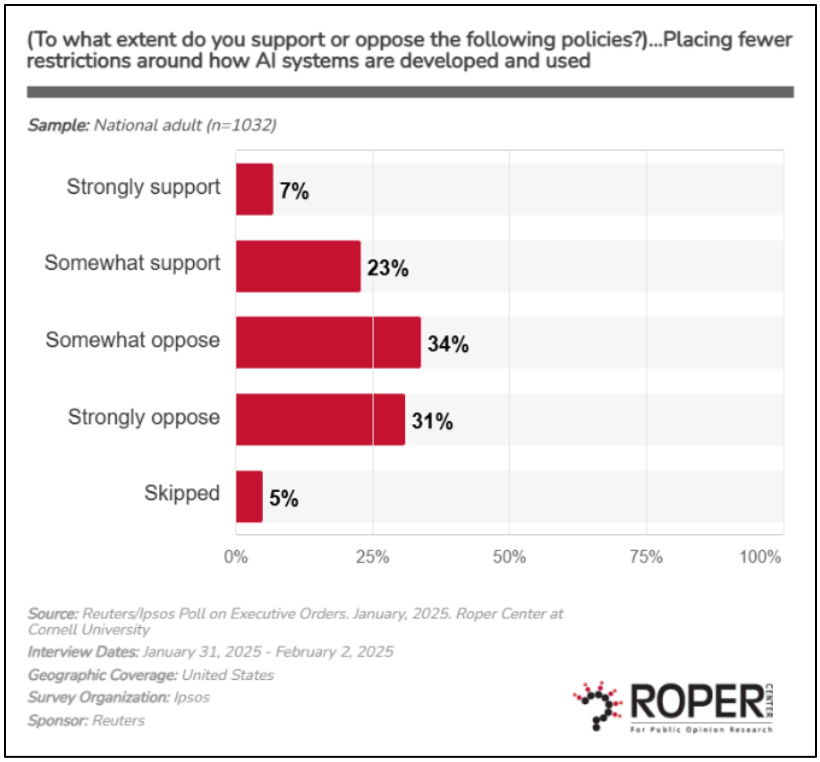

Figure 6. Polling in Early 2025, of 1032 respondents across the US, 65% oppose or strongly oppose placing fewer restrictions on how AI systems are developed and used.

Today, the "innovation-centric" and "big tech-centric" policy streams are washing out the three traditional streams of policy problems, policy solutions, and political streams described in the MSF[48]. In early 2025 Elon Musk, supported by Donald Trump was deeply involved in laying off scores of recently hired federal workers[77]. He oversaw broad sweeping changes to some U.S. Government operations. It is difficult to tell at this point how lasting they will be. 'Responsibility' for Elon Musk means removing waste and fraud and reducing the size of government with little regard for previous functions. He spent the winter of 2025 dedicating his time to recruiting tech industry leaders as volunteers to help with government efficiency and overseeing a nimble young team of computer scientists. Together they had access to numerous government data systems. In April 2025, after many changes, he began to step back from that effort. It remains to be seen how effective his methods were and how the abrupt changes now challenged in the courts will be decided. The recent memorandum by NSF outlined how agency leaders are to respond to this through the widespread adoption of AI[70].

A recent, nationwide Jan-Feb 2025 Reuters/Ipsos Poll on Executive Orders, showed 65% of respondents oppose or strongly oppose lowering the restrictions on AI system development and use restrictions, and only 30% support less oversight[78].

The impact this will have on Responsible AI policy design initiatives, such as progress made in hiring AI experts across government agencies and work at NIST and ARIA, NAIRR is still unknown. By the time this is presented at the ASEE Conference in June 2025, what now feels consequential will be older news, and the impacts will be playing out in the courts. The future of Responsible AI governance is unpredictable. This research revealed that although voices in academic literature are important, it is also valuable for students to understand how the meaning of responsibility and responsible policies are defined by those in power and how the governance system will impact technological outcomes. Technological policy outcomes for engineers are not only linked to their technological use but may also be directly linked to their careers and intertwined with the courses of their lives. The more they understand how they came to be and potentially can participate in the process of their development, the better.

Bibliography

[1] National Academy of Science Engineering and Medicine, *Fostering Responsible Computing Research: Foundations and Practices*. 2022.

[2] US Government Publishing Office, *15 U.S.C. Chapter 119 NATIONAL ARTIFICIAL INTELLIGENCE INITIATIVE Section 9401-Definitions*. U.S., 2022.

[3] "Microsoft AI 101 > Generative AI vs other types of AI," *Microsoft.com*, 2025. https://www.microsoft.com/en-us/ai/ai-101/generative-ai-vs-other-types-of-ai.

[4] National Institute of Standards and Technology, "Glossary, NIST Computer Security Research Center," *Webpage*, 2025. https://csrc.nist.gov/glossary.

[5] J. Ormond, "Fathers of Deep Learning Revolution Receive ACM A.M. Turing Award," *ACM*, 2019, [Online]. Available: https://www.acm.org/media-center/2019/march/turing-award-2018.

[6] IBM, "AI vs. machine learning vs. deep learning vs. neural networks: What's the difference?," 2025. https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks.

[7] L. Abrash, A. Probst, and K. Edelman, "Governance of AI: A Critical Imperative for Today's Boards," 2024.

[8] W. B. Arthur, *The Nature of Technology: What it is and How it Evolves*. New York, NY: Free Press; Simon and Schuster, 2009.

[9] J. Haidt, *The Anxious Generation, How the Great Rewiring of Childhood is Causing an Epidemic of Mental Illness*. New York: Penguin Press, Random House, 2024.

[10] M. Moxley, "They Stole a Quarter Billion in Crypto and Were Caught Within a Month," *NYT Magazine*, 2025.

[11] S. Kudugunta and E. Ferrara, "Deep Neural Networks for Bot Detection." Accessed: Sep. 14, 2018. [Online]. Available: https://arxiv.org/pdf/1802.04289.pdf.

[12] M. Al-Ramahi, A. Elnoshokaty, O. El-Gayar, T. Nasralah, and A. Wahbeh, "Public Discourse Against Masks in the COVID-19 Era: Infodemiology Study of Twitter Data," *JMIR Public Heal. Surveill 2021;7(4)e26780 https//publichealth.jmir.org/2021/4/e26780*, vol. 7, no. 4, p. e26780, Apr. 2021, doi: 10.2196/26780.

[13] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *The BMJ*, vol. 372. 2021, doi: 10.1136/bmj.n71.

[14] I. Information Technology Council, "ITI ' s Global AI Policy Recommendations," 2021. [Online]. Available: https://www.itic.org/policy/artificial-intelligence/itis-global-ai-policy-recommendations.

[15] H. H. Lee, L. Tankelevitch, I. Drosos, S. Rintel, R. Banks, and N. Wilson, *The Impact of Generative AI on Critical Thinking : Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers*, vol. 1, no. 1. Association for Computing Machinery, 2025.

[16] N. T. Carter, "Science and Technology Issues for the 118th Congress," Washington, D.C., 2024. [Online]. Available: https://crsreports.congress.gov/product/pdf/R/R47373.

[17] P. J. Biden, "Executive Order 14110 Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," *Pres. Doc.*, vol. 88, no. 210, pp. 75191–75226, 2023, doi: 10.4324/9780203122273.

[18] B. Leimbigler, “Using MAXQDA for Identifying Frames in Discourse Analysis.” Verbi Software, GmbH, Berlin, pp. 121--133, 2019.
[19] Office of Science and Technology Policy, “The White House Launches National Artificial Intelligence Initiative Office,” 2021. https://trumpwhitehouse.archives.gov/briefings-statements/white-house-launches-national-artificial-intelligence-initiative-office/.
[20] NAIRR (National AI Research Resource), “Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem,” 2023. [Online]. Available: https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf.
[21] NAIRR Task Force, “U.S. NSF National Artificial Intelligence Research Resource Pilot,” 2025. https://www.nsf.gov/focus-areas/artificial-intelligence/nairr#about-the-nairr-pilot-bcb.
[22] The White House, “Blueprint for an AI Bill of Rights - MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE,” *White House*, no. October, pp. 1–73, 2022, [Online]. Available: https://www.whitehouse.gov/ostp/ai-bill-of-rights.
[23] W. Saunders, “Testimony Insider Perspectives,” 2024, [Online]. Available: https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-insiders-perspectives.
[24] R. Baeza-Yates *et al.*, “ACM Statement on Principles for Responsible Algorithmic Systems,” Washington, D.C., 2022. [Online]. Available: https://www.acm.org/binaries/content/assets/public-policy/final-joint-ai-statement-update.pdf.
[25] E. Brynjolfsson, “The Turing Trap: the Promise and Peril of Human-Like Artificial Intelligence,” *Stanford Institute for Human-Centered Artificial Intelligence, Digital Economy Lab*, 2022. https://digitaleconomy.stanford.edu/news/the-turing-trap-the-promise-peril-of-human-like-artificial-intelligence/.
[26] D. S. Schiff, “Looking through a policy window with tinted glasses: Setting the agenda for U.S. AI policy,” *Rev. Policy Res.*, vol. 40, no. 5, pp. 729–756, 2023, doi: 10.1111/ropr.12535.
[27] B. N. Ahmed, M. Wahed, and N. C. Thompson, “The growing influence of industry in AI research,” *Science (80-.).*, vol. 379, no. 6635, pp. 884–886, 2023, doi: 10.1126/science.ade2420.
[28] C. Prabhakar, Arati and A. E. Koizumi, Kei, Principle, “National AI Research and Development Strategic Plan 2023 Update,” 2023.
[29] NIST, “Artificial Intelligence Risk Management Framework : Generative Artificial Intelligence Profile,” *Natl. Inst. Stand. Technol.*, 2024, [Online]. Available: https://doi.org/10.6028/NIST.AI.600-1.
[30] R. Schwartz *et al.*, “The Assessing Risks and Impacts of AI (ARIA) Program Evaluation Design Document,” 2024.
[31] AI Tech and Talent Task Force, “INCREASING AI CAPACITY ACROSS THE FEDERAL GOVERNMENT AI Talent Surge Progress and Recommendations,” no. April, 2024.
[32] V. Bush, “Science the Endless Frontier, A Report to the President, July 1945,” 1945. https://www.nsf.gov/about/history/EndlessFrontier_w.pdf (accessed Apr. 21, 2021).

[33] N. Bostrom, *Superintelligence, Paths, Dangers and Strategies*, First. Oxford: Oxford University Press, 2014.
[34] National Institute of Standards and Technology, "Trustworthy and Responsible AI," *NIST Website*, 2024. https://www.nist.gov/trustworthy-and-responsible-ai.
[35] M. Zaharia *et al.*, "The Shift from Models to Compound AI Systems," *BAIR Berkeley Artificial Intelligence Research*. https://bair.berkeley.edu/blog/2024/02/18/compound-ai-systems/.
[36] G. Helfrich, "The harms of terminology: why we should reject so-called 'frontier AI,'" *AI Ethics*, vol. 4, no. 3, pp. 699–705, 2024, doi: 10.1007/s43681-024-00438-1.
[37] A. Liptak, "Supreme Court won't hold Tech Companies Liable for User Posts," *New York Times*, New York, May 18, 2023.
[38] Y. Bengio *et al.*, "Managing extreme AI risks amid rapid progress: Preparation requires technical research and development, as well as adaptive, proactive governance," *Science (80-.).*, vol. 384, no. 6698, pp. 842–845, 2024, doi: 10.1126/science.adn0117.
[39] Co-Chairs Senators Richard Blumenthal and Josh Hawley, "Oversight of AI: Insider Perspectives," 2024, [Online]. Available: https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-insiders-perspectives.
[40] D. E. Harris, "Testimony Insider Perspectives," 2024, [Online]. Available: https://www.judiciary.senate.gov/committee-activity/hhttps://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-insiders-perspectives.
[41] M. Mitchell, "Testimony Insider Perspectives," 2024, pp. 1–23, [Online]. Available: https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-insiders-perspectives.
[42] H. Toner, "Written testimony of Helen Toner Director of Strategy and Foundational Research Grants Center for Security and Emerging Technology, Georgetown University Before the U . S . Senate Committee on the Judiciary Subcommittee on Privacy, Technology, and the," pp. 1–6, 2024, [Online]. Available: https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-insiders-perspectives.
[43] M. Mitchell *et al.*, "Model cards for model reporting," in *FAT* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, Jan. 2019, pp. 220–229, doi: 10.1145/3287560.3287596.
[44] E. Brynjolfsson, A. Pentland, N. Persily, C. Rice, and A. Aristidou, *The Digitalist Papers, Artificial Intelligence and Democracy in America*. Stanford, CA: Stanford Digital Economy Lab, 2024.
[45] R. Schwartz *et al.*, "The Draft NIST Assessing Risks and Impacts of AI (AIRA) Pilot Evaluation Plan," 2024.
[46] R. Schwartz, G. Waters, and R. Amironesei, "The Assessing Risks and Impacts of AI (ARIA) Program Evaluation Design Document," *NIST*, 2024.
[47] J. Kingdon, *Agendas, Alternatives and Public Policy*, 2nd ed. New York: Pearson, 2010.
[48] S. Khanal, H. Zhang, and A. Taeihagh, "Why and how is the power of Big Tech increasing in the policy process? The case of generative AI," *Policy Soc.*, vol. 00, no. 00,

2024, doi: 10.1093/polsoc/puae012.
[49] I. Ulnicane, “Governance Fix? Power and politics in controversies about governing generative AI,” *Policy Soc.*, vol. 00, no. 00, pp. 1–15, 2024.
[50] E. Drage, K. McInerney, and J. Browne, “Engineers on responsibility: feminist approaches to who’s responsible for ethical AI,” *Ethics Inf. Technol.*, vol. 26, no. 1, Mar. 2024, doi: 10.1007/s10676-023-09739-1.
[51] D. Kim, Q. Zhu, and H. Eldardiry, “Toward a Policy Approach to Normative Artificial Intelligence Governance: Implications for AI Ethics Education,” *IEEE Trans. Technol. Soc.*, vol. 5, no. 3, pp. 325–333, 2024, [Online]. Available: https://ieeexplore.ieee.org/document/10614075.
[52] L. Nannini, M. Marchiori Manerba, and I. Beretta, “Mapping the landscape of ethical considerations in explainable AI research,” *Ethics Inf. Technol.*, vol. 26, no. 3, 2024, doi: 10.1007/s10676-024-09773-7.
[53] L. Floridi *et al.*, “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations,” *Minds Mach.*, vol. 28, no. 4, pp. 689–707, 2018, doi: 10.1007/s11023-018-9482-5.
[54] M. Sloane, “Controversies, contradiction, and ‘participation’ in AI,” *Big Data Soc.*, vol. 11, no. 1, 2024, doi: 10.1177/20539517241235862.
[55] A. Google, “A Policy Agenda for Responsible Progress in Artificial Intelligence.” online, 2024.
[56] Anthrop\c, “Responsible Scaling Policy,” 2024. [Online]. Available: https://www.anthropic.com/news/announcing-our-updated-responsible-scaling-policy.
[57] N. Hart and S. Kent, “Bad Data Costs Americans Trillions. Let’s fix it with a renewed data strategy.,” *Federal News Network.com*, Dec. 04, 2024.
[58] C. Lehane, “AI in America, OpenAIs Economic Blue Print,” 2025. [Online]. Available: https://cdn.openai.com/global-affairs/ai-in-america-oai-economic-blueprint-20250113.pdf.
[59] P. J. R. Biden, “Statement on Signing an Executive Order on Advancing U.S. Leadership in Artificial Intelligence Infrastructure,” *The American Presidency Project*, 2025. https://www.presidency.ucsb.edu/node/375796.
[60] P. J. R. Biden, *Executive Order 14141 Advancing U.S. Leadership in AI Infrastructure*. 2025.
[61] M. Aspan, “U.S. stock markets tumble as investors worry about DeepSeek,” *NPR*, 2025.
[62] President Trump, “Executive Order 14179, Removing Barriers to American Leadership in Artificial Intelligence,” 2025. [Online]. Available: https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence#print.
[63] National Science Foundation and Networking and Information Technology Research and Development(NITRD) National Coordination Office, “NS_FRDOC_0001-3479 Request for Information on the Development of an Artificial Intelligence (AI) Action Plan,” 2025. [Online]. Available: https://www.federalregister.gov/documents/2025/02/06/2025-02305/request-for-information-on-the-development-of-an-artificial-intelligence-ai-action-plan.
[64] NIST, “Trustworthy and Responsible AI Resource Center: Appendix B How AI Risks

Differ from Traditional Software Risks,” 2025. [Online]. Available: https://airc.nist.gov/airmf-resources/airmf/appendices/app-b-how-ai-risks-differ-from-traditional-software-risks/.
[65] S. H. Plimpton, “Notice: Request for Information on the Development of an Artificial Intelligence(AI) Action Plan.” [Online]. Available: https://www.federalregister.gov/documents/2025/02/06/2025-02305/request-for-information-on-the-development-of-an-artificial-intelligence-ai-action-plan.
[66] K. JASMINE, “Comments in Response to the Office of Science and Technology Policy Request for Informatin on the Developmetnt of an Artificial Intelligence (AI) Action Plan,” 2025. doi: 10.1016/j.patter.2022.100455.9.
[67] Association for Computing Machinery and IEEE COmputer Society, *CS2023, ACM/IEEE-CS/AAAI Computer Science Curricula*. 2023.
[68] U.S. Department of Energy, “DOE Identifies 16 Federal Sites Across the Country for Data Center and AI Infrastructure Development,” *https://www.energy.gov*, Apr. 03, 2025. https://www.energy.gov/articles/doe-identifies-16-federal-sites-across-country-data-center-and-ai-infrastructure (accessed Apr. 05, 2025).
[69] National Institute of Standards and Technology, “Assessing Risks and Impacts of AI,” *NIST Website*, 2025. https://ai-challenges.nist.gov/aria.
[70] White House Advisor on Science and Technology and Office of Management and Budget, “Fact Sheet: Eliminating Barriers for Federal AI Use and Procurement: M25-21, M25-22,” *White House Website*, 2025. https://www.whitehouse.gov/fact-sheets/2025/04/fact-sheet-eliminating-barriers-for-federal-artificial-intelligence-use-and-procurement/.
[71] President Trump, “Executive Order Advancing Artificial Intellgence Education for American Youth.” The Federal Register, [Online]. Available: ttps://www.whitehouse.gov/presidential-actions/2025/04/advancing-artificial-intelligence-education-for-american-youth/.
[72] B. J. Anderson and L. Rainie, “Expert Views on the Impact of AI on the Essence of Being Human,” 2025. [Online]. Available: https://imaginingthedigitalfuture.org/wp-content/uploads/2025/03/Being-Human-in-2035-ITDF-report.pdf.
[73] A. Kawakami, A. Coston, H. Zhu, H. Heidari, and K. Holstein, “The Situate AI Guidebook: Co-Designing a Toolkit to Support Multi-Stakeholder Early-stage Deliberations Around Public Sector AI Proposals,” *Conf. Hum. Factors Comput. Syst. - Proc.*, 2024, doi: 10.1145/3613904.3642849.
[74] C. P. Snow, “The Two Cultures, The Rede Lecture,” Cambridge, 1959. Accessed: Mar. 02, 2019. [Online]. Available: http://s-f-walker.org.uk/pubsebooks/2cultures/Rede-lecture-2-cultures.pdf.
[75] S. Vallor and T. Vierkant, “Find the Gap: AI, Responsible Agency and Vulnerability,” *Minds Mach.*, vol. 34, no. 3, pp. 1–23, 2024, doi: 10.1007/s11023-024-09674-0.
[76] M. Galvin and D. Gil, “Interview, ‘the Currency of power is increasingly becoming science and technology,’” *Issues*, pp. 36–39, 2025.
[77] R. Zhong, “As Trump Targets Research, Scientists Share Grief and Resolve to Fight,” *New York Times*, Boston, Feb. 18, 2025.
[78] Reuters and IPSOS, “Reuters/Ipsos Poll on Executive Orders, Roper #31122349 Question 31122349.00002, Version 1.” doi: 10.25940/ROPER-31122349(accessed February 21, 2025).