# The Necessity of a National US AI Governance Policy

Matthew Griggs

University of Arkansas

Bachelor of Science and Business Administration in Management Information Systems

December 2024

Advisor: Dr. Mary Lacity

**Abstract**

Artificial Intelligence is transforming industries and reshaping daily life at an unbelievable level. However, its unregulated development poses noteworthy ethical challenges, including bias, lack of accountability, and unintended societal consequences. These issues raise an urgent question: *Should the responsibility for ethical AI development fall solely on organizations, or does it necessitate governmental oversight?* This paper argues that the United States must establish a comprehensive national AI policy to ensure the ethical and responsible advancement of AI.

The research begins by analyzing the ethical principles frameworks of leading AI companies and other regulatory standards, identifying shared themes. The paper presents important and frightening ethical issues and threats that currently face AI, especially in high-risk cases of AI failures where ethical breaches resulted in harmful biases, privacy violations, and discriminatory outcomes. In response, this paper calls for the creation of a national AI policy in the United States, based around existing efforts like the European Union's AI Act. A national policy would establish clear ethical standards, enforceable regulations, and mechanisms to hold organizations accountable for AI deployment. While companies play a crucial role in the ethical development of AI, government intervention is necessary to establish enforceable standards and protect societal interests.

*Keywords*: Ethics, artificial intelligence

# Table of Contents

## I. Introduction

Artificial Intelligence is arguably the most important technology of our time. It is continuously changing industries and how we live, work, and interact with the world around us. In virtually every industry, AI is driving new discoveries, solutions, and technologies such as improving disease diagnosis, detecting fraud, and optimizing logistics. However, this exponential innovation is not without challenges; among the most urgent of these concerns are ethical issues, such as bias, accountability, and potentially even larger and more dire consequences. AI is going to continue to be a larger and larger part of our lives, making it increasingly important to understand these ethical issues and prevent them. As AI continues to evolve, ensuring it follows proper ethical principles is essential to its development and to society as a whole (Zhang). Over the last decade, AI has advanced exponentially, faster than virtually any could have predicted. Today, technologies like self-driving cars and AI-powered medical tools that we once thought were simply works of fiction are becoming a reality. However, with any technology that is very quickly implemented into the daily lives of individuals overnight, there are risks. In several cases, AI systems have been launched without the proper regulations or safeguards in place, leaving gaps where ethical issues can, and often do, arise.

This paper delves into the relationship between ethics and AI development, and the importance of the synthesis of the two. The key research question this paper aims to answer is: *Who should be responsible for the ethical creation of AI: organizations or governments?* By analyzing ethical frameworks of some of the largest and most important companies in the AI revolution and several case studies involving failures of AI ethics, this research will discover how ethical considerations and the lack of them influence AI's impact on society. The goal for this research will be to determine whether the responsibility of ethical AI development can be

trusted with organizations, or if it is too risky for organizations to handle on their own, necessitating governmental intervention.

Policy makers are likely aware of the public perception of AI, shaped by movies, books, and other forms of media: often reflecting a mix of excitement and unease. Popular films such as *I, Robot* aim to understand and shed light on AI's potential to exceed human control, whereas dystopian horror tales like Harlan Ellison's *I Have No Mouth, and I Must Scream* warn of AI defying humanity due to faulty creation and implementation of the technology (Ellison, 1967). Even more optimistic portrayals, such as a movie like *Wall-E*, raise questions about the relationship between AI and humans, and how societal priorities could begin to shift with more advanced AI. These tales, whether hopeful or cautionary, contribute to a public opinion that shows both the possibilities and worst outcomes of AI. They further push the discussion of the regulation of this powerful technology, particularly as it becomes more and more embedded in society.

It is worth noting that, while AI technology may never reach this point, these stakes could become even larger and more important when dealing with the creation of Artificial General Intelligence (AGI). Unlike the artificial intelligence we know today that is designed to handle specific tasks and is unable to reason on its own, AGI would be capable of reasoning and making decisions across virtually every industry. This level of autonomy would deeply impact daily life in a way that, if unregulated and unethically trained, could potentially lead to some of these worst-case scenarios. Many would be skeptical to believe that this is a realistic possibility, however, companies leading the AI revolution, such as OpenAI, have a "[broad] goal of advancing artificial general intelligence" (OpenAI, 2024).

The AI we have today presents us with challenges still, such as opaque decision-making or systems that reinforce harmful biases. If these problems were not only unfixed, but also implemented further, it could spell something disastrous. This disastrous evolution of AI could lead to job displacement, discrimination or underrepresentation of certain groups of people, and a public distrust of new technologies (Ferrerra, 2023). Therefore, it is vital to develop ethical frameworks that are flexible, enforceable, and prepared for the continued exponential growth of artificial intelligence.

Ethical principles form the foundation of responsible AI. Frameworks like the European Union's AI Act and guidelines from organizations such as Harvard's Berkman Klein Center focus on key values, such as privacy, transparency, and fairness. These principles aim to prevent bias, ensure accountability, and build public trust in AI systems; yet turning these ideals into practice is not as simple as it seems. It is extremely important for AI developers to incorporate thoughtful strategies to integrate and intertwine ethics into every stage of AI development. By explaining common threads in ethical frameworks and studying real-world examples of the misuses of AI, this research will aim to show how ethics can shape a better and safer future for AI in society. The intention is to argue that ethical frameworks are required in ensuring the proper and safe AI creation; something that may be too important to leave simply to the organizational level.

## II. Commonalities Among Companies' AI Ethical Principles Frameworks

The recent achievements in artificial intelligence have given light to a need for the development of AI ethics and regulations to ensure AI systems operate fairly, transparently, and promote trust and accountability on a global scale. There is a clear need for ethical principles to

guide companies and organizations who lead the development of AI systems. Across almost every industry, ethical frameworks are being developed to ensure responsible AI development and usage. These ethical frameworks are continually evolving to address the many concerns and issues that arise as the AI systems are developed and progress in their evolution. A leading example of this is The Berkman Klein Center's Principled Artificial Intelligence Project, which was conducted using evidence-based research to provide a thorough and detailed overview of ethical principles that are necessary for ethically sound AI. Data points were collected amassing 36 "principles" documents sourced from governments, corporations, and academic organizations, to provide a comprehensive view of the landscape of AI governance today. Analytical review of the data uncovered eight key themes that emerged as pillars of AI governance these were: privacy, accountability, safety and security, transparency and explainability, fairness, human control, professional responsibility, and the promotion of human values (Berkman Klein Center, 2020).

These principles are shaping the form of several global frameworks, such as the European Union's AI Act, as well as corporate policies from major companies such as Meta, Nvidia, and Walmart. This section will aim to explore these eight principles in detail, with real-world examples that will demonstrate their prominence and significance. Two tables are provided to summarize the key principles and themes across the frameworks, comparing how these principles are implemented in their respective contexts.

| Principle | Berkman Klein Center | EU AI Act | ISO Standards | Meta | Nvidia | Walmart |
|---|---|---|---|---|---|---|
| Privacy | Data minimization, consent | Privacy by design | GDPR compliance | User privacy control | Data anonymization | Encryption protocols |
| Accountability | Traceability, oversight | Audit trails | Developer liability | Bias audits | QA measures | Hiring oversight |

| | | | | | | |
|---|---|---|---|---|---|---|
| Safety and Security | Reliable, secure systems | Stress testing | Risk assessments | Content safety | Vulnerability testing | Fraud safeguards |
| Transparency | Explainability, clarity | Documentation clarity | Algorithm disclosure | User feedback systems | Debugging tools | Transparent hiring |
| Fairness | Bias mitigation | Equity in design | Fairness testing | Bias evaluation | Algorithm testing | Non-discriminatory tools |
| Human Control | Oversight mechanisms | Adjustable systems | Human oversight required | Flagged content review | System parameters | Escalation systems |
| Professional Responsibility | Ethical standards | Ethical compliance | Ethics panels | Developer training | Corporate guidelines | Developer ethics |
| Promotion of Human Values | Societal alignment | Equity, justice focus | Inclusive outcomes | Trust-building tools | Inclusive design | Workforce equity |

*Table 1: Ethical Principles Across AI Frameworks and Corporations*

### 1. Privacy

The Privacy principle pertains to the ethical handling and protection of personal data when using AI systems. Privacy ensures that AI systems respect individuals protected personal information, providing safeguards for protection against unauthorized access and threats. The Berkman Klein Center states that privacy involves informed consent, data minimization, and the right to opt out of data collection. This principle is critical in AI because systems often rely on large datasets that may include sensitive personal information; therefore, data protection is of paramount importance. The EU's AI Act has similar privacy policies, as it mandates compliance with the General Data Protection Regulation (GDPR) which includes requirements for personal data protection in high-risk AI applications (European Commission, 2024). As AI becomes more integrated into daily life, privacy remains essential for building public trust; and by embedding privacy into AI design, developers can create systems that limit unnecessary data exposure.

### 2. Accountability

The Accountability principle ensures that developers and users of AI systems are responsible for the decisions that the AI systems make. Accountability is very important for maintaining ethical standards in AI. It ensures that organizations take full responsibility for the outcomes of their AI systems, including errors, biases, and harm. According to the Berkman Klein Center, accountability involves creating audit trails, enforcing oversight mechanisms, and addressing unintended consequences (Berkman Klein Center, 2020). The ISO standards state that organizations should implement clear processes to trace AI decisions and fix issues. Accountability reinforces the principle that ethical responsibility cannot be outsourced to technology but instead remains with those who design and deploy the technology (ISO).

**3. Safety and Security**

Safety and security safeguards users from harm while protecting AI systems against misuse. Safety ensures that AI performs reliably and consistently under any condition, while security protects against external threats such as hacking or breaches. The Berkman Klein Center identifies this principle as important for user trust and preventing exploitation (Berkman Klein Center, 2020). The EU AI Act shows the importance of risk assessments pre-deployment, particularly for AI used in healthcare, transportation, and critical infrastructure (European Commission, 2024).

**4. Transparency and Explainability**

Transparency and Explainability are distinctive to AI because of the opaque nature of many machine learning models of today. Before AI, computers produced deterministic outputs, meaning that there was a clear understanding and nature of how a computer came to a decision. However, AI produces probabilistic outputs that are often coined to have a “black box” nature,

where individuals would not be able to determine how AI came to a decision. Transparency requires that organizations clearly communicate an AI system's purpose and limitations, while explainability ensures that decision-making processes can be understood by stakeholders. The Berkman Klein Center states these principles are essential for building trust, as "early experience has already proven that it's not always clear when an AI system has been implemented in a given context, and for what task" (Berkman Klein Center, 2020). The EU AI Act enforces algorithmic transparency for high-risk systems, whereas ISO Standards advocate for clear documentation of AI decision-making and functionality (ISO; European Commission, 2024). Companies implement this principle in various ways as well; Meta, for example, uses user feedback systems to increase explainability, whereas Nvidia provides effective debugging tools for better understanding of their system (Meta, 2024). These practices show the necessity of addressing the traditional "black box" nature of AI models, and how to improve on them to create a better relationship between AI and the public.

**5. Fairness and Non-Discrimination**

Fairness and Non-Discrimination pertains to minimizing "algorithmic bias," which is "the systemic under- or over- prediction of probabilities for a specific population" (Berkman Klein Center, 2020). Fairness is an important principle, aimed at preventing AI from pushing biases and stereotypes. The Berkman Klein Center underscores the need to identify and eliminate discriminatory practices in both AI training and operational algorithms. ISO Standards state the importance of designing systems that promote equity and prioritize fair treatment for all, with specific importance needed surrounding the datasets the AI is trained upon (ISO). The EU AI Act mandates fairness testing for AI applications, especially those that are "high-risk" like hiring, lending, or healthcare applications (European Commission, 2024). Corporate efforts align

with these policies as well, for example: Walmart designs its hiring systems to ensure equal opportunities and attempts to mitigate bias in the hiring process (O'Connor, 2023). These measures emphasize the importance of fairness as a technical and ethical priority for corporations and policymakers, as a biased AI application could be detrimental to society.

### 6. Human Control

Human Control can essentially be defined as maintaining human oversight and control over AI systems. This is particularly important to maintain proper goals and prevent massive consequences for organizations and governments alike. The Berkman Klein Center advocates for the inclusion of mechanisms that allow human intervention in AI decision-making, especially in high-risk contexts. ISO Standards give a fairly similar recommendation, recommending that AI systems include parameters that can be adjusted by users, allowing them to modify and override outputs. The EU AI Act requires human oversight for all high-risk contexts, ensuring AI systems do not operate autonomously in their own interests. In practice, Meta includes human review processes for flagged content, and Walmart integrates their chatbots with effective escalation capabilities to send complex questions and tasks to human representatives (Berkman Klein Center, 2020; European Commission, 2024; ISO; Meta, 2024; O'Connor, 2023).

### 7. Professional Responsibility

The principle of professional responsibility describes the ethical obligations developers and organizations need to have to prioritize the well-being of the user. The Berkman Klein Center explains the importance for developers to adhere to ethical codes and maintain integrity in their work. ISO Standards recommend incorporating professional responsibility into corporate policies, and ethics training for developers so they understand the small intricacies of ethical

development. The EU AI Act has introduced ethics panels to oversee compliance during the development of AI systems, this ensures the proper, ethical development and training of AI. Corporate practices reflect these ideas as well, with many companies now requiring its developers to undergo ethics training, and many internal policies to ensure ethical AI practices (Berkman Klein Center, 2020; European Commission, 2024; ISO; Meta, 2024; O'Connor, 2023).

**8. Promotion of Human Values**

The promotion of human values principle refers to the way in which the AI was developed and designed, and the need for it to be aligned with human and societal values. According to the Berkman Klein Center, this principle is one that will become increasingly important as technology advances, with them saying: "...particularly if we begin to approach artificial general intelligence, the imposition of human priorities and judgment on AI is especially crucial" (Berkman Klein Center, 2020). The EU act advocates for a human-centric approach to AI development, saying that AI systems should be designed with European values in mind and that AI should serve as a tool for people. ISO defines this as "non-maleficence," meaning that AI systems should "avoid harming individuals, society, or the environment" (ISO).

**What is distinctive about AI that warrants a national policy?**

Many of these principles can apply broadly to various sectors and technology, like privacy, fairness, and accountability. However, some of these principles are fairly unique to the development of AI, particularly transparency and explainability. The table shown here gives a good visual understanding of this, providing a good representation of principles that are unique to AI, not unique to AI, as well as explaining which principles are more amplified by AI.

| Theme | Distinctive to AI? | Examples Across Frameworks |
|---|---|---|
| Privacy | Common to all tech | GDPR compliance in the EU AI Act; anonymization practices at Nvidia; privacy-by-design policies at Meta. |
| Accountability | Common to all tech | Audit trails in ISO standards; developer liability in the EU AI Act; algorithmic bias audits at Meta. |
| Transparency & Explainability | Distinctive to AI | Explainable AI requirements in the EU AI Act; debugging tools at Nvidia; transparent hiring systems at Walmart. |
| Fairness | Common but amplified | Bias mitigation strategies at Nvidia; fairness testing for hiring tools at Walmart; equitable outcomes in ISO. |
| Safety | Common but amplified | Risk assessments in EU AI Act; stress testing at Nvidia; fraud safeguards in Walmart's systems. |
| Human Control | Distinctive to AI | Oversight requirements in EU AI Act; adjustable parameters at Nvidia; human review protocols at Meta. |
| Professional Responsibility | Common to all tech | Ethical codes in ISO standards; developer training at Nvidia; ethical review panels at Meta. |
| Promotion of Human Values | Common but amplified | Inclusive AI models at Nvidia; equity-focused policies at Walmart; trust-building efforts at Meta. |

*Table 2: Key Themes in Ethical Principles*

By examining these frameworks, it is especially clear that ethical principles need to be one of the pillars of any corporation's development of AI. Both morally and practically, there is a need to establish quality and effective AI ethical frameworks for all corporations and governments. The integration of these principles explained across global and corporate frameworks will allow for increased innovation and public confidence around AI, as well as a shared commitment to responsible AI development.

## III. Analysis of AI Failures: Are Ethical Breaches to Blame?

As artificial intelligence penetrates further into our daily lives, failures in AI systems give real insight into the true consequences of ethical mistakes. Technical limitations can contribute to these failures, but ethical breaches often multiply the harm done. This section will examine three

recent case studies: Google Gemini's biased outputs in its image generator, Sports Illustrated's use of undisclosed AI-generated articles, and an AI-generated deepfake audio in a Maryland high school. Choosing only three cases of ethical breaches in AI that have happened in the last few years was rather difficult, as there are numerous examples of ethical dilemmas that could display the misuse of AI. It is important to choose cases that represent a variety of industries and potential outcomes, which is what these cases aim to show. These examples also give a general overview of the current state of ethical breaches in AI and show the need to prevent these failures in the future.

### Case Study #1: Google

In 2024, Google faced severe backlash for its AI image generator, part of their Gemini line of AI-driven products, after users discovered that the system produced racially biased outputs and misrepresented historical figures. The controversy started when users noticed that AI depicted figures like Norse Vikings as African American, an Asian woman in a World War II-era military uniform, and a female pope, ignoring historical evidence that would clearly state otherwise (Milmo, 2024).


(Duffy, 2024).

The model also refused, sometimes, to generate certain prompts related to marginalized groups, increasing the perceptions of bias. Google originally launched Gemini's image generator as a tool designed to rival ChatGPT's DALL-E and was, allegedly, cutting-edge. However, researchers and users identified an obvious issue with the system: its reliance on biased training data. Critics argued that Google's training process failed to correctly address these biases, causing outputs that perpetuated negative stereotypes (Milmo, 2024). Also, the system's decision-making process was opaque, leaving users with almost no understanding and explanation as to why certain outputs were generated and why some were denied. There needed to be some semblance of explainability to help users understand why the AI made its decisions. Public criticism escalated even further when academics and acceptance groups showed the harm of these biases and the tool itself. Google decided to then remove the feature to update it, acknowledging in a statement that the image generator "did not meet our high standards for

inclusivity and accuracy… and [have] temporarily paused its ability to generate images of people while we make necessary updates" (Google, 2024).

This failure shows severe ethical lapses in fairness, transparency, and accountability. The use of flawed training data showed a lack of correct fairness testing, while the opacity of the system took advantage of user trust. Additionally, Google's delayed response to public concerns showed poor accountability. This case shows the necessity for proper and intensive fairness protocols, transparent decision-making processes, and proactive engagement from company leadership in the development of AI systems.

**Case Study #2: Sports Illustrated**

In November 2023, Sports Illustrated was accused of publishing AI-generated articles under fictional writers, causing a debate over transparency and accountability in journalism. An investigation revealed that several articles by the company were written by fake authors with AI-generated profile photos and biographies. The content of the articles themselves wasn't flagged as incorrect or inaccurate, but many argued that the lack of transparency completely violated the trust in the publication and ruined their reputation, as well as raising questions about ethics in the role of AI in media. Sports Illustrated defended itself by stating that the use of AI was limited to generating drafts for these articles, which were then reviewed by human editors. However, media and industry professionals alike still criticized the company for failing to clearly disclose its AI usage. According to Sports Illustrated, the articles were produced by AdVon Commerce, a third-party company contracted to write and edit these articles. Allegedly, they assured Sports Illustrated that they were written and edited by humans, which was proven to be false. Sports Illustrated later said they "are removing the content while our internal investigation continues and have since ended the partnership [with AdVon Commerce]" (Bauder, 2023).

This case shows the critical ethical failures in transparency, accountability, and professional responsibility within journalism. The use of AI-generated content without proper notice causes the public to severely distrust those that used the AI content, which again, shows the necessity for strict ethical guidelines governing AI. It is especially concerning that the publication itself did not even realize that AI-generated content was being used in their articles, and that a publishing company was able to deceive them easily. In journalism, it is key for the public to be able to trust and reliably depend on the information they receive, and the use of AI-generated content should be specifically noted and made aware for individuals to see. In this case, however, it seems the larger issue was the deception of the publishing company towards Sports Illustrated, not so much the intended deceitfulness of the magazine to get away with using AI-generated content.

**Case Study #3: Eric Eiswert Deepfake**

In April 2024, a high school principal in Maryland, Eric Eiswert, became the victim of a malicious deepfake audio attack. For clarification, deepfakes "are videos or audio recordings that manipulate a person's likeness" (InternetMatters.org, 2024). In this case, an AI-generated recording, fabricated to mimic the principal's voice, was distributed, containing racist remarks that he did not make himself. This dishonest audio clip was created by the Athletic Director of the school, Dazhon Darien, after Eiswert criticized Darien's performance at work (Finley, 2024). The clip was spread across social media, leading to severe personal and professional consequences for the principal, including threats to his personal safety and administrative leave from his position. This incident demonstrates the growing ethical and societal concerns posed by deepfake technologies. The ease of some form of content so convincing yet entirely fabricated highlights the concern around the potential for misinformation and the erosion of public trust

surrounding AI. In response to this incident, experts are calling for enhanced regulatory measures and safeguards to fight the misuse of AI to generate intentionally deceitful content. Some proposed solutions include implementing digital watermarks to verify authenticity, enforcing stricter user verification protocols for AI services, and increasing law enforcement capabilities to fight the intentionally harmful use of deepfake technology.

This case exemplifies the urgent need for comprehensive strategies to deal with the ethical and societal risks associated with AI-generated content. This is a real threat for virtually everyone around the world, as this technology is available and those who have immoral intentions have access to it. As expert Emilio Ferrera explains it, “the widespread use of such biased AI systems can entrench discriminatory narratives and hinder efforts toward equality and inclusivity” (Ferrera, 2023). There needs to be a clear understanding that technological advancements should not compromise individual rights and the trust of the public.

| Principle | Google | Sports Illustrated | Eric Eiswert Case |
|---|---|---|---|
| Privacy | ☑ | ☑ | ☐ |
| Accountability | ☐ | ☐ | ☐ |
| Transparency | ☐ | ☐ | ☐ |
| Fairness | ☐ | ☑ | ☐ |
| Safety | ☐ | ☐ | ☐ |
| Human Control | ☐ | ☐ | ☐ |
| Professional Responsibility | ☐ | ☐ | ☐ |
| Promotion of Human Values | ☐ | ☐ | ☐ |

*Table 3: Case Study Principle Violation Summary*

These three case studies reveal patterns of ethical lapses in AI systems across industries. The reliance on biased training data from Google shows the need for strict fairness testing and bias mitigation strategies. Transparency also is a proven challenge, as shown with Sports

Illustrated's lack of disclosure surrounding its AI-generated content, however, in this case, they seem simply negligible and uneducated. Yet, this still shows the reasoning for the public's distrust of AI applications. Accountability is a common theme in all these cases, as organizations often respond reactively to public criticism, instead of addressing the ethical risks that their technology has created.

Clearly, something needs to be done on a grand scale to stop these problems from occurring. It is important to note that these cases are only three of many that have occurred in the last few years. AI-driven surveillance, how prone AI is to discriminated populations and under-representing them, and the potential loss of jobs, are all real and serious threats because of AI, making it critical to regulate on a national, or even global, level (Federspiel, 2023).

## IV. Call for a National US Policy on AI

The continuous pattern of ethical lapses in AI implementation and creation proves the need for a strict and enforceable national policy. Private sector efforts have shown progress in advancing responsible AI practices, but they would simply not have the same effect and value as a national policy. The limits of relying solely on the private sector can clearly be seen in the case studies previously analyzed. For example, the clear lack of accountability and transparency by Google in their Gemini image generation tool allowed incorrect and unethical training to be done to its AI, allowing for clear breaches of user trust. Sports Illustrated's failure also revealed a lack of accountability, as the company deflected fault and responsibility of the issue onto the publishing company, AdVon. These examples prove that voluntary corporate frameworks are only somewhat effective, since they are still not enough to protect public interests when businesses and corporations prioritize the speed of their innovations over their ethical obligations.

A national policy would provide a cohesive and standardized framework, ensuring that AI development aligns with the betterment of society and makes potential risks less likely. This would also create clear standards that are easy to enforce, based on the key principes laid out previously in section two. This policy could, for example, require organizations to conduct pre- and post-deployment risk assessments for AI systems, as seen in the European Union's AI Act (European Commission, 2024). These assessments would force companies to do proper and correct fairness testing, bias mitigation strategies, and transparency protocols. This would ensure that AI systems are evaluated for ethical compliance throughout their entire lifecycle. By holding organizations accountable for the outcomes of their AI systems and the issues they create, a national policy would incentive companies to prioritize ethical practices for their AI development.

The role of government oversight is vital, especially in applications where failures have severe consequences for the public. For companies, a national policy would offer the benefit of both clarity and predictability. With the lack of standardized regulations, organizations face uncertainty about how to align their AI practices with ever-changing ethical expectations. A national framework would provide clear benchmarks for compliance, letting companies continue innovating at the same rate, just within defined ethical boundaries. This clarity would reduce the risk of reputational damage for companies and legal liabilities, as seen in the Sports Illustrated case study. Furthermore, compliance with a national policy would signal a company's commitment to responsible and ethical AI.

Ethical frameworks also play an important role in minimizing AI ethical failures and fostering innovation. These frameworks, like those outlined by the ISO and the European Union, provide key foundational principles for risk management. However, without the enforcement of

these frameworks, their impact remains fairly limited. A national policy would integrate these frameworks into regulatory requirements, creating a baseline for ethical AI development for all industries. By enforcing and requiring adherence to these established principles, this policy could ensure that ethical considerations and needs are being embedded into every stage of AI development.

Public trust in AI technologies depends on transparency, proactive accountability, and clear, demonstrable fairness. A national policy would require developers to disclose key information about what their system does, the decision-making processes of its system, and its limitations. This transparency would allow users to make educated decisions and hold organizations accountable for faulty AI systems.

Minimizing AI failures requires an approach from several different perspectives that combine both innovation in the private sector and strong public sector governance. Companies must remain proactive in adopting ethical frameworks, but government oversight is fundamental for enforceability and consistency of the framework. A national policy for AI would be the foundation of this governance, providing the tools needed to address future challenges and prevent future harm. Through prioritizing ethical development and aligning AI systems with societal values and ethical frameworks, this policy would foster innovation while protecting the ideals of the general public.

The responsibility for ethical AI development cannot rest only with the private sector, there needs to be governmental oversight. The creation and enforcement of a national AI policy is necessary to address the limitations that voluntary frameworks have and ensures that AI technologies align with public interests and the betterment of society. By establishing clear

standards for fairness, accountability, transparency, privacy, and safety, the government can play a pivotal and extremely important role in shaping the future of AI development.

Despite the great benefits that a national AI policy may have, it would be foolish and naive to believe that a significant decision such as this would be perfect and flawless. One must consider the pushback many companies would have against this, as it would undoubtedly make AI development more difficult and less rapid. Some would most likely argue that the United States, with the implementation of a national AI policy, would be falling behind other global superpowers due to a need to follow strict ethical guidelines that would potentially limit innovation. The ideal policy would not hinder innovation, but many would see this is a foregone conclusion no matter the benefits of the policy. These individuals and organizations who believe this could be correct in this assumption; as there is no real understanding of how great of an hinderance a national AI policy could have on the speed of technological AI advancement. However, what remains to be true is the risks that could occur from failing to adhere to strict ethical guidelines. Intergovernmental bodies such as the OECD recommend that AI development follow similar ethical principles to the principles defined by the Berkman Klein Center and, furthermore, advocate for international co-operation in the ethical development of AI (OECD, 2019).

## V. Conclusion

The exponential integration of artificial intelligence into almost every aspect of society has brought exciting opportunities for innovation and efficiency, but equally as detrimental ethical vulnerabilities and risks. This thesis has explored how ethical lapses, including failures in accountability, privacy, transparency, and safety, have amplified the risks of AI adoption. After a thorough analysis of ethical frameworks and case studies, it is increasingly clear that addressing

these risks is crucial for ensuring AI systems benefit the human race, instead of being to its detriment.

The case studies show significant patterns of ethical shortcoming regarding AI; the absence of proper accountability, fairness considerations, and transparent AI systems contributed to substantial harm for individuals, breaching their trust and the company's reputation. Sports Illustrated's failure demonstrated how ignorance and a lack of responsibility could lead to this exact outcome. Similarly, Google's reliance on biased training data shows off the dangers of discriminatory AI systems, while Sports Illustrated's negligence showed how easily a company can be deceived through AI-generated content. These examples show that, while there may be technical issues in all of these failures, the ethical lapses are what truly intensify and inflame their impact.

The need for a comprehensive, consistent, and enforceable national policy on AI ethics is the central point to understand from this research. While private sector initiative and recommendations have made progress in advancing responsible and ethical AI, their clear issue and shortcoming is that they are voluntary and often inconsistent between organizations. Hence, a national policy could provide a truly standardized framework for addressing key ethical principles. This policy would involve a requirement to conduct risk assessments, fairness testing protocols, and transparency in their AI systems and their decision-making, ensuring clear expectations for ethical development of AI. Government oversight would ensure that these standards are uniformly applied, maximizing consistency and reducing the risk of widespread AI implementation.

For the future, the challenges of ethical AI development will only increase as technologies like AGI become more than fiction. AGI's potential to reason and make decisions

for itself amplifies the necessity of ethical frameworks. Ensuring that AI development aligns with the priorities of society will require collaboration between governments, researchers, and industry leaders; something that might be too difficult of a task to implement. Governments must take a leading role in the ethical development of AI through policies that follow the key principles of privacy, accountability, safety/security, transparency, fairness, human control, professional responsibility, and the promotion of human values. Public trust in AI depends on the consistent application of these principles, along with clear disciplinary actions for when failures occur. At the same time, governments should invest in education and research to ensure that they have the knowledge and tools to handle the complex environment of AI governance.

To conclude, the ethical development of AI must become a shared priority for both governments and organizations. By integrating clearly defined ethical frameworks, enacting policies with clear enforcement rules, and promoting collaboration between industries, we can ensure that AI technologies are deployed ethically and responsibly. The future path requires innovation, caution, and an unwavering commitment to aligning AI systems with the ethical principles that are best suited to ensuring AI is developed with the best intentions for humans. Through this, the future of AI can be one that advances humanity and prospers innovation and creativity, while protecting the values and principles that define us.

## Works Cited

Bauder, D. (2023, November 29). *Sports illustrated is the latest media company damaged by an AI experiment gone wrong.* AP News. apnews.com/article/journalists-ai-counterfeit-writers-479cc3869c0638df5bbb26d4b1e4f18f.

Berkman Klein Center for Internet & Society. *Principled Artificial Intelligence Project: Developing Trends and Best Practices for AI Governance.* Harvard University, 2020. cyber.harvard.edu/publication/2020/principled-ai.

Ellison, H. (1967). *I Have No Mouth, And I Must Scream*. IF: Worlds of Science Fiction.

European Commission. *AI Act.* European Union, 2024. digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

Ferrara, E. (2023, December 26). *Fairness and bias in Artificial Intelligence: A brief survey of sources, impacts, and mitigation strategies.* MDPI. www.mdpi.com/2413-4155/6/1/3.

Finley, B. (2024, April 30). *Deepfake of principal's voice is the latest case of AI being used for harm.* AP News. apnews.com/article/ai-maryland-principal-voice-recording-663d5bc0714a3af221392cc6f1af985e.

Federspiel, Frederik, et al. "Threats by Artificial Intelligence to Human Health and Human Existence." *BMJ Global Health*, vol. 8, no. 5, May 2023, p. e010435, gh.bmj.com/content/8/5/e010435, https://doi.org/10.1136/bmjgh-2022-010435.

Internet Matters. *What is a deepfake?* (2024, November 13). www.internetmatters.org/resources/what-is-a-deepfake/.

ISO. *Building a responsible AI: How to manage the AI ethics debate.* International Organization for Standardization. www.iso.org/artificial-intelligence/responsible-ai-ethics.

Milmo, Dan, and Dan Milmo Global technology editor. "Google Pauses AI-Generated Images of People after Ethnicity Criticism." *The Guardian*, 22 Feb. 2024, www.theguardian.com/technology/2024/feb/22/google-pauses-ai-generated-images-of-people-after-ethnicity-criticism.

O'Connor, N. (2023, October 17). *Our responsible AI pledge: Setting the bar for ethical AI.* Walmart. corporate.walmart.com/news/2023/10/17/our-responsible-ai-pledge-setting-the-bar-for-ethical-ai.

OECD. "Recommendation of the Council on Artificial Intelligence." Oecd.org, 2019, legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

OpenAI. *Evaluation of OpenAI o1: Opportunities and Challenges of AGI.* OpenAI, 2024. arxiv.org/pdf/2409.18486.

Responsible AI. *AI at Meta.* (2024). ai.meta.com/responsible-ai/.

Thorbecke, C., & Duffy, C. (2024, February 22). *Google halts AI Tool's ability to produce images of people after backlash.* CNN. www.cnn.com/2024/02/22/tech/google-gemini-ai-image-generator/index.html.

"Improve the grammar of this paper, especially with the word selection, to sound more professional and studious" prompt. ChatGPT, 22 Nov. version, OpenAI, 29 Nov. 2024. chat.openai.com/chat.

"Improve the writing of this paper, particularly the sentence structure and flow of the paragraphs" prompt. ChatGPT, 22 Nov. version, OpenAI, 29 Nov. 2024. Chat.openai.com/chat.

"Improve the introduction and conclusion's flow between paragraphs, and the word choices to sound more intelligent" prompt. ChatGPT, 22 Nov. version, OpenAI, 11 Dec. 2024. Chat.openai.com/chat.

"Improve on the abstract's word choices to align more closely with the message of the paper" prompt. ChatGPT, 22 Nov. version, OpenAI, 14 Dec. 2024. Chat.openai.com/chat.

Zhang, Baobao, et al. *Artificial Intelligence: Ethics, Risks, and Strategies.* arxiv.org/abs/2105.02117.