The Ethics of Artificial Intelligence in Software Engineering: Balancing Innovation with Responsibility

A Research Paper Submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Abram Johan

Fall 2024

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

"The rapid expansion of AI is often seen as inevitable, but this perspective overlooks the critical need for intentional, ethically grounded regulation." - IBM

I. Introduction

According to Forbes, "Artificial intelligence could add as much as $15.7 trillion to the global economy by 2030" (Trefis Team, 2019). In today's rapidly advancing technological landscape, artificial intelligence (AI) is revolutionizing various industries, including software engineering. It is being used to streamline important processes such as testing, debugging, scripting, and general code creation. AI boosts productivity for students, engineers, and everyday users. Its applications in customer service, home automation, and insurance demonstrate its widespread benefits. With applications in customer service, thermostats, insurance, and many more everyday commodities, it is apparent that AI can be very beneficial. Its financial impact on the world can be seen below in Figure 1.

Despite its advantages, the rapid rise of AI and its integration in major systems raises ethical concerns. Its massive influence and potential have led experts to question its underlying systems, with concerns about algorithmic bias, transparency, and accountability. If these ethical issues are not addressed, unchecked usage of AI could cause some significant harm. Misuse of the technology could lead to flawed education, privacy violations, large scale software failures, and a general loss of trust in AI, disproportionately impacting marginalized groups.

This research examines the AI landscape to identify relevant institutions, policies, and pressures. The research will follow Geels' multi-level approach, focusing on a certain few key ethical concerns in the space. It aims to pioneer a shift in how people around the world create, display, and use artificial intelligence.

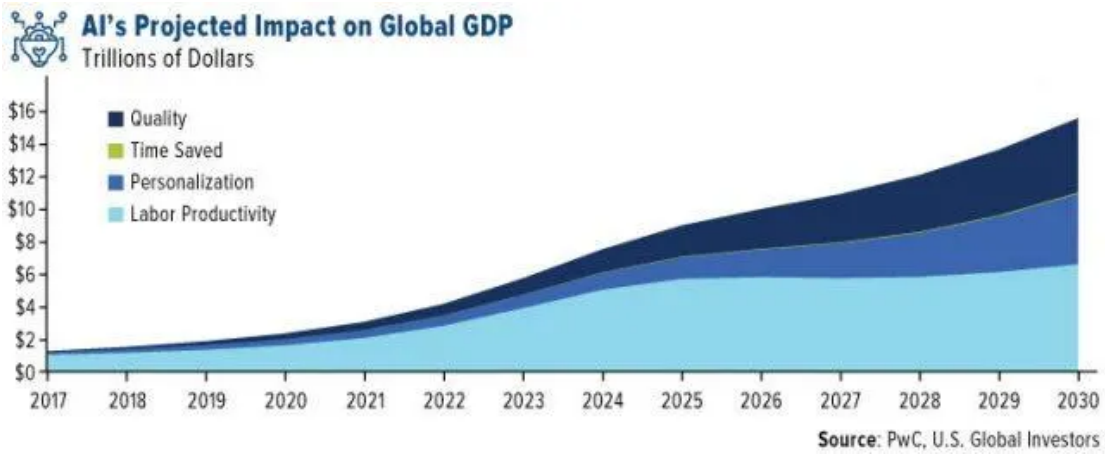

Figure 1. AI's impact on global GDP

II. Problem Definition

AI's growing impact in the last few years is undeniable. With the advent of generative AI and the great strides made with machine learning, many existent systems are changing rapidly and drastically. It's impacting students, the job market, and even the everyday lives of most users. The issue is it is unclear how these different technologies are operating. It's unknown how the systems use user data, whether they're trained accurately, if it's secure etc. There are many ethical concerns, namely the concepts of algorithmic bias, lack of transparency, and potential privacy violation. Ethical considerations should be heavily involved in the conception, design, and creation of novel technologies, but it is not exactly required of the AI technology that is being created currently. Neglecting these considerations can lead to significant harm for both developers and end-users, which could contribute to the worsening of existing societal inequalities.

Existing ethical computing guidelines are imperfect in both content and enforcement. To strive for a solution, it is important to start with the root of the problem. This will be centered around the analysis of the AI technological landscape to identify the relevant players that impact

the integration of ethics into AI. Specifically, this research aims to better understand how large sociotechnical systems like AI-driven software engineering evolve over time.

III. Research Approach

The Problem the Source Addresses

To explore the AI realm, the study utilizes Frank W. Geels' work on the "Transformations of Large Technical Systems," which uses a multi-level perspective (MLP) to analyze system transitions. Geels' study of the Dutch highway system offers insights into how technological changes are shaped by interactions between niches, regimes, and landscapes. The multi-level perspective offers a promising approach for examining how AI can be ethically integrated into software engineering, as it allows for a comprehensive understanding of the pressures that drive both innovation and the need for ethical safeguards.

In Geels's paper "Transformations of Large Technical Systems", he advocates for the use of a multi-level perspective approach to evaluate change among large sociotechnical systems(LTS). Geels' case study of the Dutch highway system demonstrates how his multi-level perspective can be used to understand the evolution of sociotechnical systems. Specifically, Geels examines how technological change is driven by interactions between the three different levels: niche innovations, regime shifts, and broader socioeconomic landscapes. His paper also emphasizes that technological transformations cannot be understood in isolation, but they should instead be understood as a complex interaction between actors, institutions, and pressures. In the case of the Dutch highway system, Geels revealed that infrastructure, policies, social practices, and the economic landscape at the time all had a unique impact on the evolution of the system.

As this multi-level perspective considers various impact levels, it is apparent how it could be applied to understand the ethical integration of a large-scale technology like artificial intelligence.

The Research Approach Used

At its core, Geels's research approach is centered around the multi-level perspective(MLP), which essentially breaks technological transformations into the three levels: niche, regime, and landscape. The niche level is at the micro level, where "radical novelties emerge." This level is where the new technologies are being developed, usually shielded from the mainstream market pressures. However, these new innovations could interfere with the systems in place and the status quo. Next, at the meso level, is the technological regime. This level consists of the institutional structures that link together technical artifacts. Lastly, at the macro level is the landscape regime. This level refers to the external socioeconomic or political factors that impact the regime. Geels's approach emphasizes that the interplay between these levels is what makes up technological transformation. It begins with change in the regime level, with the conception of new technology. These technologies are then integrated into the regime. The regime is constantly being impacted by the volatility of the landscape level, deriving its stability from the landscape pressures. The dynamic between all these levels ultimately determines the impact of technological transformations. A more detailed visualization of this process can be seen below in Figure 2.

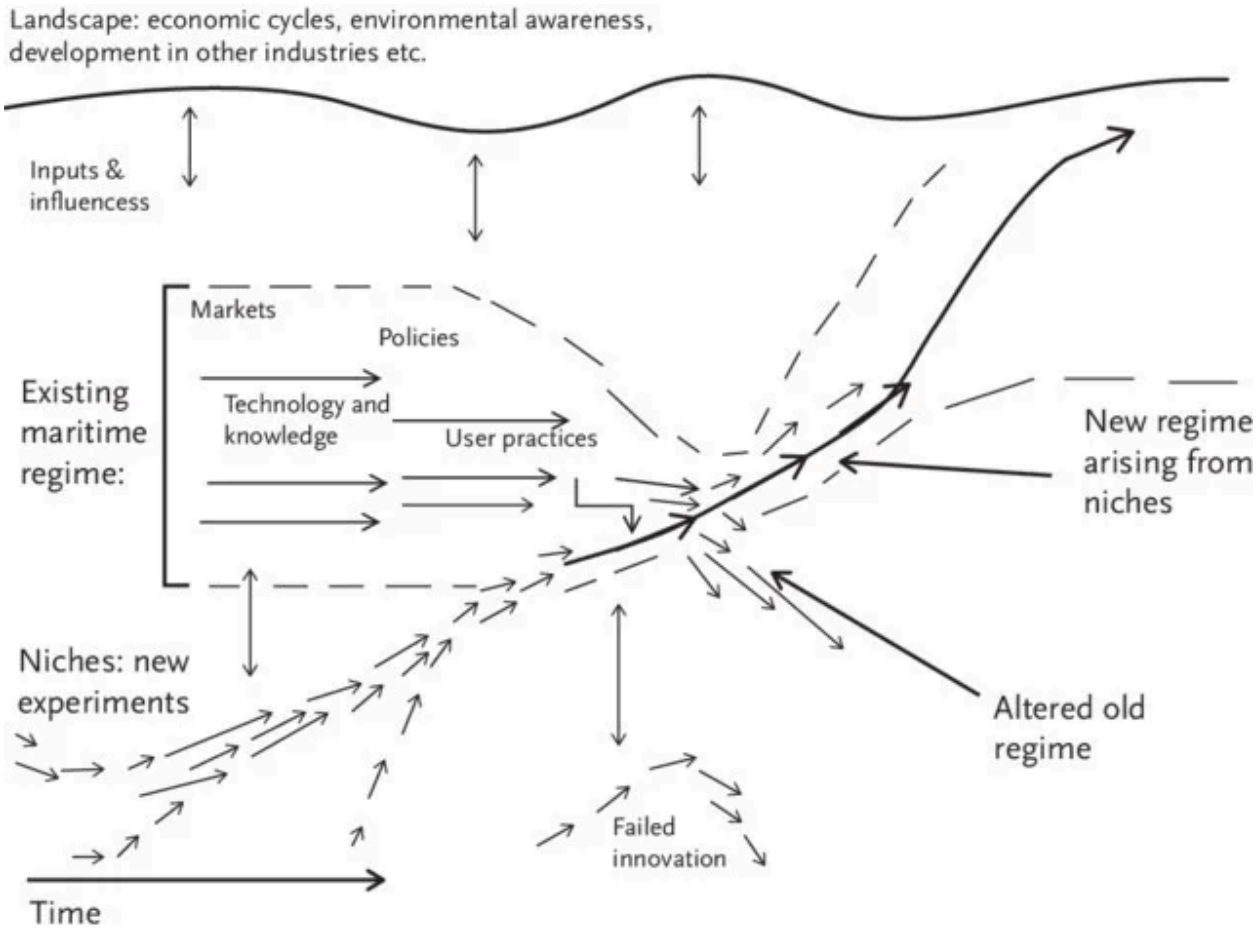

Figure 2. Detailed visualization of how the multilevel approach operates

As shown in the image, there is a general chain of progress when new innovations are brought into the market. An assortment of different regime level factors impact the integration of the technology into the mainstream market, with the landscape pressures impacting the established system. Geels used his multi level approach to evaluate the Dutch highway system, seeking to understand how the roads changed so much between 1950-2000. His analysis consisted of a plethora of relevant material, including the adoption of new traffic management technologies and with changes in policy. He even mentioned the landscape level pressure that resulted from World War II, citing the different legislation that was passed at the time.

Geels' approach effectively captures the complexity of technological transitions by examining interacting factors driving change. With the case study of the Dutch highway system, he was able to convey that point both powerfully and articulately. Applying this to the research, it will first explore the various factors that impact the artificial intelligence regime. After identifying relevant institutions, policies, and other pressures, the paper will highlight key ethical concerns, emphasize the dangers of neglecting them, and loosely articulate where change needs to happen. As to remain within the scope of the project, it will no longer seek to create a framework for ethical AI usage. Instead, it will strive to comprehensively investigate the regime to find as much information about the AI landscape as possible. Using this, the research aims to inspire larger forces to create an enforceable framework that will ethically regulate AI usage.

Conclusion

In conclusion, Frank W. Geels' multi-level perspective approach serves as a great tool to help attain a complete understanding of a technological market. Similar to actor network theory, it requires an understanding of contextual factors. In using his approach, the paper aims to gather a comprehensive understanding of the AI space to determine what pressures are hindering safer, more ethical AI usage. As Geels did in his Dutch highway analysis, this research hopes to use the analysis of the artificial intelligence landscape to pioneer a shift in how people around the world create, display, and use artificial intelligence.

IV. Results

MLP Analysis

Regarding the regulation of artificial intelligence usage, the research revealed a variety of different actors in play. As was expected, the AI landscape is complex and filled with conflicting

parties. From students to developers to large corporations to even the white house, many different actors have their own vested interests with the technology. Consequently, their conflicting pressures are what made the AI regime what it is today. This research will help identify key actors, explain the current regulations in place, detail some of the current ethical issues with AI, speak on the discourse of inevitability, and ultimately push for change in the currently flawed regulations.

Using Geels' multi-level perspective, this paper will first analyze AI's initial integration in society, identifying the primary players involved. Artificial intelligence was 'introduced' at the niche level with its roots generally traced back to the work of Alan Turing, an extremely prominent computer scientist and historical figure. At a 1956 conference at Dartmouth college, computer science researchers from across the US met "to discuss groundbreaking concepts on an emerging branch of computing called artificial intelligence, or "AI." The term "AI" was, in fact, coined at this Dartmouth conference" (NCSL 2023). AI has drastically changed over the years, evolving from simple programs to simulate a game of checkers to modern day generative AI. A general timeline of key AI developments is shown below in Figure 3.

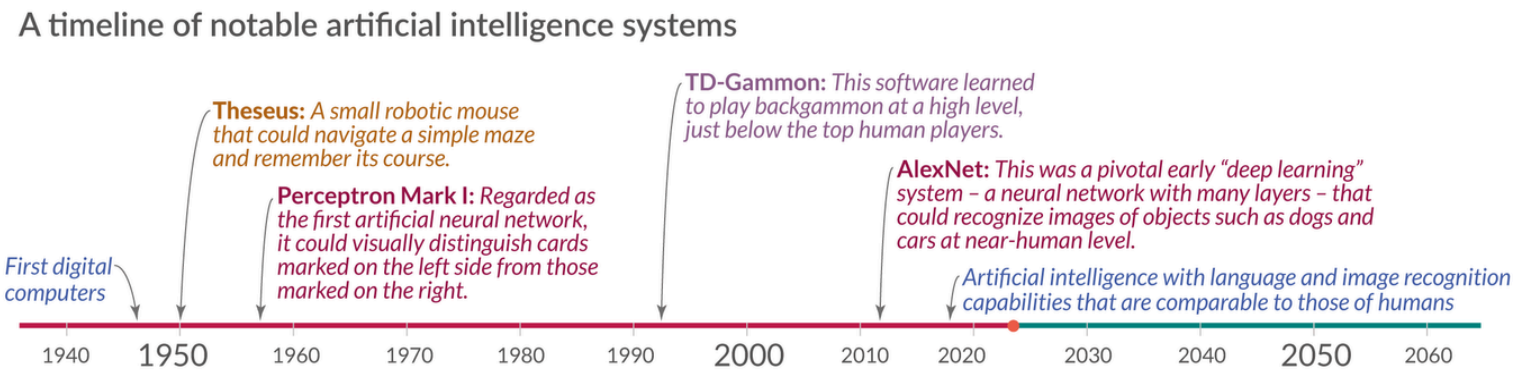

Figure 3. Timeline of notable AI developments

These advancements led us to modern day, where AI is far more capable than before. At the regime level, it appears there are three primary actors: tech companies/developers, academia and research institutions, and industry trade groups. Each of them play an important role in how AI is used and regulated. The tech corporations typically focus on innovation and profitability, often neglecting certain ethical safeguards. Moreover, many of these companies' AI systems rely on usage of user data, which raises privacy concerns. On the other hand, academic/research institutions tend to emphasize more analytical development, actively working to identify potential ethical issues early in the development process. Lastly, the industry trade groups and lobbyists usually push for regulations that favor industry growth, which often conflicts with the general public's desire for more ethical regulation. Together, the three create a dynamic that prevents real regulatory change. The innovative drive to push for rapid advancement generally overpowers the ethical concerns of users, providing for the flawed regulations that currently exist.

At the landscape level, larger societal forces and external pressures decide how AI regulation is enforced. The main actors at the landscape level are the government policymakers, public interest groups, and international regulatory groups. The government is what sets the laws

and policies for the AI space. In other words, the government's regulations are what influence actors in the regime, defining what is considered an acceptable practice. Next, public interest groups argue for more critical restrictions on AI development. This is because they represent the interests of the general public, who naturally desire safer AI systems. International regulatory groups serve to balance the interest of both innovation and public concern. These actors, with their individual interests, all place pressure on the regime to promote more accountability and safeguards in the AI development space. These landscape level influences create an environment that challenges regime actors to adopt more ethical practices.

Ethical Concerns

The rapid rise of AI raises critical ethical concerns, particularly regarding the issues of algorithmic bias, transparency, and accountability. Algorithmic bias occurs when AI systems trained on human sampling data show discrimination, reflecting existing human biases in the training data. One of the most notable examples of this was in 2018, when Amazon's Rekognition faced an external audit. The AI service was sold to law enforcement as a facial analysis system intended to help police find criminals. Though Amazon claimed their service was bias free, the external review proved otherwise. One study conducted by the American Civil Liberties Union(ACLU) showed "28 members of Congress, disproportionately people of color, were incorrectly matched with mugshot images" (Najibi 2018). Later that year, a follow up study by Gender Shades "also showed racial bias against darker-skinned women (31% error in gender classification)" (Najibi 2018). Considering that this product was released by a corporation as large, well-known, and reputable as Amazon, one can only imagine the type of harm an amateur company could do. An image highlighting the variable accuracies can be seen below in Figure 4.

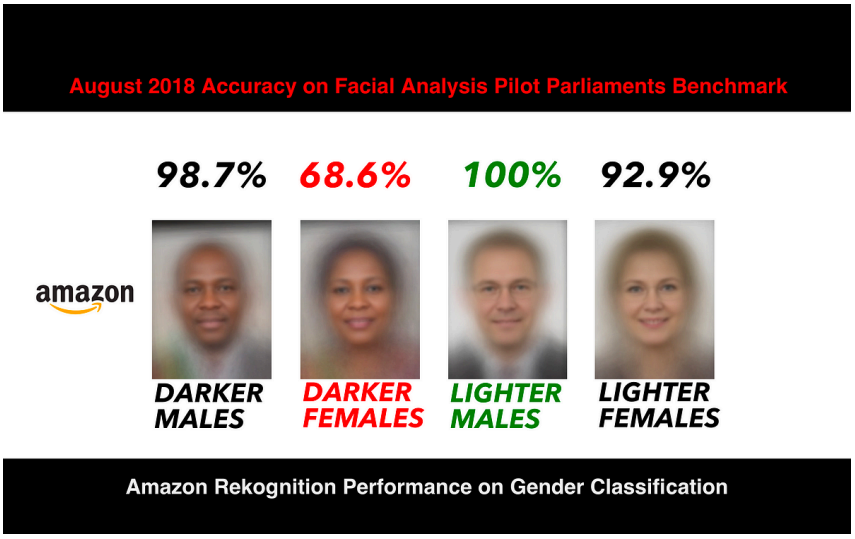

Figure 4. Side by side view of different demographics' accuracy

This same case also brings up questions about accountability and privacy. In the Amazon Rekognition case study, Amazon was aggressively marketing its technology to police, "boasting it could identify up to 100 faces in a single image"(Snow 2018). Similarly, In 2019, the New York Times reported that facial recognition firm Clearview AI scraped billions of images from social media platforms to create a facial recognition tool used by law enforcement. These cases raised major ethical and legal concerns about consent and surveillance.

Transparency and accountability are also essential for ethical AI. AI technologies are typically "black box", which means the underlying code for the algorithms are not known to the public. As the white house describes it, "AI systems are often 'black boxes' where even developers can't fully explain how they reach specific decisions." A lack of transparency makes it difficult to understand why certain outcomes occur, which is especially dangerous in cases like Rekognition. People wrongly identified by the technology could be falsely accused of crimes they did not commit. However, because the algorithm's underlying code is hidden, the reason for

all its faults remain unknown. Mandatory transparency standards for algorithms, particularly in sectors where AI influences life-changing decisions, are urgently needed.

A fundamental concern about all of these issues is the discourse of inevitability surrounding AI. The discourse of inevitability implies that because AI's advancement is so rapid, that it should be accepted. In other words, it suggests that there is nothing that can be done. This narrative can minimize important discussions about AI's ethical implications and protect guilty corporations from taking accountability, dismissing valid concerns as a means to an end. With Amazon Rekognition, executives were able to deny that the algorithm was flawed, and they were able to avoid any consequences. In fact, most people aren't even aware of the scandal at all. When we treat AI's impact as inevitable, we fail to address the many issues that it may be causing. Taking accountability in engineering can help address detrimental issues in advance, providing for a safer end product.

Areas of Change

To move toward ethical AI regulation, there will need to be changes made in the niche, regime, and landscape levels of Geels' multi-level approach. First, at the regime level, tech companies, academic institutions, and trade groups must shift their priorities to include more robust ethical standards in AI development. Major tech companies, which drive most of AI development, currently prioritize financial profit and speed over ethical implementation. This type of development typically ignores key issues such as data privacy and algorithmic bias, and it could result in more harm than good. As a result, companies should consider adopting mandatory transparency measures. For critical tools in industries such as policing, surveillance, and hiring, flaws in AI systems could be extremely detrimental. By making these black-box algorithms more

tested, tech companies can work to foster more trust in their products. Naturally, this would never happen as corporations would never compromise profitability. They would need to face some persuasion from the landscape and other actors within the regime.

Academic and research institutions could be crucial in the push for emphasis on ethical AI development. Universities and research groups already contribute by identifying potential issues with AI early in the development cycle. However, they must use this influence to sway regulatory groups. In collaboration with public policymakers, research institutions can provide data-driven evidence that informs AI legislation. Their expertise could help shape policies that emphasize the importance of transparent, fair AI practices. Industry trade groups and lobbyists should also be pressured to advocate for industry standards that prioritize user protection alongside business growth. By promoting such standards, trade groups could help balance the power dynamics in the current regime, reducing the dominance of tech companies over public interests. However, it is unlikely that these groups would speak up on the issue unless it became a politically contentious topic. They would also need some pressure from the landscape to act on the issue.

At the landscape level, government policymakers, public interest groups, and international regulatory bodies all have a role in reinforcing ethical AI standards. In fact, given its vast influence, this is likely the most important level in creating lasting change. Government policymakers must go beyond flimsy guidelines and focus on creating regulations that prioritize public safety and fairness. In particular, standards for AI transparency need to be codified into law so that companies are required to explain how their algorithms reach decisions that affect individuals. Policymakers should also institute accountability frameworks, ensuring that developers and executives are held responsible for harmful AI outcomes, as seen in the Amazon

Rekognition case. Public interest groups can support this movement by amplifying the voices of affected communities, highlighting the societal impacts of unethical AI usage, and advocating for more stringent privacy and anti-bias laws.

International regulatory groups can also play a pivotal role by encouraging global AI standards that address both innovation and ethical responsibility. As the AI field grows, ethical concerns will transcend national borders, meaning that AI regulations must be as global as the technology itself. Organizations like the European Union have already taken strides in setting comprehensive AI guidelines, but a more unified international stance could strengthen these efforts. With aligned global standards, multinational companies would face consistent regulatory expectations, pushing them to develop safer and more accountable AI products.

As all these levels are fundamentally interconnected, no singular change could create a lasting change. As Geels' asserted, the key to technological transformation is through the multi level approach. Technological giants need some sort of mandate to implement their innovations conscientiously. Naturally, the bulk of the burden is on the government to enact a legislative change. However, to raise the issue for the government to take action, it would require pressure from interest groups, academics, and lobbyists. As such, it requires a large increase of focus and priority on the issue of ethical AI. Ultimately, that is the purpose of this research. It aims to bring awareness to the topic and instruct on the steps necessary to fix this issue.

Conclusion

As artificial intelligence is only getting increasingly widespread, it is imperative that changes be made now. As Geels's work emphasized, technological change does not happen on its own. Rather, it will require involvement from various different fields, institutions, and regulatory

bodies. The first step to eliciting such a movement is by raising awareness on the topic. The majority of consumers use these AI tools and products, completely oblivious to the harmful underlying processes. This paper aims to alert the public of these ethical concerns, identify the involved parties, and describe what a shift to ethical AI would constitute.

As shown in the research, while government bodies have the most direct agency, improving the ethical standards of AI development is a process that will require aid from a number of actors. The most realistic road to progress is by way of raising awareness. By raising awareness on the issue, ethics in AI can become a politically charged topic. With this publicity, research institutions, public interest groups, and lobbyists can work to sway regulatory groups and government bodies. With an official, centralized framework or policy to dictate the usage of artificial intelligence, the major tech corporations that create our computer software, websites, everyday home goods, and more will be forced to clean their acts up. They will be forced to implement their AI technologies with users' interests in mind.

Like all political change, making a change to the current development of AI will be very challenging. With such massive amounts of money vested in this industry, it's not likely that large powers will be swayed. However, the conducted research demonstrates that this topic is large enough to fight for. Our society mustn't fall victim to the discourse of inevitability, which implies that there is nothing that can be done. Instead, it is important to realize what is at stake with massive, potentially dangerous technologies. Failure to implement ethical safeguards could lead to drastic problems as AI continues to grow. It is up to us to incite the change we want to see with AI.

References

IBM. (n.d.). *Shedding light on AI bias with real-world examples.* IBM Think. Retrieved from https://www.ibm.com/think/topics/shedding-light-on-ai-bias-with-real-world-examples

Lohr, M. A., & Kohler, A. (2020). *Racial discrimination in face recognition technology.* Science in the News. Retrieved from https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/

National Conference of State Legislatures. (2023). *Approaches to regulating artificial intelligence: A primer.* Retrieved from https://www.ncsl.org/technology-and-communication/approaches-to-regulating-artificial-intelligence-a-primer#toc6

National Conference of State Legislatures. (2024). *Artificial intelligence 2024 legislation.* Retrieved from https://www.ncsl.org/technology-and-communication/artificial-intelligence-2024-legislation

Roser, M., Ritchie, H., Ortiz-Ospina, E., & Rindermann, H. (2023). *A brief history of artificial intelligence.* Our World in Data. Retrieved from https://ourworldindata.org/brief-history-of-ai

Trefis Team. (2019, February 25). *AI will add $15 trillion to the world economy by 2030.* Forbes. Retrieved from https://www.forbes.com/sites/greatspeculations/2019/02/25/ai-will-add-15-trillion-to-the-world-economy-by-2030/

White House Office of Science and Technology Policy. (n.d.). *Blueprint for an AI Bill of Rights.* Retrieved from https://www.whitehouse.gov/ostp/ai-bill-of-rights/

White & Case LLP. (n.d.). *AI Watch: Global regulatory tracker – United States.* Retrieved from https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-states#:~:text=Laws%2FRegulations%20directly%20regulating%20AI,AI%20albeit%20with%20limited%20application