Bipartisan Policy Center

AI and Ethics

AUGUST 2020

Introduction

Artificial intelligence is changing our lives on a daily basis. From voice recognition to movie recommendations to robo-advisors, many of these services are all around us. AI has great potential to create new opportunities and greatly improve lives, but many ethical challenges that previously existed—like bias, privacy, and power asymmetries—will evolve and can be greatly exacerbated by the emerging uses of AI technologies. Therefore, issues of civil rights and liberties must be front and center in discussions about the development, deployment, and oversight of AI technologies and systems. Ideally, with the right policies and an inclusive approach, AI technologies would not exacerbate but rather mitigate the existing challenges in protecting civil rights and liberties.

Fortunately, AI technologies are not an autonomous force beyond human control. All stakeholders—developers, users, consumers, and policymakers—should have the power to determine how these technologies evolve, how they are applied, and ultimately how they impact people and societies. These stakeholders should work to ensure public policy, regulation, and governance structures are well-designed to meet the challenge. Many ethical questions need to be answered.

Take the example of an AI system used to help allocate hospital resources. The AI system can be trained to diagnose patients and suggest what resources should be allocated to their treatment. The potential for such a system to greatly improve human health is vast. However, many thorny ethical questions arise. Should

the AI system make final decisions about a treatment, or should only humans? How much should the algorithm focus on saving lives versus improving the quality of life for terminally ill patients? What rights should the patient have? What governance structures should the hospital have to oversee the system throughout its life cycle? The AI system can unintentionally become biased because of the data it is trained with or by algorithmic design, so what should be done if the AI system appears to exhibit bias against a protected group that threatens to perpetuate or amplify existing societal inequities?

At its best, the AI system will help improve the allocation of medical resources and can improve health outcomes, lower costs, and ultimately save lives in an inclusive manner. But there are serious risks, such as an AI system picking up and exacerbating human biases, worsening inequities in the health care system, and harming those who are most vulnerable.

Health care is only one area where ethical concerns about AI are being raised. In critical areas ranging from criminal justice to financial services to national defense, people are grappling with their own set of questions. Identifying common themes and differences amongst industries can help guide Congress going forward to ensure a thoughtful and well-tailored approach for promoting AI ethics. The failure of the United States to lead will result in other countries setting global AI ethics standards that might not be aligned with American values.

However, public concerns about emerging technologies are not novel to AI. Pharmaceuticals and automobiles are examples of technologies that have benefited society but also raised ethical concerns. In addressing these ethical challenges, neither denialism nor sensationalism was the correct response. The best policies came about when policymakers consulted the relevant stakeholders and experts, then raised public awareness and put in place thoughtful policy solutions that addressed legitimate concerns.

This is the public policy approach the United States must take for AI ethics. In this spirit, the Bipartisan Policy Center, in consultation with Reps. Will Hurd (R-TX) and Robin Kelly (D-IL), has worked with government officials, industry representatives, civil society advocates, and academics to better understand the major AI Related ethical challenges the country faces. This paper hopes to shed more clarity on these challenges and provide actionable policy recommendations, to help guide a U.S. national strategy for AI. BPC's effort is primarily designed to complement the work done by the Obama and Trump administrations, including President Barack Obama's 2016 *The National Artificial Intelligence Research and Development Strategic Plan*,[1] President Donald Trump's Executive Order 13859, announcing the *American AI Initiative*,[2] and the OMB's subsequent *Guidance for Regulation of Artificial Intelligence Applications*.[3] The effort is also designed to further advance work done by Kelly and Hurd in their 2018 Oversight and Government Reform Committee (Information Technology Subcommittee) white paper *Rise of the Machines: Artificial Intelligence*

and its Growing Impact on U.S. Policy.[4] Our goal through this effort is to provide the legislative branch with potential actions it can take to advance AI building on the work being done by the Obama and Trump administrations.

I. Key Principles

Over the past several months, BPC has conducted a series of roundtables and convenings with experts, academia, industry representatives and civil society organizations to examine concerns about AI and fairness, bias, and privacy. Based on these discussions, BPC has identified the following key principles:

1. The federal government should further fund and encourage research and development projects that address bias, fairness, and privacy issues associated with AI.
2. The federal government should encourage more diversity in AI talent to help mitigate unfair bias and promote fairness in AI practices.
3. The federal government should encourage the development of voluntary standards frameworks to help create shared conceptual foundations, terminology, and best practices for fairness and bias based on a cooperative and multi-stakeholder approach.
4. In promoting ethics and mitigating unintended bias, the regulation of AI should build on existing regulation when possible and be tailored to different use cases using a risk-based approach.
5. AI and privacy should not be conflated, but AI-specific considerations should inform and influence privacy legislation.

The remainder of this white paper is organized as follows. Section II provides a broad overview of AI and ethics to summarize the common foundations identified by BPC in its discussions with stakeholders from industry, civil society groups, government and academia. Subsequent sections describe each of the five key principles listed above, including a brief overview along with specific recommendations the United States can pursue to accelerate and sustain global leadership in AI while minimizing the likelihood of adverse impacts to civil liberties, civil rights, and innovation.

II. Overview of AI and Ethics

In 1956, Computer scientist John McCarthy held a conference where he coined the term "artificial intelligence." According to *AI Magazine*, the experts could not agree on a common definition but classified the field of AI as "the shared vision that computers can be made to perform intelligent tasks."[5] There remains no definitive definition of AI, and the meaning of the term has evolved over the years. For the purpose of this paper, BPC will use a description from the fiscal year 2019 National Defense Authorization Act.[6] That act defines AI as inclusive of the following:

1. Any artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets.

2. An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action.

3. An artificial system designed to think or act like a human, including cognitive architectures and neural networks.

4. A set of techniques, including machine learning, that is designed to approximate a cognitive task.

5. An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision making, and acting.

The rise of AI has great potential to improve quality of life but also raises important ethical issues. Serious solutions to address these issues are necessary to help promote a more just society. BPC will focus on the issues of fairness, bias, and privacy in this paper. BPC will not define these terms, given the lack of consensus on what these terms mean. However, we expect that society will continuously debate how to define these terms in the context of AI. Ensuring these discussions are inclusive is a critical goal because AI ethics should be of concern to anyone who interacts with this technology.

The issue of fairness raises two overarching questions. First, what is a fair use of an AI system? AI systems are ultimately tools that people use. Much like a hammer, an AI system can be used for good or evil. This raises important ethical questions about whether specific use cases for AI are fair. For example, when (if ever) is it appropriate for an employer to use an AI system to monitor employees? Society will have to make value judgements on whether such use

cases are appropriate and what rules should guide them.

Second, how can society ensure human values of fairness are best encoded into AI systems? AI systems need instructions to operate, but any attempt to encode fairness into an AI system will be imperfect, since fairness can never truly be defined mathematically. Ensuring fairness in an AI system is further complicated by the fact that different definitions of fairness can run into conflict. For instance, how should an AI system that makes recommendations on allocating hospital resources to medical professionals balance the desire to save lives versus improve the quality of life for terminally ill patients? The formal definitions and tradeoffs that best reflect human ideals of fairness are not something an AI system can realistically make without human input.

The issue of bias is interrelated with fairness. Algorithms can become harmfully biased (often unintentionally) because of the data they are trained with or the algorithmic design. This can lead to unfair outcomes in a variety of areas including lending, housing, criminal justice, and health care. A police facial recognition technology that is trained with few examples of women or people of color will likely misidentify these protected groups. This can result in deeply unfair outcomes, such as increasing the likelihood that people of color will be wrongfully detained after a burglary.

The issue of bias is further complicated by the fact that it is hard to determine what constitutes bias and whether it was caused by unfair human bias, unrepresentative datasets, or other factors. For instance, an AI algorithm used in a medical setting may exhibit bias against a protected group due to being trained with data reflecting unfair human bias in making medical decisions. However, this might instead be the result of data reflecting legitimate physiological differences, such as women being more prone to breast cancer relative to men.

AI also poses new privacy challenges. While the privacy debate is broader than just AI, specific AI considerations should help guide it. Modern AI systems are heavily reliant on quality data to be effective, which magnifies many existing questions about consumer privacy. Further, AI technologies, such as facial recognition, raise unique privacy concerns that have not existed before. Determining what is meant for something to be private, how to protect privacy, and what tradeoffs should be acceptable between privacy and other objectives is critical. Ensuring good governance structure and a proper regulatory framework for privacy is critical for building trust in AI technologies.

Ultimately, AI ethics should be based on human ethics. AI has merely brought many existing questions about human ethics and values further to the forefront. An inclusive discussion about ethics is necessary to determine what values society wants AI to reflect.

III. Key Takeaways

1. **The federal government should further fund and encourage research and development projects that address bias, fairness, and privacy issues associated with AI.**

Increased federal support for research and development (R&D) on ways to improve fairness, reduce bias, and protect privacy is critical. There are many areas of technical research that are worth pursuing further, such as differential privacy and explainable AI.

Differential privacy is a technique that can help reveal features about a population, while masking which individuals have these features. This technique uses randomness to help preserve anonymity and privacy, but it does sacrifice some level of accuracy. Differential privacy is not a panacea, has its limits, and is not suitable in all cases, but it does have the potential to further improve user privacy in various situations.

Explainable AI is an emerging field in machine learning that aims to bring transparency to some of the most complicated AI systems. Many AI system are often seen as a "black box." An expert might be able to explain how an AI system works in theory but may not be able to explain how the system reached a specific decision at a more granular level, given the complexity in the interactions between the input data and the algorithm.[7] Explainable AI can give people a method to better understand the way an algorithm arrived at a solution to foster more transparency.

Research funding should also cover non-technical solutions to address ethical issues, since roots of many of these issues are based on human values and context specific. As such, multi-disciplinary efforts that bring together various experts from different domains should be supported. For instance, research by criminologists and sociologists may provide important insights about challenges in the criminal justice system that can inform guidelines for when the use of facial recognition technology by law enforcement is appropriate.

Testing and evaluation are also critical. Due to the complexity of autonomous systems and the large number of potential system responses, traditional testing and evaluation methods are difficult to apply to AI technologies. The Institute for Defense Analysis categorized four ways to test AI systems: formal methods, cognitive instrumentation, adversarial testing, and run-time monitoring.[8] Research should include understanding the type of testing and evaluation an algorithm should be subject to, based on the level of risk the use case poses, the real-world context of its use, and the type of outcome the tool is expected to produce. Testing and evaluation should also cover the up-front governance processes that are used to

build the AI tool, as the decisions around the use of AI is often where bias occurs.

Further R&D can help better develop the technology and ethical standards in ways that improve outcomes for individuals and promote greater justice. A better understanding of techniques, policies, and governance structures can help society find ways to better promote the ethical use and design of AI systems.

> **Recommendation #1:** Funding should be allocated towards research and development for technical and non-technical solutions to address ethics issues in AI. Diversity among researchers should be taken into account when allocating these funds.

In addition to funding, dataset accessibility and quality is important. There are a limited number of publicly available datasets upon which researchers rely, and any issues with these datasets can lead to suboptimal results and have ripple effects. Further action should be taken to improve access to robust datasets, with proper safeguards, and to reduce the bias in existing datasets, since problems with these datasets can later be reflected in the AI systems that are trained on them. Additionally, in addressing bias, access to data about protected classes could be helpful to test, evaluate, and refine an AI system. Existing regulations that limit the collection of such data should be reevaluated in this context.

> **Recommendation #2:** Federal agencies should build off the Open Government Data Act to further develop and release publicly available benchmark datasets, with proper safeguards to protect privacy, mitigate bias, and promote inclusivity.
>
> **Recommendation #3:** Existing regulations that limit the collection of demographic information should be revisited to consider how organizations can access the information to mitigate algorithmic bias and promote fairness.

2. The federal government should encourage more diversity in AI talent to help mitigate unfair bias and promote fairness in AI practices.

A key tool for mitigating unfair bias and promoting fairness is a more diverse workforce.[a] A diverse workforce can better identify and address sources of bias in a dataset or algorithm, and the potential for an AI tool to inadvertently disadvantage a protected group.[9] For instance, facial recognition software used in an airport may regularly misidentify people wearing certain garments, such as a hijab, yarmulke, or turban, because the dataset was not trained with images of people wearing them. A diverse workforce may better identify the range of garments people wear across the world, so the dataset is more complete and fewer people are misidentified. Further, broader questions around the appropriate use cases for facial recognition would be better addressed by a diverse workforce that more accurately reflects the various views on fairness across society.

Unfortunately, the existing AI workforce suffers from poor representation of

a See BPC's paper *AI and the Workforce* for a more thorough discussion of workforce issues.

society, though there are active efforts underway to build additional diversity.[10] While an imbalance of representation among groups risks the perpetuation of historic inequalities,[11] the United States is uniquely positioned to leverage its diverse workforce toward improving innovation in the shadow of numerous studies demonstrating the accrued benefits to innovation through diversity and inclusivity.[12]

The challenge of creating a diverse workforce requires a holistic approach, starting from early education and throughout a person's career. It must focus not just on recruiting talent, but also on developing and retaining existing talent, which requires looking at the culture of an organization and whether it is inclusive. It also must include efforts to diversify organization leadership and not just those reporting to them. Different interventions should be tailored for optimal effect at different points in the pipeline. A significant amount of research has been done to understand the prevalence and consequences of discrimination, and approaches to mitigate human bias and help under-represented communities and marginalized groups realize their potential.[13] For instance, studies of interpersonal contact show increased exposure and social contact between different groups in schools are effective in changing discriminatory attitudes and behaviors.[14] There is evidence that the presence of minority and women leaders or role models can shift prejudices about the competence of minority groups,[15] and positively influence attitudes of minority groups about their own ability to succeed.[16] Policy solutions should be adapted to address the different drop off points for under-represented groups in the pipeline, from education through to careers.

> **Recommendation #4:** The federal government should expand funding to existing technology education programs and, where gaps exist, create new programs, particularly within under-represented communities and marginalized groups.
>
> **Recommendation #5:** Congress should authorize additional grants for programs designed to experiment on ways to increase workforce diversity and retain diverse talent at all levels of an organization.
>
> **Recommendation #6:** Federal agencies should review their current policies for recruiting and retaining talent from under-represented communities and marginalized groups at all levels of the organization to determine whether these policies need specific modifications for technology workers.

3. **The federal government should encourage the development of voluntary standards frameworks to help create shared conceptual foundations, terminology, and best practices for fairness and bias based on a cooperative and multi-stakeholder approach.**

Fairness and bias are terms that are not well-defined mathematically (and can never be solely defined mathematically). An AI system cannot simply be told to act more fairly and then be expected to act in a manner that humans find fair.

In addressing fairness and bias issues for AI, standards frameworks can help

better encode human values into AI systems. Standards frameworks can be used to help develop common language and terminology to guide discussions about how to incorporate evolving societal values into AI design. They can also highlight best practices and ways to think about the different societal impacts of AI systems. In setting standards, public engagement is necessary to ensure diverse perspectives are considered. Mindfulness about the diverse context in which an AI system is used and its various features is also important. Finally, it should be acknowledged that standards and frameworks have their limits and will not solve the various challenges with fairness and bias by themselves,[17] but they can become the basis for future government regulation.

The National Institute of Standards and Technology, an institute within the Department of Commerce, plays a key role in developing standards. NIST released a well-received, voluntary privacy framework in 2020 that was guided by input from public and private sector stakeholders.[18] NIST should continue undertaking similar efforts as part of the federal government's efforts to address these issues.

> **Recommendation #7:** Congress should authorize and provide robust funding to NIST to develop voluntary standards frameworks to help address the issues of bias and fairness based on a cooperative and multi-stakeholder approach and authorize grants through the National Science Foundation to qualified academic institutions to test qualified AI systems and develop AI testbeds.

4. In promoting ethics and mitigating unintended bias, the regulation of AI should build on existing regulation when possible and be tailored to different use cases using a risk-based approach.

The Civil Rights movement ushered in a wave of new laws to fight discrimination and promote fairness. These laws helped promote equal opportunity and countered discrimination in a variety of areas, including voting, housing, lending, and employment. Regulations to protect civil rights and liberties have been necessary for ensuring fairness and equal opportunity in a market-driven economy.

New incidences and stories about algorithmic bias and privacy violations will create a push to put in place new regulations and regulatory agencies. However, there is already a rich, existing body of applicable regulatory authorities across each sector of the economy. For instance, the Federal Trade Commission has been using and reviewing its use of the Fair Credit Reporting Act and the Equal Credit Opportunity Act (ECOA) for regulating AI systems.[19] Creating new agencies and laws could add unnecessary complexity and unintended consequences that might be avoided by building on existing regulatory frameworks. Further, effective regulatory approaches for AI will require deep, sector-specific knowledge to sufficiently evaluate solutions and understand associated risks. For instance, in the health care sector, the scientific community may be challenged in identifying causality while addressing gender bias in heart attack risk assessment, but they are better positioned to do so than people with no domain expertise. Therefore, regulation should be modernized when appropriate but be well-tailored and generally focus on building off existing frameworks and policies.

However, there are legitimate concerns about potential gaps and ambiguities in existing laws that hinder enforcement efforts. These concerns merit serious attention and greater congressional oversight to ensure enforcement is adequate and that regulators have the tools necessary to enforce the laws.

> **Recommendation #8:** Congress should conduct a study to review the range of existing federal regulations and laws that identifiably apply to AI and determine where existing laws apply and if gaps exist. Recommendations should also be provided that include specific changes that are needed to ensure laws apply to AI and are appropriately modernized.
>
> **Recommendation #9:** Congress should ensure federal agencies have adequate resources, including funding and staffing, to meet their regulatory obligations in the context of AI.
>
> **Recommendation #10:** At the start of each Congress, all committees should include in their oversight agendas an examination of AI and AI-related policies and issues in the executive branch within their jurisdiction if justified.

Regulatory approaches must consider the level of risk associated with different AI applications. For instance, an AI system used to place clothing ads does not carry the same risk of harm as one used to diagnose patients. As such, the nature of the risk will vary by sector and application. Higher risk areas, such as health care, lending, criminal justice, and housing, should be treated differently than lower risk ones.

This regulatory approach is consistent with the recent draft guidance for AI released by the Office of Management and Budget, which prescribes, "a risk-based approach… to determine which risks are acceptable and which risks present the possibility of unacceptable harm, or harm that has expected costs greater than expected benefits."[20] Pursuant to executive regulatory authorities in accordance with Executive Order 12866, the guidance continues, "…while the broader legal environment already applies to AI applications, the application of existing law to questions of responsibility and liability for decisions made by AI could be unclear in some instances, leading to the need for agencies, consistent with their authorities, to evaluate the benefits, costs, and distributional effects associated with any identified or expected method for accountability."[21] Importantly, the OMB draft memo does not address Congress' role in ensuring that the agencies are effectively and appropriately applying federal laws and regulations to AI uses.

The risk-based approach will require different regulatory schemes for different use cases and industries. Several regulatory approaches have been proposed, but the decision of whether and how to apply them should be context specific.

Given a risk factor, one approach to regulation is to focus on outcomes. The outcome-based approach gives industry more flexibility in determining ways to meet an objective, but it holds them accountable for the outcome of actions when they cause harm. This approach can be less burdensome than more prescriptive approaches and foster more innovation, but it can also be less proactive and

preventative in trying to mitigate harm in certain use cases. This can be concerning in high-sensitivity areas where life and liberty are at stake. However, an outcome-based approach should be combined with other approaches in higher sensitivity areas, such as health care, criminal justice, housing, and lending.

The first line of defense should continue to be existing laws. For instance, the ECOA requires people be given an explanation for why they were rejected for a loan, if they request it. Existing regulations, such as ECOA, can enhance an outcome-based approach to help more proactively prevent harm.

The discussions around whether, when, and how to complement the outcome-based approach (beyond applying and modernizing existing law) have been a source of major debate at BPC's convenings. To ensure a robust regulatory regime, several questions need to be further explored: For what use cases do the outcome-based approach and existing regulations fall short? How can regulation balance the need for transparency with the role trade secrets play in fostering innovation? What are the costs and benefits of the different regulatory remedies and do they justify their use in specific contexts? How should power asymmetries be managed? These are all questions that need to be answered.

Recommendation #11: Regulation related to AI should follow a risk-based approach. Federal agencies should report to Congress about how they are dealing with AI issues, including how they are using the full suite of enforcement tools available to them to address discriminatory outcomes.

Recommendation #12: Congress should conduct oversight by reviewing agencies' implementation and enforcement of federal regulations, including any OMB guidance and regularly hold oversight hearings to examine how agencies enforce their laws within their jurisdiction in AI cases.

Congress and federal agencies can further set up programs, such as "regulatory sandboxes," that allow select products to be tested and monitored for a limited period in a controlled environment. Innovation often requires experimentation. Regulatory sandboxes can be used to test a product designed to mitigate unintended bias or promote fairness in a small-scale environment and under the supervision of regulators.[22] Federal sandboxes do pose of risk of allowing products that might cause harm, so appropriate safeguards and oversight processes are vital.

Recommendation #13: Congress should support funding for agencies interested in adopting programs (such as regulatory sandboxes) for temporarily approving, testing, and monitoring innovative AI tools in limited markets. Programs should have necessary safeguards and oversight processes.

5. **AI and privacy should not be conflated, but AI-specific considerations should inform and influence privacy legislation.**

The debate surrounding comprehensive privacy legislation is much broader than AI, but AI-specific considerations and potential tradeoffs should inform this debate. AI-driven technologies, like facial recognition, bring novel privacy challenges that need to be addressed. Modern AI systems are also heavily dependent on vast amounts of quality data to be effective. Ensuring privacy is protected is important to build trust in AI systems, so good regulation and governance are important. If people trust an AI system will respect their privacy, they will be more willing to grant access to their data when appropriate. As a result, this trust can enable more robust, quality datasets that will improve the accuracy and trustworthiness of AI systems.

A national privacy framework should help build trust with the public. The U.S. federal government should take the lead in shaping domestic and global privacy standards. The failure to assert U.S. leadership on privacy issues may result in competing frameworks, such as the European Union's GDPR or China's state-centric model, becoming the de-facto global standard. This creates the risk that U.S. values will not be reflected in a global AI regulatory regime.

Privacy rules and regulations should be context-specific, with use cases evaluated based on the level of risk they pose. For example, a decision-making tool in the criminal justice system can have significant consequences for civil liberties, whereas a chatbot for an online retailer represents less consequential risk.

> **Recommendation #14:** The United States must enact federal legislation overseeing data privacy to build trust, prevent harm, and avoid ceding leadership on the issue to the EU or China for international standards.

IV. Conclusion

Promoting fairness, countering unfair and unintended bias, and protecting privacy are core American values. The rise of AI has created new challenges for ensuring these values. Policymakers should continuously strive to improve AI ethical standards and reduce harmful bias. Congress, an elected body that represents the various constituencies that makeup the American public, is well-positioned to debate and tackle the ethical challenges the rise of AI brings through developing a congressional AI national strategy. Congress, the executive branch, and the private sector should work together to build trust in AI by addressing legitimate concerns that consumers have about the use of AI in their daily lives.

Concerns about fairness, bias, and privacy should be addressed through a combination of research and development, workforce diversity, standard setting, and regulatory modernization. By developing a robust framework setting forth legal norms and policy surrounding AI ethics, the United States can help foster a global regime that upholds human rights, advances common prosperity, and promotes the pursuit of happiness as a guiding goal for the age of AI.

Endnotes

1 Executive Office of the President. *The National Artificial Intelligence Research and Development Strategic Plan*, October 2016. https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf

2 Executive Office of the President. *Executive Order 13859 Maintaining American Leadership in Artificial Intelligence*, Pub. L. No. 2019–02544, 84 FR 3967 E.O. 12859 3967 (2019). https://www.federalregister.gov/d/2019-02544.

3 Vought, Russell T., *Memorandum for the Heads of Executive Departments and Agencies: Guidance for Regulation of Artificial Intelligence Application*, January 7, 2020, https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf.

4 Hurd, Will, and Robin Kelly. "Rise of the Machines: Artificial Intelligence and Its Growing Impact on U.S." Subcommittee on Information Technology, Committee on Oversight and Government Reform: U.S. House of Representatives, September 2018. https://hurd.house.gov/sites/hurd.house.gov/files/AI%20White%20Paper%20Clean.pdf.

5 Moor, James. "The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years." *AI Magazine* 27, no. 4 (2006): 87. https://www.aaai.org/ojs/index.php/aimagazine/article/view/1911

6 Thornberry, Mac. John S. McCain National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. H.R.5515 (2018). https://www.congress.gov/bill/115th-congress/house-bill/5515/text.

7 Kearns, Michael, and Aaron Roth. *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. New York, NY: Oxford University Press, 2020, pp.9-11.

8 Haugh, Brian A., David A. Sparrow, and David M. Tate. "The Status of Test, Evaluation, Verification, and Validation (TEV&V) of Autonomous Systems." Alexandria, VA: The Institute for Defense Analyses, 2018. https://www.ida.org/-/media/feature/publications/t/th/the-status-of-test-evaluation-verification-and-validation-of-autonomous-systems/p-9292.ashx.

9 Shellenbarger, Sue. "A Crucial Step for Averting AI Disasters." *The Wall Street Journal*. February 13, 2019. https://www.wsj.com/articles/a-crucial-step-for-avoiding-ai-disasters-11550069865.

10 Xia, Rosanna. "Most Computer Science Majors in the U.S. Are Men. Not so at Harvey Mudd." *Los Angeles Times*. January 4, 2017. https://www.latimes.com/local/lanow/la-me-ln-harvey-mudd-tech-women-adv-snap-story.html.

11 West, Sara M., Meredith Whittaker, and Kate Crawford. "Discriminating Systems: Gender, Race and Power in AI." AI Now Institute, 2019. https://ainowinstitute.org/discriminatingsystems.pdf.

12 Mayer, Roger C., Richard S. Warr, and Jing Zhao. "Do Pro-Diversity Policies Improve Corporate Innovation?" *Financial Management* 47, no. 3 (January 2018).

13 Bertrand, Marianne, and Esther Duflo. "Field Experiments on Discrimination." *National Bureau of Economic Research, Inc* NBER Working Papers 22014 (January 2016). https://economics.mit.edu/files/11449.

14 Rao, Gautam. "Familiarity Does Not Breed Contempt: Diversity, Discrimination and Generosity in Delhi Schools" NBER Working Paper (May 2018). https://scholar.harvard.edu/rao/publications/familiarity-does-not-breed-contempt-diversity-discrimination-and-generosity-delhi.

15 Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. "Powerful Women: Does Exposure Reduce Bias?" *Quarterly Journal of Economics* 124, no. 4 (November 2009): 1497–1540.

16 Drury, Benjamin J., Oliver Siy, and Sapna Cheryan. "When Do Female Role Models Benefit Women? The Importance of Differentiating Recruitment From Retention in STEM." *Psychological Inquiry* 22 (2011): 265–269, and Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India." *Science (New York City, NY)* 335, no. 6068 (February 3, 2012): 582–86. https://doi.org/10.1126/science.1212382.

17 Whittaker, Meredith. Facial Recognition Technology (Part III): Ensuring Commercial Transparency & Accuracy, Written Testimony of Meredith Whittaker, United States House of Representatives Committee on Oversight and Reform (2020). https://ainowinstitute.org/oversight-committee-testimony-whittaker.pdf.

18 "NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management." U.S. Department of Commerce: National Institute of Standards and Technology, January 16, 2020. https://www.nist.gov/system/files/documents/2020/01/16/NIST%20Privacy%20Framework_V1.0.pdf.

19 Smith, Andrew. "Using Artificial Intelligence and Algorithms." Federal Trade Commission. *Business Center* (blog), April 8, 2020. https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms.

20 Vought, Russell T., *Memorandum for the Heads of Executive Departments and Agencies: Guidance for Regulation of Artificial Intelligence Application*, January 7, 2020, https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf.

21 Executive Office of the President. Executive Order 12866: Regulatory Planning and Review. 30 September 1993. *Federal Register*, vol 58 no. 51735, 4 October 1993 https://www.archives.gov/files/federal-register/executive-orders/pdf/12866.pdf.

22 Turner Lee, Nicol, Paul Resnick, and Genie Barton. "Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms." Washington DC: Brookings Institution, May 22, 2019. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/.