Unfairness In Machine Learning: Reaching an Ethical Closure

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

Henry Todd

Spring 2023

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

Bryn E. Seabrook, Department of Engineering and Society

Most people today wield state-of-the-art artificial intelligence in their pockets, in the form of smart assistants such as Siri. The growth of AI in society is not stopping there: the number of cars on the road that drive themselves continues to go up, and increasingly powerful tools like Chat-GPT are being developed each year. However, since the foundation of artificial intelligence with the "Perceptron" in 1958, unlike other important realms of science and technology artificial intelligence has seen a distinct lack of legislation governing its development and implementation. As the foothold of artificial intelligence in everyday life grows, ensuring its algorithms are developed and used ethically is paramount.

Despite the dearth of policy specific to artificial intelligence, artificial intelligence is intrinsically tied to the data sets on which it is developed, and as a result falls under numerous data protection and privacy laws. However, the ways in which artificial intelligence draws conclusions using that data is much more complex than in typical statistical calculations. As a result, it is unclear whether preexisting legislation is sufficient to ensure ethical usage of artificial intelligence, or if more nuanced policy is required. This paper analyzes the social construction of artificial intelligence by considering the social groups from which AI has developed and which influenced its implementations and considers if current legislation governing data protection is sufficient to ensure ethical development and usage of AI, or whether more nuanced legislation is required to find an ethical closure in the social construction of artificial intelligence.

RESEARCH METHODS

This paper takes sources gathered from journals, articles, and research papers, found primarily through online databases. These articles consist of information regarding the current legislation governing artificial intelligence, proposed future legislation, and perspectives on the sufficiency of these policies in ensuring ethical AI. Research for the paper can be separated into categories. The first category is the background and history of artificial intelligence, both to provide background information to the reader as well as to conduct a Social Construction of Technology (SCOT) analysis of the technology as it developed. Second, research was conducted on ethical considerations in the realm of artificial intelligence. The third area of research involves the past and current policy on data and AI, as well as perspectives on the potential for future regulation. Finally, research was conducted on bias in artificial intelligence, both its sources and how it impacts the output of algorithms, as well as potential means of minimizing bias. Additionally, research was conducted on the history and use of the social construction of technology as a framework for analysis of a technology.

The data collected by the research detailed above is outlined in the following sections with discussions on findings, and further analyzing using the social construction framework by breaking down the social groups relevant to artificial intelligence and considering how their interpretations of the technology vary. A discussion of the closure of artificial intelligence as a technological artifact follows, considering the cycle of interpretive flexibility at different stages. The paper uses the analysis of interpretive flexibility provided by SCOT to consider the predominant form of artificial intelligence today, and its perceptions, to analyze both current regulations, and how future regulations might be imposed.

ORIGINS OF AI AND CURRENT DATA PROTECTION RIGHTS

The preliminary concept from which field of artificial intelligence was born is the "Perceptron," first published in 1958 in a paper by Frank Rosenblatt. As a psychologist, Rosenblatt sought to represent the human brain electronically. His Perceptron is an abstraction of a human neuron, which simulates brain activity by performing calculations on an input to provide output. The concept is deceptively simple, when many Perceptrons are used in tandem they have the ability to "learn." Even at the time, Roseblatt's conception was able to recognize and relay information about shapes it perceived through a lens. The Perceptron is the foundation of modern machine learning, a subdomain of artificial intelligence in which large systems of perceptrons are used in conjunction to "learn" from training data in order to provide output. They do so by combining nodes (each like a single Perceptron) into a network. Although modern methods learn in the same manner, they are distinctly different from early Perceptron networks in two ways: 1) the number of nodes combined is much greater, forming vastly deeper and wider layers; and 2) these more complex systems of nodes require immense quantities of data ("big data") to sufficiently train the model.

In order to consider the effectiveness of current data privacy legislation ensuring ethical artificial intelligence, it is important to understand the wide variety AI algorithms and their differences from other uses of data sets. One such example is the use of artificial intelligence in children's toys to emulate and respond to emotion, which authors Andrew McStay and Gilad Rosner discuss in an article titled *Emotional AI in children's toys*. The two define emotional AI to be "technologies that use affective computing and artificial intelligence techniques to try to sense, learn about, and interact with human emotional life," and argue that unfairness and injustice relating to emotional data of children is a new issue that must be considered (McStay, A., Rosner, G., 2021, p. 2). Specifically, the use of emotional AI on children can lead to personal detriment in the form of "personalities that [are] heavily influenced by algorithms," as well as the manipulation of emotions for economic gain of companies (McStay, A., Rosner, G., 2021, pp. 4-5).

These effects are distinctly different from typical statistical analysis in that artificial intelligence has the power to interact with and use new data over an extended period of time. In the case of children's toys, emotional AI algorithms were trained on a preexisting dataset, but continue to interact with children after the algorithms have been trained. The continued interaction between users and AI leads to more complex ethical considerations than previous uses of data. Despite these complicating factors, McStay and Rosner state after conducting interviews with field experts in data privacy that new regulation is not necessary as emotion data is already protected under legislation regarding sensitive data and/or biometric data (McStay, A., Rosner, G., 2021, p. 7).

The current data protection rights in the US are governed by the Data Protection Act of 2021, which established the Data Protection Agency to govern data protection in place of the Federal Trade Commission. The subjects of oversight include (1) "automated decision systems…, (2) [large scale profiling], and (3) personally identifying biometric information" (Congressional Research Service, 2021). Since the introduction of this act two years ago, the use of machine learning is loosely protected under the first point, however no specific legislation regarding the technology is in place. Additionally, new legislation has been proposed as the American Data Privacy and Protection Act (ADPPA) in late 2022. The act builds on previous legislation by adding protections against “potentially discriminatory impacts of algorithms” (Patel et al., 2021).

THE SOCIAL CONSTRUCTION OF TECHNOLOGY

The social construction of technology is a framework in STS first pioneered by Bijker et al. and expanded upon by Bijker and Pinch in two papers published in 1984. The framework discusses the social groups which influence the construction of technology by providing interpretations that drive development. As technology is developed, there is “interpretive flexibility” between different social groups, leading to divergent development until eventually a predominant form prevails.

The construction of artificial intelligence technology is interesting due to its arrival out of a number of fields. As mentioned earlier, the initial conception of machine learning came from a psychologist, Frank Rosenblatt, who sought to emulate the human brain electronically. From there, artificial intelligence developed greatly alongside the rise of computing while also taking inspiration from statistics. Different applications of artificial intelligence grew from various backgrounds, for example the Bellman-Ford equation used in Q-learning is also used in general shortest-path algorithms in computer science, which have been used in practical applications such as packet routing algorithms for computer networks.

Previous studies of the social construction of artificial intelligence describe the “three central stakeholders,” or “relevant social groups” in Bijker’s terms, as “academia, industry, and government” (Enyon, R., Young, E. 2020). This paper investigates the social construction of artificial intelligence as it pertains to AI’s use in education. Its analysis therefore considers the interpretive flexibility of artificial intelligence between these groups in that context, considering how these groups view artificial intelligence as a means of aiding education. The paper performs a similar analysis, but rather in the context of how artificial intelligence influences at risk social groups, e.g. children and minorities, due to biases arising from data and through monitoring and interacting with users. The interpretation of artificial intelligence from various social groups leads to insight into the current policy, for example if some groups view AI as similar to other data analysis and therefore current legislation is sufficient, and to the need for new legislation from these different points of view.

RESULTS AND DISCUSSION

The analysis section of this paper addresses the ways government regulations should minimize bias in artificial intelligence applications to reduce disproportionate classification of minorities in its applications. First, an analysis of the social construction of artificial intelligence is performed. The overview of SCOT analysis introduces the current day interpretive flexibility of AI. The following sections overview ethics and bias in artificial intelligence, with regards to both data used to train artificial intelligence algorithms, as well as the bias propagated by the algorithms themselves. The potential impact of such biases is discussed with examples. Following the discussion of potential pitfalls of artificial intelligence design, the paper presents the current regulations in place governing data protection and artificial intelligence and considers potential future governance. SCOT is used throughout to analyze the place of AI in society, and how the view of AI might reach closure to aid in the reduction of bias.

SOCIAL CONSTRUCTION OF ARTIFICIAL INTELLIGENCE

This section provides an analysis of the social construction of artificial intelligence to gain insight into the social context from which artificial intelligence has developed. With that analysis in mind, we consider how artificial intelligence needs to develop now to be ethical, and how it might be regulated. This analysis aids the following section which considers the ethics of biased AI and how it might be regulated.

First Era of Artificial Intelligence

Artificial intelligence has seen tremendous growth from its beginnings as an esoteric concept from academia to a technology used in almost every industry today, including entertainment, advertising, finance, technology. As a result, the relevant social groups for AI as an artifact have changed over time as well. Artificial intelligence in its early stages was conceived in academia from several fields. Warren McCulloch, a neurophysiologist, and Walter Pitts, a mathematician, developed early models of neural networks in 1943 which involved networks of binary devices (on or off) activated at some threshold (Macukow, 2016). Then, Rosenblatt devised the “Perceptron” in 1958, an electronic device modeled after neurons in the human brain that “showed an ability to learn,” from which modern machine learning is descendant (Macukow, 2016).

From this early stage of AI, there are several relevant social groups: (1) the developers of theory, usually researchers from fields such as mathematics, psychology, and neuroscience, (2) the public, who awaited the rise of sentient robots, and (3) the investors who saw potential for practical applications and funded research. The interpretive flexibility of artificial intelligence is clear. Researchers formed two interpretations, one set worked on modeling the human brain to attempt some form of intelligence, while others such as Widrow and Hoff developed more mathematical models focusing on simply the characteristic of a "learning procedure" by which regressions become more accurate (Macukow, 2016). The public at the time associated artificial intelligence with robots capable of sentient thought, more akin to science fiction, while investors interpreted artificial intelligence as a means of either financial gain or military dominance (Fast & Horvitz, 2017).

The interpretive flexibility of artificial intelligence at began to reach closure in 1969 with the book published by Minsky and Papert, in which "some limitations for complexity of the problem that can be solved by perceptrons were established" (Fradkov, 2020). This book helped shift the dominant view of artificial intelligence to nothing more than theoretical concepts without much practical application by demonstrating some limitations of the early technology, resulting in both public disappointment in a lack of sophistication and dwindling investor support. This era is described as "the winter of artificial intelligence," due to the "reduction of funding of AI research… for more than two decades" (Fradkov, 2020).

Following the Winter of AI

The interpretive flexibility of artificial intelligence reopened as a result of major technological advancements, namely "the invention of backpropagation learning algorithm," Google's MapReduce which "made it possible to distribute the processing of huge amounts of data between simple processors," and the advent of 'big data' as "the cost of RAM significantly decreased, which opened the possibility to work with large amounts of data in memory"

(Fradkov, 2020). The increased computation power led to a new sect of artificial intelligence, namely deep machine learning, which utilizes big data to train large, multi-layered algorithms.

With the advent of (deep) machine learning models, artificial intelligence began to have real practical applications. As a result, the relevant groups in this second stage include technology companies with access to enormous amounts of data in their databases, algorithm developers, and again researchers, the public (users), and investors. With artificial intelligence moving from its theoretical beginnings to practical implementations which make use of the drastically increased amount of data available in today's world, large technology companies have come to lead the way in terms of further development. As a result, the interpretation of artificial intelligence as a technology has changed for many of the relevant social groups. Researchers now view machine learning as a tool to be used and focus on developing its use in various applications rather than developing the tool itself. Tech companies who collect and own the data see artificial intelligence to make profit by selling access to their databases, as well as developing their own implementations as a product. Public perceptions have shifted towards artificial intelligence, which is now viewed as both a tool for their own use, but also as a means for companies to exert influence over users of the technology. This shift in perception has resulted in a growing desire for regulation as the influence of AI technology over society grows.

Modern Interpretive Flexibility

The technology of artificial intelligence as an artifact today has not yet reached a second closure. Its place in society is varied, as are its interpretations by different social groups. The need for increased regulations governing ethical artificial intelligence is clear with regards to the above SCOT analysis. At present, there is a disparity between the views of artificial intelligence of various relevant social groups. Specifically, corporations and other developers of artificial intelligence often use the technology as a means for financial gain, or as a product. As a result, these companies have incentive to make the most revenue possible at the potential risk of users. Users on the other hand are concerned with the potential misuse of AI implementations by the companies, with the public becoming increasingly anxious about the influence of AI over society.

When considering the discussion of bias and unfairness which can be amplified by artificial intelligence, it is important to recognize the interpretations of artificial intelligence from the perspective of the companies and developers versus that of the users. Without regulation, or other strong incentive to do so, companies are not likely to make necessary changes to their algorithms at the cost of profits. The view of many companies that their algorithms are unbiased and indifferent must shift to the viewpoint that AI can perpetuate or even create social harm in order to reach a second closure on interpretive flexibility. The following sections explore the various interpretations of AI today and the issues of bias in more detail, furthering the argument for regulation in order to reach a desired closure on ethical AI.

BIASED DATA AND ETHICAL AI

Ethics in artificial intelligence presents itself in many forms. To consider how regulations might address issues of bias and unfairness in certain applications we must understand the ethical considerations that go into developing smart algorithms. Unlike other technologies, the ethical considerations for artificial intelligence go beyond the development of the application alone. Before the algorithm is even designed, there exists the underlying data sets upon which the AI will be trained. This dataset poses ethical problems itself since “machine learning datasets need to be large…” for algorithms to function properly, and “often-used clinical trial research databases are largely derived from majority populations,” according to Safdar et al., clearly the data itself may contain biases inherent within (2020).

Bias From the Dataset to the Model

These biases do not rest within the dataset, however, as author Cath describes, “AI systems… apply learning techniques from statistics to find patterns in large sets of data and make predictions based on those patterns,” such that when bias resides in the data, artificial intelligence algorithms will perpetuate and even exaggerate those biases in its output (Cath, C. 2018). The propagation of bias in artificial intelligence algorithms from their underlying data makes defining an ethical implementation more convoluted. Following the SCOT discussion above, there are some interpretations of AI which argue that artificial intelligence algorithms are separate from the data and as a result should be judged separately. However, choosing to evaluate these algorithms without considering how they handle bias ignores the issue that the bias will propagate to the output of the AI and will lead to continued misclassification of minorities as a result (Buolamwini & Gebru, 2018). Therefore, when considering potential policy governing AI algorithms, we must treat the underlying biases as part of the algorithm itself in order to reach our desired closure.

The Dangers Inherent in AI

Yet ethical dilemmas do not arise solely from the data inputted into artificial intelligence, but from within the algorithm as well. To understand the potential ethical dilemmas in AI decisions let us consider the fictitious example given by professor and researcher of law, Pauline Kim, in which an artificial intelligence algorithm is used to rank candidates for a job. In this scenario, the number of women hired by the company are disproportionately less than the number of men. Yet because the algorithm is “data-driven” and “gender is not [an explicit] factor in the… hiring algorithms, [the company] assumes the process is fair” (Kim 2019).

Although fictitious, the example illustrates another important area of varying interpretations of artificial intelligence, namely the belief that the algorithm itself is biased or unbiased. The idea that AI cannot be source of bias is, however, a misconception. Michael Sandel, a political philosopher and professor of government theory at Harvard Law, describes the danger of this interpretive flexibility, that “AI not only replicates human biases, it confers on these biases a kind of scientific credibility” and gives them “an objective status” (Pazzanese, 2020). Such a belief would tempt a developer to believe that if the data set is unbiased then all previously mentioned issues would be resolved, since the algorithm could never be a source of bias. This belief that AI cannot be the cause of bias, of course, is incorrect and misleading.

Consider that an algorithm which does not use gender as a factor in its decision yet may implicitly deem other gendered characteristics as preferable. Indeed, there are real cases of gender disparity at the hands of artificial intelligence. Buolamwini and Gebru describe an example of gender and bias in a commonly used machine learning algorithm, Word2Vec (2018). The algorithm was used to build an analogy generator that likened men to the term “programming” and women to “housekeeping.” Going beyond the dataset, the algorithm itself must be designed properly to account for potential biases to prevent the propagation of unfairness.

These biases are clearly harmful, and the algorithms which fail to prevent bias can have serious consequences. Take, for example, a software designed by Wu and Zhang (2016) which purports to identify characteristics such as IQ and likelihood of committing a crime solely from an individual's face. A misclassification by this algorithm would lead to valid candidates for a job being ignored, or an individual falsely targeted for crime. Buolamwini and Gebru discuss how "clients of such software include governments" and other "face detection and classification algorithms are also used by US-based law enforcement for surveillance and crime prevention" (2018). The stakes are no longer simply an offensive analogy generator as artificial intelligence is already being used to assess individuals' place in society. Failure to prevent or reduce bias will lead to perpetuating and even accentuating preexisting social inequities, especially for minorities. In order to prevent the amplification of injustice, we must lead artificial intelligence to reach a second closure on the interpretations described above.

REGULATIONS PRESENT AND FUTURE

Given these significant issues that arise both within the algorithm implementation as well as from the data it uses, it is crucial to understand the current regulations on data protection and artificial intelligence to understand what is already protected, and to reveal where regulation might be improved. Unfortunately, "while the European Union already has rigorous data-privacy laws and the European Commission is considering a formal regulatory framework for ethical use of AI, the US government has historically been late when it comes to tech regulation" (Pazzanese, 2020). Despite lackluster legislation governing artificial intelligence specifically in the US, there are laws governing data protection and privacy, specifically as part of the Data

Protection Act of 2021 (Library of Congress, 2021). This bill established the Data Protection Agency, which "among other functions… must oversee the use of high-risk data practices" which includes "automated decision systems, such as machine learning" as well as "profiling individuals on a large scale" (Library of Congress, 2021). Clearly, although the bill serves to govern many types of data uses, it is attempting to oversee the use of artificial intelligence as well.

Additionally, on July 20, 2022, the American Data Privacy and Protection Act (ADPPA) was approved, which lays out "national standards and safeguards for personal information collected by companies, including protections intended to address potentially discriminatory impacts of algorithms" (Patel et al., 2022). Although approved, this bill has not yet been enacted and has been placed on the Union calendar for the time being. Yet, the consideration of the bill is following "a growing trend calling for federal regulation of AI" (Patel et al., 2022). Specifically, the bill protects against the "collect[ion], process[ing], or transfer [of] covered data in a manner that discriminates in or otherwise makes unavailable the equal enjoyment of goods or services" (Patel et al., 2022). The wording of the bill is important as it clearly states that both collecting and processing/transmitting biased data will be protected against, helping to prevent both core issues discussed previously: (1) the creation of biased data sets and (2) the propagation of bias through algorithms.

The Need for Better Regulation

Despite the growing trend calling for regulation of artificial intelligence, the bill(s) passed and currently under review are both vague and focus primarily on data itself, albeit with some consideration of propagation of bias as well. These bills lack, however, a clear discussion on how artificial intelligence algorithms can be held accountable for unethical use. Given the state of current legislation, further regulation will be required to ensure AI can achieve ethical closure.

In addition to protecting data privacy and minimizing bias, potential regulations could demand algorithmic transparency, explainability, and ability to audit (Cath, 2018). Specifically, transparency speaks on the clarity of intentions for an algorithm; explainability is a "possible mechanism to increase algorithmic fairness, transparency, and accountability" by giving individuals the right to an explanation on their classification, consider the example given by Cath where an applicant would be entitled to the 'reasoning' behind their algorithmic refusal of a loan application; and auditing speaks on "accountability mechanisms" for "highly complex algorithmic systems" whose methods of classification cannot properly be described, and instead the inputs and outputs of the algorithm alone are examined for potential bias and unfairness (Cath, 2018).

Authors Bostrom and Yudkowsky reiterate the importance of "develop[ing] algorithms that are not just powerful and scalable, but also transparent to inspection" (2018). An important quality for transparency of AI, they state, is to be "predicable to those they govern," arguing that optimization is not always as valuable as being able to anticipate how the algorithm will act (Bostrom & Yudkowsky, 2018). In addition to using these concepts to hold algorithms accountable, Yu et al. describes how "well-established technique of reward shaping" in machine learning could be used to "to incorporate ethical values" into the algorithm itself (Yu et al., 2018). Going forward, policy should look to enforce some of these constraints to ensure ethical use of AI, although more work is needed to develop the technology, even mandate that algorithms learn to be ethical on their own.

LIMITATIONS

This paper defines and identifies the ethical considerations in the field of artificial intelligence; however it is unable to consider all possible ethical dilemmas in such a widespread technological field. Additionally, the overview of the current and future regulations regarding data protection and AI given in this paper pertain exclusively to the United States and may not represent the position of other countries. The analysis performed on the social construction of artificial intelligence hopes to account for the various social groups and interpretations, however artificial intelligence is a broad umbrella term for many forms of algorithms and innumerable implementations, so works with a vague definition of artificial intelligence means and is unable to consider all perspectives. Further research should seek to perform further analysis of the social construction of artificial intelligence in its current form and consider ways in which the technology might reach a preferable closure. Additionally, new algorithms and implementations are being developed. The scope of this paper is limited to AI historically and cannot assess what the impact of algorithms such as Chat-GPT and others not yet released.

CONCLUSION

New regulations are necessary to ensure ethical use of artificial intelligence and minimize biases and unfairness in its applications. Currently artificial intelligence is viewed through with interpretive flexibility, of which many interpretations can be harmful. The implications of allowing these interpretations to pervade are severe as the power and prevalence of artificial algorithms continues to grow. Current regulations governing data protection and AI are in place, but insufficient. Areas in which artificial intelligence governance need improvement include enforcing transparency, explainability, or, in cases of higher complexity, auditing. These methods are necessary, as indicated by an analysis of the social construction of artificial intelligence, in order to reach an ideal closure on artificial intelligence as an artifact, in which its potential pitfalls are recognized and accounted for. Without incentive to do so, companies will not reliably reduce the misclassification of minorities due to biased data and methods. Regulations can be passed, as specified before, to mandate such reforms. Research on the place of AI in society is significant towards minimizing misclassification by indicating the need for regulations by analyzing the problems at hand and the lagging policies in the US surrounding artificial intelligence.

REFERENCES

Blackman, R. (2020). A practical guide to building ethical AI. *Harvard Business Review*. https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai

Bostrom, N., Yudkowsky, E. (2018). The ethics of artificial intelligence. *Artificial Intelligence Safety and Security* (pp. 57-69). Chapman & Hall. https://doi.org/10.1201/9781351251389-4

Buolamwini, J., Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research, 81*(1), 77-91. http://proceedings.mlr.press/v81/buolamwini18a.html?mod=article_inline

Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions Royal Society A.* http://dx.doi.org/10.1098/rsta.2018.0080

Congressional Research Service. (2021, June 17). *Data Protection Act 2021 (S.2134)*, Library of Congress. https://www.congress.gov/bill/117th-congress/senate-bill/2134

Eynon, R., Young, E. (2021). Methodology, Legend, and Rhetoric: The Constructions of AI by Academia, Industry, and Policy Groups for Lifelong Learning. *Science, Technology, & Human Values*, *46*(1), 166–191. https://doi.org/10.1177/0162243920906475

Fast, E., Hovitz, E. (2017). Long term trends in the public perception of artificial intelligence. *AAAI Press*, *17*(1), 963-969. https://dl.acm.org/doi/proceedings/10.5555/3298483

Fradkov, A. L. (2020). Early history of machine learning. *IFAC, 53*(2), 2405-8963. https://doi.org/10.1016/j.ifacol.2020.12.1888.

Kerry, C. F. (2020). Protecting privacy in an AI-driven world. *Brookings.*

https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/

Kim, P. T. (2019). Big data and artificial intelligence: New challenges for workplace equality. *University of Louisville Law Review, 57*(2), 313-328. https://heinonline.org/HOL/P?h=hein.journals/branlaj57&i=325

Macukow, B. (2016). Neural networks: State of Art, Brief History, Basic Models and Architecture. *Computer Information Systems and Industrial Management, 9842*(1). https://doi.org/10.1007/978-3-319-45378-1_1

McStay, A., Rosner, G. (2021). Emotional artificial intelligence in children's toys and devices: Ethics, governance, and practical remedies. *Big Data and Society*, *8*(1). https://doi.org/10.1177/2053951721994877

Patel, N. K.., Shinohara, T. K., Rosa, J. M., Harrington, B. J., Kourinian, A., & Waltzman, H. W. (2022). The american data privacy and protection act: Is federal regulation of AI finally on the horizon? *Mayer Brown*. https://www.mayerbrown.com/en/perspectives-events/publications/2022/10/the-american-data-privacy-and-protection-act-is-federal-regulation-of-ai-finally-on-the-horizon

Pazzanese, C. (2020). Great promise but potential for peril. *The Harvard Gazette*. https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/

Pinch, T. J., Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Sciences of Science, 14*(3), 399-441. https://www.jstor.org/stable/285355

Raub, M. (2018). Bots, bias, and big data: Artificial intelligence, algorithmic bias and disparate impact liability in hiring practices. *Arkansas Law Review, 71*(2), 529-570.

https://heinonline.org/HOL/P?h=hein.journals/arklr71&i=549

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386-408. https://psycnet.apa.org/doi/10.1037/h0042519

Safdar, N. M., Banja, J. D., Meltzer, C. C. (2020). Ethical considerations in artificial intelligence. *European Journal of Radiology, 122*(1). https://doi.org/10.1016/j.ejrad.2019.108768

Sato, T. (2021). *Technological Frame and Best Praxis in the Age of Artificial Intelligence*. Management, Education, and Automation. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003017707-5/technological-frame-best-praxis-age-artificial-intelligence-toyoko-sato

Yu, H., Shen, Z., Miao, C., Leung, C., Lesser, V. R., Yang, Q. (2018). Building ethics into artificial intelligence. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. https://doi.org/10.48550/arXiv.1812.02953

Wu, X., Zhang, X. (2016). Automated Inference on Criminality using Face Images. *ArXiv*, arxiv.org/abs/1611.04135.