Fairness in Machine Learning: Regulation or standards?

Mike H.M. Teodorescu and Christos Makridis

Brookings (Feb. 2024)
© 2024 Brookings. All rights reserved.
Reprinted with permission.

BROOKINGS

RESEARCH

Fairness in machine learning: Regulation or standards?

and

February 15, 2024

Abstract

Machine Learning (ML) tools have become ubiquitous with the advent of Application Programming Interfaces (APIs) that make running formerly complex implementations easy with the click of a button in Business Intelligence (BI) software or one-line implementations in popular programming languages such as Python or R. However, machine learning can cause substantial socioeconomic harm through failures in fairness. We pose the question of whether fairness in machine learning should be regulated by the government, as in the case of the European Union's (EU) initiative to legislate liability for harmful Artificial Intelligence (AI) and the New York City AI Bias law ↗, or if an industry standard should arise, similar to the International Standards Organization (ISO) quality-management manufacturing standard ISO 9001 ↗ or the joint effort between ISO and the International Electrotechnical Commission (IEC) standard ISO/IEC 27032 ↗ standard for cybersecurity in organizations, or both. We suggest that regulators can help with establishing a baseline of mandatory security requirements, and standards-setting bodies in industry can help with promoting best practices and the latest developments in regulation and within the field.

Introduction

The ease of incorporating new machine learning (ML) tools into products has resulted in their use in a wide variety of applications, including medical diagnostics, benefit fraud detection, and hiring. Common metrics in optimizing algorithm performance,

such as algorithm Accuracy (the ratio of correctly predicted to total number of attempted predicted), do not paint the complete picture regarding False Positives (algorithm incorrectly predicted positive) and False Negatives (algorithm incorrectly predicted negative), nor do they quantify the individual impact of being mislabeled. The literature has, in recent years, created the subfield of Machine Learning Fairness, attempting to define statistical criteria for group fairness such as Demographic Parity or Equalized Opportunity, which are explained in section II, and over twenty others, described in comprehensive review articles like the one by Mehrabi et al (2021).[1] As the field of ML fairness continues to evolve, there is currently no one standard agreed upon in the literature for how to determine whether an algorithm is fair ↗, especially when multiple protected attributes are considered.[2] The literature on which we draw includes computer science literature, standards and governance, and business ethics.

Fairness criteria are statistical in nature and simple to run for single protected attributes—individual characteristics that cannot be the basis of algorithm decisions (e.g., race, national origin, and age, among other individual characteristics). Protected attributes in the United States are defined in U.S. federal law and began with Title VI of the Civil Rights Act of 1964 ↗. However, in cases of multiple protected attributes it is possible that no criterion is satisfied. Furthermore, oftentimes a human decision maker needs to audit the system for compliance with the fairness criteria with which it originally complied at design,[3] given that a machine learning-based system often adapts through a growing training set as it interacts with more users. Moreover, no current federal law nor industry standard mandates regular auditing of such systems.

However, precedents exist in other industries for both laws and standards where risk to users exists. For example, in both the U.S. (the Federal Information Security Modernization Act ↗ [FISMA], the California Consumer Privacy Act ↗ [CCPA]) and EU (the General Data Protection Regulation ↗ [GDPR]) laws protect user data. In addition, the industry has moved to reward those who find cybersecurity bugs and report them to companies confidentially, for example through bug reward programs like Google's Vulnerability Reward Program ↗. In this article, we propose that a joint effort between lawmakers and industry may be the best strategy to improve the fairness of machine learning systems and maintain existing systems so that they adhere to fairness standards, with higher penalties for systems that pose greater risks to users.

Before elaborating further on existing regulations, we will briefly summarize what ML fairness is and illustrate why it is a complex problem. (#_ftnref1)

Defining ML fairness

ML fairness—and AI fairness more broadly—is a complex and multidimensional concept, and there are several definitions and metrics used to measure and evaluate fairness in ML-based systems. Some of the most common definitions include:[4][5][6]

1. **Demographic parity (statistical parity):** Demographic parity requires that the algorithm predicted outcomes are independent of a specified protected attribute. In other words, the ML system should treat different demographic groups equally, resulting in similar outcomes for each group.

2. **Equalized odds[7] (conditional procedure accuracy equality):** Equalized odds require that the AI system's true positive and false positive rates are equal given different demographic groups when conditioned on the true label. This means that the system should perform equally well for each group in terms of correctly predicting positive and negative outcomes.

3. **Equality of opportunity:** Equality of opportunity is a simplified version of equalized odds as it only equalizes true positive rates. This metric specifies that the system's true positive rate is equal across different demographic groups when conditioned on the true label. This means that the ML system should provide equal chances for each group to receive positive outcomes when it deserves them.

4. **Predictive parity[8] (outcome test fairness):** Predictive parity requires that the system's positive predictive value (defined as the proportion of the true positives among all the predicted positive outcomes) is equal across different demographic groups. This means that the system should be equally accurate in predicting positive outcomes for each group.

5. **Calibration (group calibration—see Footnote 8):** Calibration requires that the system's predicted probabilities of a positive outcome are accurate for each demographic group. In other words, if the system predicts a 70% chance of a positive outcome for a specific group, then 70% of the cases in that group should indeed result in a positive outcome.

The fairness criteria should equally apply to procedural and distributive aspects.[9] Other approaches are also possible; among others, the use of balanced accuracy (and its related measures)[10] should be explored.

These definitions of fairness often require trade-offs, as optimizing for one may negatively impact another.[11] Chouldechova (2017)[12] showed that it is not possible for three group-fairness criteria to be satisfied at once, so determining the appropriate fairness metric for a specific AI system depends on the context, the domain, and the societal values at stake. This is a decision that human designers of an ML system need to make, ideally prior to the system's release to actual users. Involving several stakeholders, including users of the ML system, in deciding the fairness metric is helpful to ensure the end-product aligns with the system's ethical principles and goals.[13] There is an extensive literature in ethics and psychology revolving around principles of procedural fairness.[14][15][16] Throughout the literature, procedural fairness has been broadly understood as perceived "fairness of the methods used to make the decisions,"[17] and involves high-level principles such as correctability (if a decision is perceived as incorrect, the affected party has a mechanism to challenge it), representativeness, and accuracy (in the field of ML algorithms, this means the algorithms rely on "valid, high-quality information"[18]), among others (for a detailed explanation of procedural fairness applied to ML see Footnote 17). These principles are found in regulations; for example, the correctability principle is found in GDPR privacy law as a "right to rectification" and in the "right to object" found in both CCPA and GDPR.[19]

Given the rapid advances in machine learning, we recognize that legal frameworks by nations or local governments may be more difficult to develop and update. Thus, we propose first looking at industry standards, which can be used to incentivize companies to perform better while also collaborating on developing standards. Certification of a product to a standard can provide customers with a signal of quality and thus differentiate among ML-based solutions in the market.

Companies who are early adopters of a novel standard of ML fairness may be able to use that standard to gain market share as well as establish a competitive advantage compared to newcomers. For example, a company that invests early on in an auditing team for its ML system may produce more transparent software, which could be apparent to the discerning customer and thus appease concerns customers might

have regarding the use of their data. Industry collaboration organizations such as Institute of Electrical and Electronics Engineers ↗ [IEEE] have developed standards for recent technologies with the help of leaders in the industry that have resulted in benefits for customers. For instance, the IEEE 802 ↗ wireless standards provided a foundation for the now-widespread Wi-Fi technology and, in the early days, provided a badge to signal to customers purchasing computers that the manufacturer complied with the latest Wi-Fi standard. Current updates to the standard enable "a new wave of innovation ↗" including in areas of indoor mapping. The same incentives for standardization advocated by organizations like IEEE and ISO in manufacturing quality management, electrical safety, and communications may apply to ML.[20]

In the following section, we include some additional background information on standards and parallels to the cybersecurity standards that could serve as a reference for developing standards for ML fairness. (#_ftnref1)

Background: Standards and regulations

Standards are guidelines or best practices developed by industry and federal regulators in collaboration to improve the quality of products and services. Although they are often voluntary, organizations choose to adopt them to signal their commitment to security and quality, among other reasons. Some of the most prolific and widely adopted standards in the technology sector are the International Organization for Standardization (ISO) / International Electrotechnical Commission (IEC) 27000 series ↗, the National Institute of Science and Technology (NIST) Cybersecurity Framework ↗, and the Center for Internet Security (CIS) Critical Security Controls ↗.

While standards are recommended, regulations are enforced. Procedures and standards must be regulatorily compliant.[21][22] Standards are often included "by reference" in U.S. federal regulations,[23] which means that publishing of regulations "in the Federal Register and the Code of Federal Regulations (CFR) by referring to materials already published ↗ elsewhere" is lawful, as long as those materials, like international standards documents, can be accessed by the public. Since some of the standards of organizations like ISO can be hundreds of pages long and are accessible in electronic form, this approach is sensible. Regulations are legally binding rules imposed by governments or regulatory authorities that are mandatory and specify

penalties, fines, or other consequences for non-compliant entities. Examples of cybersecurity and privacy regulations include the GDPR in the European Union, and the CCPA and FISMA in the United States.

The following are criteria for evaluating the use of standards and regulations:

- **Flexibility:** Standards generally offer more flexibility than regulations, allowing organizations to tailor their approaches to their specific needs, whereas regulations tend to be more prescriptive. Regulations often follow standards set by industry and academia.
- **Speed of development:** Standards can often be developed more quickly than regulations since they go through fewer steps than required by federal agencies.
- **Enforcement:** Regulations may provide a stronger enforcement mechanism than standards because they come with legal consequences for non-compliance.
- **International harmonization:** Standards can help facilitate international cooperation and consistency because they can be adopted across borders more easily than regulations, which are typically specific to a particular jurisdiction.

Considering these factors, a combination of standards and regulations may be the most timely and effective approach in many cases. Regulations can establish a baseline of mandatory ML security and ethics requirements, while standards can provide guidance on best practices and help organizations stay current with the latest developments in the field.

Example of prior standards in software: Cybersecurity standards

In the United States, the process of developing cybersecurity standards often involves collaboration among the federal government, industry stakeholders, and other experts. This collaborative approach helps ensure that the resulting standards are practical, effective, and widely accepted.

One key example of this is the development of the NIST Cybersecurity Framework. NIST ↗, an agency within the U.S. Department of Commerce, plays a key function in developing cybersecurity standards and guidelines, yet is not endowed with

regulatory enforcement capabilities. NIST often solicits input from industry stakeholders, academia, and other experts to ensure that its guidance is comprehensive and current. Following the issuance of Executive Order 13636 ↗, "Improving Critical Infrastructure Cybersecurity," in 2013, NIST was tasked with developing a framework "to help organizations better understand, manage, and communicate cybersecurity risks."[24] To do so, NIST engaged in an open, collaborative process that included the following:

- **Request for Information (RFI):** NIST issued an RFI ↗ to gather input from stakeholders on existing standards, guidelines, and best practices in cybersecurity. This allowed companies and experts to share their knowledge and experiences with NIST.
- **Workshops:** NIST organized a series of workshops ↗ across the country to gather additional input from stakeholders and foster discussion on key cybersecurity issues. These workshops provided opportunities for participants to engage in face-to-face discussions with NIST and other experts.
- **Public review and comments ↗:** NIST allowed stakeholders to provide feedback on the proposed framework and suggest improvements.
- **Ongoing collaboration:** After the release of the Cybersecurity Framework, NIST continued to collaborate with stakeholders through workshops, webinars, and other events to gather feedback and keep the framework current.

These examples demonstrate the value of public-private partnerships in the development of cybersecurity standards. By involving industry stakeholders in the process, the federal government can help ensure that the resulting standards are practical, widely accepted, and effective in addressing the challenges posed by cybersecurity threats.

Regulations have also played a significant role in governing cybersecurity and privacy and have often been used to set mandatory requirements for organizations to protect sensitive information and ensure privacy.

General Data Protection Regulation (GDPR) – European Union: GDPR, implemented in 2018, is widely recognized as one of the most comprehensive data protection regulations worldwide. It requires organizations to protect the personal data of EU

citizens, ensuring privacy and security. GDPR has been effective in raising awareness about data protection and pushing organizations to improve their cybersecurity posture and has been seen "as setting a global standard"[25] for user privacy. A research report ↗ on GDPR impacts published by the UK government found that "GDPR compliance had resulted in more investment in cybersecurity" by a majority of surveyed European businesses and that organizations generally prioritized cybersecurity following the new regulations. However, challenges include the complexity of the regulation as well as high compliance costs (further details can be found here ↗).

Health Insurance Portability and Accountability Act (HIPAA) – United States: HIPAA ↗, passed in 1996, sets standards for protecting sensitive patient data and requires health care organizations to implement cybersecurity measures for electronic protected health information (ePHI) . Although HIPAA has been successful in improving the safeguarding of patient data, it has faced criticism for being overly complex for patients who may assume it applies to contexts where it may not offer protections (such as in health mobile apps) and for the fact that patients and caregivers of patients can have difficulty accessing necessary records.[26] Furthermore, when cybersecurity breaches of private health data occur, it may be difficulty for consumers to know what options they have for recourse. The law as was written in 1996 may require updating in the face of rapidly evolving cybersecurity threats.

California Consumer Privacy Act (CCPA) – United States: Implemented in 2020, the CCPA ↗ grants California consumers specific data privacy rights, such as the "right to know" what information is stored, as well as an option for a consumer to "opt-out" of data sharing. The CCPA has been praised for empowering consumers and raising the bar for privacy protection in the United States. For companies that have customers in California and other states, the CCPA has resulted in a standard for consumer privacy rights that will likely be applied by companies to other states and create "dynamics that contribute to shaping the U.S. privacy regulatory framework."[27] However, the CCPA does face criticism for several issues including its complexity: the nine exceptions to the right of consumers to delete data may not give consumers the protection that they expect; the burden it places on businesses; and the potential conflicts with other state or federal privacy regulations.[28] Some lessons may be drawn from the current complexity of privacy laws to regulation of algorithmic fairness: Consumers may not have the time to read every opt-out notice or legal

disclaimer and understand on the spot what rights they may be giving up or gaining from accepting terms of service.

Federal Information Security Management Act (FISMA) – United States: Signed into law in 2002, FISMA ↗ "requires ↗ each federal agency to develop, document, and implement an agency-wide program to provide information security for the information and systems that support the operations and assets of the agency." Although FISMA has led to improved cybersecurity within federal agencies, it has been criticized for being overly focused on compliance rather than continuous risk management and for not keeping pace with the evolving threat landscape. FISMA does establish generally applicable principles of cybersecurity in government, such as requiring the NIST to establish ↗ "federal information processing standards that require agencies to categorize their information and information systems according to the impact or magnitude of harm that could result if they are compromised," which are codified in NIST standards ↗. Broad principles like these are sensible for businesses to adopt as well (for example, at the business-unit or organization-unit levels).

Although regulations can be effective in setting mandatory requirements for cybersecurity and raising awareness, they may also face challenges such as complexity, high compliance costs, outdated provisions, and potential conflicts with other regulations. To address these issues, policymakers should consider periodically reviewing and updating regulations to ensure they remain relevant and effective in the face of rapidly evolving cybersecurity threats. Additionally, harmonizing regulations across jurisdictions and providing clear guidance to organizations can help alleviate some of the challenges associated with compliance. Existing cybersecurity regulations may be a template for ML fairness regulations, as well.

A similar approach can be applied to ML ethics, where regulations can set a legal framework and minimum requirements for ethical ML development and deployment, while standards can provide detailed guidance on best practices, allowing for flexibility and adaptation to new technological advancements. (#_ftnref1)

AI standards and regulations

Until recently, there have been no comprehensive AI-specific regulations in the United States. However, there have been efforts to establish guidelines, principles, and standards for AI development and deployment, both by the U.S. government and by various organizations and industry groups. Some examples include the following:

Executive Order on Maintaining American Leadership in Artificial Intelligence (2019): Issued by the White House in 2019, this order ↗ aimed to promote sustained investment in R&D and enhance the United States' global leadership in AI. Although it did not establish specific regulations, it directed federal agencies to create a national AI strategy and develop guidance for AI development and deployment.

Defense Innovation Board (DIB) AI Principles (2019): The DIB is an advisory committee to the U.S. Department of Defense (DoD). In 2019, it released a set of ethical principles ↗ for AI in defense, covering areas such as responsibility, traceability, and reliability. Although they are not legally binding, these principles provide a basis for the ethical deployment of AI within the DoD.

Organization for Economic Co-operation and Development (OECD) AI Principles (2019): In 2019, the OECD established a set of standards ↗ for use of AI that are respectful of human rights and democratic values.

NIST AI Risk Management Framework (2023): NIST has been active in AI research and standardization efforts, focusing on topics such as trustworthy AI, AI testing and evaluation, and AI risk management. In 2023, NIST published the AI Risk Management Framework ↗ (AI RMF 1.0), which aims to provide a systematic approach to managing risks associated with AI systems. This framework, once finalized, could serve as a foundation for future AI standards and guidelines in the United States.

Partnership on AI (PAI) Tenets: The PAI is a multi-stakeholder organization ↗ that brings together industry, academia, and civil society to develop best practices for AI technologies. The partnership has published a set of tenets ↗ to guide AI research and development, including principles such as ensuring AI benefits all, prioritizing long-term safety, and promoting a culture of cooperation.

Industry-specific guidelines and standards: Several organizations and industry groups have developed their own guidelines and principles for AI development and deployment within specific sectors, such as health care, finance, and transportation. For example, PrivacyCon ↗ is an annual conference to bring together industry, government, and academia that serves the development of such guidelines – in 2020 the theme of the event was health data. In a recent article ↗, Accenture's global health industry lead provided some "best practices" for "generative AI in healthcare" which are likely just a beginning of guidelines as generative AI grows in adoption in that industry. Bain and Company put together "design principles ↗" for the use of "generative AI in financial services," again likely a topic of growing interest in the coming years. In the transportation industry, a wide consortium of auto manufacturers, chipmakers, and other industry members put together guidelines ↗ for "automated driving" back in 2019. These examples of guidelines, although they are not legally binding, can help set expectations and establish best practices for AI governance within their respective industries. The National AI Institute at the Department of Veterans Affairs ↗ has also been building on and harmonizing these frameworks for trustworthy AI and operationalizing them in the health care sector.

AI-related concerns, such as data privacy and algorithmic fairness, may be addressed by existing regulations and guidelines that are not AI-specific, such as GDPR, CCPA, and guidelines ↗ on algorithmic transparency and fairness from the Federal Trade Commission (FTC). As AI continues to evolve, it is likely that more targeted regulations and standards will be developed to address AI-specific concerns and ensure ethical and responsible AI governance. Comprehensive AI auditing processes will need to be developed in a timely manner and updated periodically. Additionally, a system of incentives may be needed to encourage companies to actively develop tools to address and solve AI fairness concerns.

Discussion and recommendations

Standard-setting bodies work well when there are mechanisms for accountability built-in. For instance, external audit committees[29] can provide an accountability mechanism as long as the audits are performed periodically (e.g., quarterly or annually) and if the auditors are not influenced by the position of those who are being audited (no "revolving door" scenario). To ensure accountability, such auditors may be hired by testing organizations such as the Technical Inspection Association (TUV),

Underwriters Laboratories (UL), or Intertek, among others. Alternatively, auditors may be part of a volunteer community, similar to code maintainers in the Linux open-source community who control the quality of any changes to the codebase. Therefore, we suggest creating fairness audits by external auditors to the firm and codifying the type of audit and frequency in an industry standard.

We propose an approach where regulations complement industry standards by adding to them tools for enforcement, rather than as a one-size-fits-all tool. Unfortunately, firms—at least right now—do not have a strong commercial incentive to pursue ML fairness as a goal of product development and incur additional liability in the absence of agreed upon standards. In fact, it is well known that standards may not reach their potential if they are not effectively enforced.[30] Since consumers do not currently have widespread visibility into which products abide by fairness criteria and fairness criteria are not yet accessible to the general consumer since they require specialized knowledge (e.g., data science and programming skills in addition to knowing the literature), it is perhaps not feasible that a majority of consumers could themselves test for unfairness in a product or service. Furthermore, it is not yet the case that most consumers have access to training sets or company proprietary algorithms to prove whether they have been harmed by an ML system, which is required for damages under the newest regulations such as the EU AI Act (see the commentary in Heikkilä, 2022 ↗). The literature on ML fairness is a complex, multidisciplinary one, so computer scientists, lawyers, ethicists, and business scholars are needed to be part of driving regulations.

Under such circumstances, it is not surprising that companies do not perceive a direct financial incentive to maintain specialized staff to audit ML fairness or supervise with a fairness-oriented goal the development of ML based products, especially in a downturn in the markets. Recently, many leading companies have unfortunately laid off ↗ a number of specialized ML fairness engineering staff, in some cases closing entire departments, which results in loss of company-specific knowledge and will mean a much slower adoption of fairness principles in industries in the future. Although regulations can provide general principles (for example, meeting at minimum an equality of opportunity fairness criterion and performing an annual fairness audit; see the guidelines ↗ by the Consumer Financial Protection Bureau (CFPB) in auditing compliance to the Equal Credit Opportunity Act [ECOA] in the United States as well as expectations regarding transparency in algorithmic decision-making ↗ when related to credit applications) and provide some consumers relief in cases of egregious violations

of basic fairness criteria, they are insufficient to provide incentives to companies to perform better than such minimums and do not incentivize companies to innovate beyond meeting the regulatory requirements. Although companies may wish to implement fair ML as part of every stage of their software development processes to ensure they meet the highest ethical standards,[31] and we encourage that in related work,[32] we recognize that people, and thus companies at large, respond to incentives and the current "rules of the road" are still in their infancy when it comes to ML.

Market forces can provide incentives to companies to innovate and produce better products provided the advantages of the innovation are clear to a majority of consumers. A consumer may choose a product and pay more if it satisfies a fairness standard set by a leading, recognizable standard-setting body and if the benefits of that standard are apparent. For example, a consumer may prefer a lender that markets itself as ensuring no discrimination based on subgroup fairness (i.e., combinations of categories of race, age, gender, and/or other protected attributes) if the alternatives only guarantee group-level fairness. If the consumer is aware of the higher standard this product is satisfying—for example through a standard that the product is displaying—the consumer may choose to pay a higher price for two feature-equivalent products if one satisfies a fairness standard. Thus, we call on the industry and organizations such as the Information Systems Security Association (ISSA), the IEEE, the Association for Computing Machinery (ACM), and the ISO, among others, to invest in developing an ML fairness standard, communicate their rationale, and interact with policymakers over these standards as they deploy them into products over the next five years.

We also suggest that firms create bug bounty programs ↗ for fairness errors, in which users themselves can file bug reports with the producer of the ML system, much like what exists in cybersecurity. For example, if the ML system is a voice assistant that often misunderstands the user based on a speech impediment due to a disability, the user should be able to report that experience to the company. In another example, a user utilizing an automated job resume screening tool (as some companies now have implemented in their hiring processes) who gets consistently denied and suspects the reason may be because of a protected attribute, should be able to request a reason from the service provider. In yet another example, a mobile phone application allowing the user to test for melanoma by taking a picture should allow the user to report false positives and false negatives following consultation with a physician should such consultation prove the application misdiagnosed the user, which would allow the

developers to diagnose the root cause, which may include information covered by protected attributes. A researcher or independent programmer should also be able to report bugs or potential fairness issues for any ML-based system and receive a reward if that bug report is found to reveal flaws in the algorithm or training set related to fairness. (#_ftnref1)

Conclusions

In this report, we shared some background information on the issues of ML fairness and existing standards and regulations in software. Although ML systems are becoming increasingly ubiquitous, their complexity is often beyond what was considered by prior privacy and cybersecurity standards and laws. Therefore, we expect that what norms should specify will be an ongoing conversation requiring both industry collaboration through standardization and new regulations. We recommend a complementary approach to fairness by creating a fairness standard via a leading industry standard-setting body to include audits and mechanisms for bug reporting by users and a regulation-based approach wherein generalizable tests such as group-fairness criteria are encoded and enforced by national regulations.

AUTHORS

Mike H. M. Teodorescu **Assistant Professor** - University of Washington
@miketeod

Christos Makridis **Associate Research Professor** - Arizona State University

Footnotes

1. Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2022. "A Survey on Bias and Fairness in Machine Learning." *ACM Computing Surveys* 54 (6): 1–35. https://doi.org/10.1145/3457607 ↗.

2. Chouldechova, Alexandra, and Aaron Roth. 2020. "A Snapshot of the Frontiers of Fairness in Machine Learning." *Communications of the ACM* 63 (5): 82–89. https://doi.org/10.1145/3376898 ↗.

3. Teodorescu, Mike, Lily Morse, Yazeed Awwad, and Gerald Kane. 2021. "Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation." *MIS Quarterly* 45 (3): 1483–1500. https://doi.org/10.25300/MISQ/2021/16535 ↗.

4. Kusner, Matt, Chris Russell, Joshua Loftus, and Ricardo Silva. 2019. "Making Decisions That Reduce Discriminatory Impacts." In *Proceedings of the 36th International Conference on Machine Learning*, edited by Kamalika Chaudhuri and Ruslan Salakhutdinov, 97:3591–3600. Proceedings of Machine Learning Research. PMLR. https://proceedings.mlr.press/v97/kusner19a.html ↗.

5. Wang, Xiaomeng, Yishi Zhang, and Ruilin Zhu. 2022. "A Brief Review on Algorithmic Fairness." *Management System Engineering* 1 (1): 7. https://doi.org/10.1007/s44176-022-00006-z ↗.

6. Alikhademi, Kiana, Emma Drobina, Diandra Prioleau, Brianna Richardson, Duncan Purves, and Juan E. Gilbert. 2022. "A Review of Predictive Policing from the Perspective of Fairness." *Artificial Intelligence and Law* 30 (1): 1–17. https://doi.org/10.1007/s10506-021-09286-4 ↗.

7. Hardt, Moritz, Eric Price, and Nathan Srebro. 2016. "Equality of Opportunity in Supervised Learning." In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 3323–31. NIPS'16. Barcelona, Spain.

8. Makhlouf, Karima, Sami Zhioua, and Catuscia Palamidessi. 2021. "On the Applicability of Machine Learning Fairness Notions." *ACM SIGKDD Explorations Newsletter* 23 (1): 14–23. https://doi.org/10.1145/3468507.3468511 ↗.

9. Morse, Lily, Mike Horia M. Teodorescu, Yazeed Awwad, and Gerald C. Kane. 2022. "Do the Ends Justify the Means? Variation in the Distributive and Procedural Fairness of Machine Learning Algorithms." *Journal of Business Ethics* 181 (4): 1083–95. https://doi.org/10.1007/s10551-021-04939-5 ↗.

10. Brodersen, Kay Henning, Cheng Soon Ong, Klaas Enno Stephan, and Joachim M. Buhmann. 2010. "The Balanced Accuracy and Its Posterior Distribution." In *2010 20th International Conference on Pattern Recognition*, 3121–24. Istanbul, Turkey: IEEE. https://doi.org/10.1109/ICPR.2010.764 ↗.

11. See Makhlouf, Zhioua, and Palamidessi, "On the Applicability of Machine Learning Fairness Notions."

12. Chouldechova, Alexandra. 2017. "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments." *Big Data* 5 (2): 153–63.

https://doi.org/10.1089/big.2016.0047 ↗.

13. Shahriari, Kyarash, and Mana Shahriari. 2017. "IEEE Standard Review — Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems." In *2017 IEEE Canada International Humanitarian Technology Conference (IHTC)*, 197–201. Toronto, ON, Canada: IEEE. https://doi.org/10.1109/IHTC.2017.8058187 ↗.

14. Tyler, Tom R. 1997. "Procedural Fairness and Compliance with the Law." *Swiss Journal of Economics and Statistics (SJES)* 133 (II): 219–40.

15. Van Den Bos, Kees, Henk A. M. Wilke, and E. Allan Lind. 1998. "When Do We Need Procedural Fairness? The Role of Trust in Authority." *Journal of Personality and Social Psychology* 75 (6): 1449–58. https://doi.org/10.1037/0022-3514.75.6.1449 ↗.

16. Brockner, Joel, Ya-Ru Chen, Elizabeth A. Mannix, Kwok Leung, and Daniel P. Skarlicki. 2000. "Culture and Procedural Fairness: When the Effects of What You Do Depend on How You Do It." *Administrative Science Quarterly* 45 (1): 138–59. https://doi.org/10.2307/2666982 ↗.

17. See Morse et al., "Do the Ends Justify the Means?," p. 1089.

18. See Morse et al., "Do the Ends Justify the Means?," p. 1089.

19. Barrett, Catherine. 2019. "Are the EU GDPR and the California CCPA Becoming the de Facto Global Standards for Data Privacy and Protection?" *Scitech Lawyer* 15 (3): 24–29.

20. See, for example, some of the papers cited in Teodorescu et al., "Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation."

21. Peres, S. Camille, Noor Quddus, Pranav Kannan, Lubna Ahmed, Paul Ritchey, William Johnson, Samina Rahmani, and M. Sam Mannan. 2016. "A Summary and Synthesis of Procedural Regulations and Standards—Informing a Procedures Writer's Guide." *Journal of Loss Prevention in the Process Industries* 44 (November): 726–34. https://doi.org/10.1016/j.jlp.2016.08.003 ↗.

22. Landoll, Douglas J. 2017. *Information Security Policies, Procedures, and Standards: A Practitioner's Reference*. 1st edition. CRC Press.

23. Mendelson, Nina. 2014. "Private Control over Access to Public Law: The Perplexing Federal Regulatory Use of Private Standards." *Michigan Law Review* 112 (5): 737–807.

24. Bartock, Michael, Joseph Brule, Ya-Shian Li-Baboud, Suzanne Lightman, James McCarthy, Karen K. Reczek, Doug Northrip, Arthur Scholz, and Theresa Suloway. 2021. "Foundational PNT Profile: Applying the Cybersecurity Framework for the Responsible Use of Positioning, Navigation, and Timing (PNT) Services." NISTIR 8323. Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8323 ↗.

25. Kalman, Laurence. 2019. “New European Data Privacy and Cyber Security Laws: One Year Later.” *Communications of the ACM* 62 (4): 38–38. https://doi.org/10.1145/3310326 ↗.

26. Solove, Daniel. 2013. “HIPAA Mighty and Flawed: Regulation Has Wide-Reaching Impact on the Healthcare Industry.” *Journal of American Health Information Management Association* 84 (4): 30–31.

27. Baik, Jeeyun (Sophia). 2020. “Data Privacy against Innovation or against Discrimination?: The Case of the California Consumer Privacy Act (CCPA).” *Telematics and Informatics* 52 (September): 101431. https://doi.org/10.1016/j.tele.2020.101431 ↗.

28. Palmieri III, Nicholas F. 2020. “Who Should Regulate Data?: An Analysis of the California Consumer Privacy Act and Its Effects on Nationwide Data Protection Laws.” *Hastings Science and Technology Law Journal* 11 (1): 37–60.

29. Tarafdar, Monideepa, Mike Teodorescu, Hüseyin Tanriverdi, Lionel P. Robert, and Lily Morse. 2020. “Seeking Ethical Use of AI Algorithms: Challenges and Mitigations.” In *ICIS 2020 Proceedings*. https://aisel.aisnet.org/icis2020/panels/panels/1 ↗.

30. Magat, Wesley A., and W. Kip Viscusi. 1990. “Effectiveness of the EPA’s Regulatory Enforcement: The Case of Industrial Effluent Standards.” *The Journal of Law and Economics* 33 (2): 331–60. https://doi.org/10.1086/467208 ↗.

31. Martin, Kirsten. 2013. “Designing Ethical Algorithms.” *MIS Quarterly Executive* 18 (2): 129–42.

32. See Teodorescu et al., “Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation.”

Copyright 2024 The Brookings Institution