Regulation and Artificial Intelligence Ethics: The State of Play

Onur Bakiner, Ph.D.
Faculty Fellow, Center for Business Ethics
Associate Professor of Political Science
Seattle University
June 28, 2022

SEATTLEU
ALBERS SCHOOL OF BUSINESS AND ECONOMICS
CENTER FOR BUSINESS ETHICS

Executive Summary

This discussion identifies major risks of Artificial Intelligence (AI) systems, introduces the reader to the terms of today's debates on the ethical uses of AI, defined as attempts to understand, criticize, reform and transform AI systems for social good, and then examines proposed solutions to mitigate or eliminate these risks. It pays specific attention to the interaction of business- and industry-level solutions with government regulation and draws upon the latest academic research on AI ethics as well as news coverage of the topic.

The main findings include:

- Although there is divergence in what is understood as the risks of AI systems and how they should be addressed, some fundamental values have emerged. Accuracy, privacy, non-discrimination, transparency, explainability, trust, responsibility and sustainability stand out as a basic (but by no means exhaustive) list of fundamental values guiding the implementation of ethics in AI.

- Adherence to ethical AI principles through companies' internal ethics boards has yielded positive results in some companies, but high-profile cases of unethical behavior reveal the limits of self-regulation. A mixture of movements within companies by employees and engineers, along with external pressures from other stakeholders, has increased the likelihood of compliance in some cases.

- U.S. companies have straddled between seeking broad exemptions from direct regulation and making themselves regulation-compliant, in great part as a response to the European Union's (EU) relevant laws and proposed bills.

- As the race to regulate AI has gained pace on both sides of the Atlantic, researchers' and practitioners' perspectives on ethical and responsible AI will increasingly be incorporated into policy debates.

Introduction

The application of ethical standards in the responsible development and use of artificial intelligence (AI) technologies, hereafter referred to as AI ethics, is fast becoming an integral part of the larger call for greater ethical responsibility in business. A panel survey of 21 expert academics and practitioners finds a slight majority that agrees a company's responsible AI efforts should be tied to its corporate social responsibility work; another survey of over 1,000 executives in major firms finds that executives think this linkage is already taking place (Renieris et al., 2022). Dozens of ethical (or responsible) AI guidelines have been published in the past decade by private businesses, academic researchers, government entities, and non-governmental organizations (NGOs). Does this mean there are, and should be, fundamental values that guide ethical conduct in the design, development and deployment of AI systems?

Skeptics point out that AI companies describe their product with words like "ethical" and "responsible" to deflect attention away from criticism and to control the terms of the debate (Greene et al., 2019; Hickok, 2021; Metcalf et al, 2019). While this critique captures an essential truth about some of the uses and abuses of AI ethics, it can also be argued that a community of developers, researchers and activists with bona fide commitment to shared principles and practices has emerged in the recent past. Thus, this paper summarizes state-of-the-art conversations in AI ethics as an emerging academic and practical field that brings in enormous conceptual, epistemic and methodological diversity, and some shared concepts, meanings and methods.

Human-AI Interaction

Since the very early days of AI research, one way of thinking about AI has been in terms of conflict and cooperation between humans and machines (Friend, 2018). The darkest speculation, of course, is some form of Artificial General Intelligence, no longer subject to human oversight, unilaterally deciding to eradicate the human race. Another idea of the human-machine relationship that appears more realistic in the present envisions some sort of mutual coexistence in which increasingly capable machines interact with humans in sophisticated ways (Stone et al., 2016). What happens when inevitable conflicts arise? Some advocate a defense of humans in the face of potential threats by smart machines at any cost and under any circumstance. Underlying this human-centered approach are doubts around the notion that AI systems can experience sentience, trust, love, harm, and so on in ways that humans do (Baria & Cross, 2021; Ryan, 2020). Others argue that if AI systems approach sentience, they should be seen not merely as tools, but rather as beings endowed with a set of rights (Gunkel, 2018; Heller, 2016).

The human-machine interaction has also been conceptualized as a labor-market concern. One prediction is that a large number of jobs, including ones that rely on cognitive or affective labor, will be lost to computer systems capable of improving their performance through automatic learning (Walz & Firth-Butterfield, 2018). The job-replacement debate brings out pessimism around the future of the labor market, as well as optimism in what technology can do: perhaps the takeover by AI will lead to mass unemployment, but others hope that it will eliminate only the most tedious of jobs, while liberating humans to take up new and boundlessly more interesting jobs whose content we cannot even imagine today. According to this view, artificial intelligence will "augment" rather than "supplant" human intelligence (Markoff, 2016).

Ethical Concerns and Shared Values

AI as a Source of Moral Dilemmas

Futuristic warnings about massive disruption are only one part of debates within AI ethics. Another way of framing the question of values in AI involves hard ethical questions. Echoing different versions of the Trolley Problem, this perspective considers AI problems as zero-sum situations involving the possibility of real-time sacrifices. Should driverless cars prioritize the lives of passengers over those of passersby? What is an acceptable rate of success before we let robot surgeons take over from human ones? Lethal autonomous weapons systems from human soldiers? How many errors are too many? Do failures justify total bans, or should we grant AI systems as much lenience as we do human decision makers? For those interested in working through – and creating – scenarios involving machine intelligence, an MIT team has put together Moral Machine (moralmachine.net), a massive Internet survey that captures popular responses to ethical dilemmas (Awad et al., 2018).

The Social Impact of Algorithms

Views of AI as an existential threat or a source of dilemmas will always animate the imagination, but AI ethics has evolved in the past decade to cover far-reaching economic, social and political problems (Coeckelbergh, 2020). Consider one type of AI system, powered by machine-learning algorithms and now ubiquitous in business. Algorithmic decision-making may weed out successful candidates for school admissions or jobs (Dastin, 2018); falsely predict some individuals will commit crimes again when granted parole (Angwin & Larson, 2016); fail to recognize faces in situations where misrecognition may affect a person's access to public spaces (Tucker, 2017); and spread outright lies (Milmo, 2021). Worse yet, those affected by AI's failures tend to be those who already face discrimination, exclusion and marginalization – the poor, ethnic and racial minorities, women, gender minorities, and the disabled (Eubanks, 2018; Noble, 2018; O'Neil, 2016).

What, or who, is at fault here? One explanation is that the datasets that power algorithms reflect societal biases. Chatbots pick up the sexist and racist things people say in the dark – and not-so-dark – corners of the Internet; companies that have historically hired men disproportionately signal equivalence between information found in men's resumés and professional success to unassuming mathematical functions; racist bias ends up embedded in datasets to mirror a society that incarcerates its Black members disproportionately. What makes algorithmic output even more problematic is the fact that in the eyes of most people, biased results are blessed with an aura of objective truth simply because their production involves complex mathematics.

Biased datasets that in turn mirror a biased society definitely constitute part of the problem. A closer look into how algorithms work suggests that they do more than reflect bias, however. Some algorithms amplify existing problems by channeling them to ever newer audiences (LaFrance, 2021). Anyone can attest to having queried a simple term, like "COVID-19" or "elections", in order to read or watch more about a matter of interest, only to find oneself exposed to some of the most unhinged conspiracy theories imaginable three clicks later. A recommendation algorithm designed to maximize engagement would do just that, with no regard to the quality or desirability

of the content. Thus, if producers of AI systems fail to consider the potential for conflict between what algorithms optimize (which ultimately comes down to the company's bottom line) and the social good, serious risks arise.

Data and Transparency

AI ethics involves more than debating the merits of outcomes; the procedure matters, too. The complex algorithms used to process large chunks of data do not lend themselves to easy interpretation or explanation; in more technical terms, neither the numbers embedded in the "hidden layers" of neural networks that convert data input into output nor parameters that should be fine-tuned for accurate results make much intuitive sense.

Algorithmic complexity becomes a problem when those adversely affected by the output cannot appeal to an algorithm's opaque "reasoning" process – an idea captured in the term *contestability* (Lyons et al., 2021). Controversy about a recidivism-prediction algorithm, for example, reached a climax when a defendant's appeal before the Wisconsin Supreme Court was rejected on the grounds that the district court's decision was not merely automatic in *State of Wisconsin v. Eric L. Loomis* (State v. Loomis, 2016). The ruling did not find any reason to scrutinize the mathematical iterations that led the algorithm to recommend no parole; however, one concurring opinion asked to maintain the distinction between considering algorithmic output and relying on it, and another asked a court to take into account the strengths and weaknesses of the relevant technology (Elrod, 2020).

The Court's decision not to get involved with the nuts and bolts of a proprietary algorithm only further galvanized explainable AI advocates to force the producers of software with potentially life-transforming powers to make their data and algorithms more transparent. Naturally, businesses want to capitalize on proprietary data and algorithms, but the combination of secrecy and complex algorithms challenge the fundamentals of business law and ethics, such as the "reasonableness standard" (i.e., the idea that companies should not impose unreasonable risks to clients): "As a result of the secrecy, we know little of what individual companies have learned about the errors and vulnerabilities in their products. Under these circumstances, it is impossible for the public to come to any conclusions about what kinds of failures are reasonable or not" (Selbst, 2020).

The broader ecosystem of data production and processing within which AI is embedded has come under scrutiny, as well. Companies cannot always find all the data they need by adhering to privacy, consent, confidentiality and quality standards, which tempts them to commit privacy violations – as a recent court settlement in which Clearview AI has agreed not to sell its database of photos at will suggests (Mac & Hill, 2022). Thus, considerations of data privacy, which of course may arise with or without AI, are typically treated as an integral part of AI ethics because AI companies so often stretch ethical standards to collect, store and process more data.

AI systems may produce negative impact on the ecosystem in a more literal sense: data-intensive algorithms use enormous compute power, thereby consuming energy and leaving behind carbon footprint. As companies with virtually limitless access to data and vast warehouses have taken over state-of-the-art research in fields like natural language processing, critical voices question if breakthroughs in practical applications like chatbots and translators result less from improvements

in algorithmic efficiency than what can be called "throwing compute power at the problem". As a result of these worries, sustainability in AI has become one of the overarching themes advocated by ethicists (Crawford, 2021; van Wynsberghe, 2021).

Emergent Shared Principles

In conclusion, concepts, problem definitions and proposed solutions associated with the mitigation of these ethical risks in AI do not always generate consensus. This is in great part because different actors have different priorities (Rudschies et al., 2020). What complicates the picture is the high degree of correspondence between AI ethics and concepts with a much longer history, such as machine ethics, robot ethics, digital ethics, bioethics, and data privacy (Gordon, 2020; Hanna & Kazim, 2021; Segun, 2021; Wallach & Allen, 2009), not to mention human rights, Sustainable Development Goals, and so on (Risse, 2019; Stahl et al., 2021; Vinuesa et al., 2020). In addition, every single term in the AI ethics vocabulary (accountability, explainability, fairness, privacy, responsibility, sustainability, transparency, trust) is essentially contested, and therefore open to multiple and at-times contradictory definitions. Nonetheless, some shared ideas do exist (for conceptual scholarship, see: Berendt (2019); Canca (2020); Floridi et al. (2018); Morley et al. (2021)). Summarizing current debates, below is a minimal, and therefore non-exhaustive, list of shared principles that have been voiced by academics, businesses, governments and advocacy groups:

- Algorithmic output should be accurate at a level that justifies its use
- Data collection and storage should respect data subjects' privacy
- Data collection and analysis should not cause or amplify discrimination against vulnerable and marginalized groups in society
- The data and algorithms that make up an AI system should be made as transparent and explainable as possible
- Trust should be at the foundation of relationships between AI companies and stakeholders, including the end users, data subjects and others affected by AI systems
- When an AI system produces harmful effects, responsibility should be traceable
- Awareness of algorithms' environmental impact should guide their use

Enforcing Shared Principles

Debates about what ethical AI entails naturally spill over to questions about who should take charge to mitigate or eliminate risks. AI systems themselves? Researchers? Companies? Governments? So far, the answer seems to be that a growing research community has carried the conversation forward on accountable, fair, responsible and transparent AI. Company practices have varied with respect to the line of production, the centrality of AI systems to the business model, and company culture. Governments have been relatively slow to regulate AI, but a series of EU laws is about to change that. In what follows, company practices and legal regulation are discussed.

Self-Regulation and Business Culture

Some companies producing AI systems have adopted ethics guidelines or established in-house ethics boards to self-regulate. Their motivations include preemptively keeping government regulation at bay, protecting the company's reputation from potential scandals, and responding to criticism when things indeed go wrong (Hagendorff, 2020).

Failures of the self-regulation model, especially after the fallout at Google's Ethical Artificial Intelligence team, when Timnit Gebru's contract was terminated in December 2020 and co-lead Margaret Mitchell was dismissed two months later, have without doubt drawn public attention. A recent survey of 82 AI professionals finds a "significant responsibility gap between the beliefs and the actions of the AI professionals" (Nersessian & Mancha, 2020). Of course, high-profile failures receive more media attention than quiet cases of ethical behavior, for example when an ethics board recommends against a lucrative yet ethically questionable pursuit, and the company as a whole agrees. All in all, however, even at its best, in-house regulation should be seen as a necessary yet insufficient mechanism for ethical behavior (for a complex discussion of what in-house teams can and cannot do, see: Lazzaro (2021)).

Restricting self-regulation to company-approved guidelines and boards misses out another way in which employees make a difference: protest action in the form of whistleblower disclosures, walkouts and petitions have sometimes forced companies to reconsider the ethical implications of their business. In June 2018, Google announced it would not renew a Department of Defense contract for Project Maven, a month after employees sent a letter to the CEO warning against the project's weaponization of AI (Latonero, 2018). Around the same time, a coalition of civil-rights organizations and employees also pressured Amazon to stop selling facial-recognition software to police departments and immigration enforcement (Molnar, 2019) – to no avail.
In addition to protest behavior, whistleblowers have at times exposed companies' disregard for ethical and legal standards, evidenced by the string of Facebook investors and employees who spoke out in the past decade (Fried, 2021); as a result, whistleblower protection is recommended as a key mechanism to uncover, and hopefully address, algorithmic bias (Katyal, 2019). All in all, organizational culture is a main factor in understanding the potential for, and limits of, ethical behavior in companies (Lauer, 2021).

European Regulation

It is often argued that today's AI systems have been developed in the absence of legal regulation, especially in the United States (Denno & Surujnath, 2019). A more accurate way of rephrasing the argument is that existing statutes on data privacy, consumer protection, business liability and digital content moderation provide little guidance in the face of challenges posed by the combination of data availability and efficacious machine learning algorithms. Section 230 of the Communications Decency Act of 1996 has not only survived the Act itself, but has in many ways shaped the Internet we have today (Edelman, 2021) – yet it is from a time when Google did not exist, Amazon was an online bookstore, and the Internet itself was known to very few people around the world. Across the Atlantic, the European Union's (EU) *Data Protection Directive* regulated data privacy from 1995 to 2018, when it was finally supplanted by the *General Data Protection Regulation* (GDPR). Around the globe, legal regulation of AI began to gain momentum only in the second half of the 2010s, presumably as a result of the multiple high-profile

controversies, some of which were described earlier, but even then, the rate of legislative change has remained comparatively slow (Larsson, 2020).

At the time of this discussion, the legal regulation of AI systems looks more likely than ever. The EU's multifaceted approach to AI regulation is the closest any political entity has come to setting specific rules on the development of AI systems. The *AI Regulation*, proposed by the European Commission in April 2021, will be the first legal document targeting AI systems specifically. In addition, the EU's political bodies have come close to legislating *the Digital Services Act*, *Data Governance Act*, and *Digital Markets Act*, which will regulate the publication of digital content, data sharing, and fairness in digital competition, respectively. Thus, the European digital sphere may look quite different as early as 2024, with implications for all companies doing business in the EU, American ones included. So far, the United States has by and large maintained the status quo (for example, Section 230 is still in effect despite criticisms), with legal change taking the form of targeted and limited alterations while more comprehensive legislative proposals are failing to make it to Congressional debates. Whatever happens in Washington, D.C., American companies will still have to adapt to EU law, unless they opt out of the European market altogether.

The American Response

The regulatory disconnect across the Atlantic raises the prospects of yet another great divergence – not necessarily of fundamental values but rather, regulatory approaches. Advocates of no regulation have long argued that American competitiveness relies on a legal infrastructure that imposes little or no regulatory burden on business. In fact, perceived cultural contrasts between the regulation-loving Brussels and innovative Silicon Valley have become something of a cultural trope. Is some kind of legal and cultural divergence really observed, or is it the stuff of myth?
Research suggests that U.S. companies have adopted various strategies to harmonize their business with EU regulation. These strategies include: advocating for an intercontinental "shield" to avoid direct regulation; making the business EU-compliant; lobbying to shape policy; and litigation when other strategies fail to deliver results.

A Shield. Five years after the EU's adoption of the *Data Protection Directive*, which allows companies operating in the EU to send personal data to third-party countries only if "adequate" levels of protection are guaranteed, the *International Safe Harbor Principles* were announced in 2000 to allow companies regulated by the Federal Trade Commission or the Department of Transportation to certify themselves as EU-compliant. The framework was updated in July 2016 as the EU-US Privacy Shield as a result of the Court of Justice of the European Union (CJEU) ruling that invalidated it. The updated Shield faced the same fate as the Safe Harbor, as the CJEU's *Schrems II* ruling on July 16, 2020, invalidated it on the grounds that it did not provide adequate protections in the case of government surveillance.

The failure of earlier efforts did not stop negotiations, however: in late March 2022, the two sides announced an agreement on the *Trans-Atlantic Data Privacy Framework*. In its current form, the *Framework* sounds like a statement of shared principles, rather than a set of concrete legal guidelines (European Data Protection Board, 2022). So far, mutual cooperation has been limited to voluntary participation, and in the area of data collection and storage only. There are calls for a Digital Alliance among liberal democracies, along the lines of what the Bretton

Woods system achieved in the post-WWII era (Wheeler, 2022), but that degree of cooperation across multiple issue areas is difficult to imagine now.

Compliance. Expressions of discontent and bravado in the face of regulation may have grabbed the headlines, but U.S. companies typically signal their compliance with existing regulatory standards, international ones included. Apple, for example, was quite vocal about its data and privacy policies, which came into effect a short time before the GDPR did (Evans, 2018). Facebook (currently Meta) announced one month after the *Schrems II* ruling that they would continue to operate through Standard Contractual Clauses (Meta, 2020), which has drawn criticism because such Clauses could violate European citizens' privacy in the case of government takeover of data (Hodge, 2020). A recent article published by McKinsey recommends that companies take stock of their AI tools and risk-mitigation measures, and make their businesses compliant with the EU's proposed AI Regulation – before the Regulation has even cleared the legislative process (Benjamin et al., 2021)! In conclusion, compliance is a fundamental – if sometimes understated – mechanism through which U.S. companies operate in the global marketplace.

***Lobbying*.** Research suggests an increase in U.S.-company lobbying activity in the European Union. Corporate Europe Observatory's 2021 report states: "20 per cent of the 599 companies lobbying digital policy issues in Brussels have their head office in the United States," which surpasses individual European countries as well as China (Bank et al., 2021). Eight of the top ten lobbying spenders in Europe are U.S.-based companies or their European satellites (Bank et al., 2021). It is worth noting that the EU's mixture of supranational and intergovernmental organizations, especially the multi-party, 705-strong European Parliament, makes successful lobbying by a small number of companies difficult. Nonetheless, this strategy will probably become more prominent as EU legislation picks up speed. It has been reported that Big Tech companies managed to bring down a provision in the proposed Digital Services Act that would have banned tracking-based advertising entirely.

Litigation. Even if not a tool of choice, American companies have been subject of litigation in the EU for a while. Amazon, Google, WhatsApp and Facebook have all received record-setting fines due to GDPR violations between 2019 and 2021. The French National Commission on Informatics and Liberty fined Google, Inc. for €50M in 2019. Thus, it can be predicted that where lawsuits cannot be avoided, companies will gear up legal strategies to ward off fines and other forms of punishment in Europe.

To sum up: U.S. companies' behavior in the face of EU regulation does not amount to a story of intercontinental divergence. Quite to the contrary, these companies have so far attempted to comply in their own terms (as the various "shields" had attempted to do), and when that failed, sought ways to shape the legislative process through lobbying, face litigation, or simply comply. For all the talk of the regulatory burden, U.S. businesses of various sizes have decided to stay in the European market rather than leave it in the wake of regulation.

Conclusion

On June 3, 2022, the *American Data Privacy and Protection Act* was introduced as a draft bill before the House of Representatives (American Data Privacy and Protection Act, 2022). This came soon after the creation of the National Artificial Intelligence Research Task Force in mid-2021 (Snyder, 2022), and three other legislative proposals, namely the *Digital Services Oversight and Safety Act*, the *Platform Accountability and Transparency Act* and the *Digital Platform Commission Act*, which were introduced in December 2021, February 2022 and May 2022, respectively. Thus, the future of AI ethics in the U.S. will be linked to legal regulation, whether legislated at home soil, abroad or both.

There is neither one single set of principles to define nor a silver bullet to address all the risks of AI systems, but shared understandings and convergent regulatory paths have been emerging. Accuracy, privacy, non-discrimination and fairness, transparency and explainability, trust, responsibility and sustainability will shape the future discourse of AI ethics, and the harmonization of legal regulation and business practice will shape its future practice. The European Union has taken the global lead in legislating principles around data rights, algorithmic accountability and online platforms, forcing businesses and legislators elsewhere to either comply with its rules or become more proactive in shaping the agenda.

References

American Data Privacy and Protection Act, H.R. 8152, 117th Cong. (2022).

Angwin, J., & Larson, J. (2016, July 29). *ProPublica responds to company's critique of machine bias story*. ProPublica. https://www.propublica.org/article/propublica-responds-to-companys-critique-of-machine-bias-story

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J., & Rahwan, L. (2018). The moral machine experiment. *Nature*, *563*(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Bank, M., Duffy, F., Leyendecker, V., & Silva, M. (2021). *The lobby network: Big Tech's web of influence in the EU*. Brussels, Belgium: Corporate Europe Observatory. https://corporateeurope.org/en/2021/08/lobby-network-big-techs-web-influence-eu

Baria, A. T., & Cross, K. (2021). *The brain is a computer is a brain: neuroscience's internal debate and the social significance of the computational metaphor*. arXiv. https://doi.org/10.48550/arXiv.2107.14042

Benjamin, M., Buehler, K., Dooley, R., & Zipparo, P. (2021, August 10). *What the draft European Union AI regulations mean for business*. McKinsey & Company. https://www.mckinsey.com/business-functions/quantumblack/our-insights/what-the-draft-european-union-ai-regulations-mean-for-business

Berendt, B. (2019). AI for the common good?! Pitfalls, challenges, and ethics pen-testing. *Paladyn, Journal of Behavioral Robotics*, *10*(1), 44–65. https://doi.org/10.1515/pjbr-2019-0004

Canca, C. (2020). Operationalizing AI ethics principles. *Communications of the ACM*, *63*(12), 18–21. https://doi.org/10.1145/3430368

Coeckelbergh, M. (2020). *AI Ethics*. MIT Press.

Crawford, K. (2021). *Atlas of AI*. Yale University Press.

Denno, D. W., & Surujnath, R. (2019). Foreword: Rise of the machines: Artificial Intelligence, robotics, and the reprogramming of law. *Fordham Law Review*, *88*(2), 381–404. https://ir.lawnet.fordham.edu/flr/vol88/iss2/1/

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Edelman, G. (2021, May 6). Everything you've heard about Section 230 is wrong. *Wired*. https://www.wired.com/story/section-230-internet-sacred-law-false-idol/

Elrod, H. J. W. (2020, May 28). Trial by Siri: AI comes to the courtroom. *Houston Law Review*, *57*, 19. https://houstonlawreview.org/article/12948-trial-by-siri-ai-comes-to-the-courtroom

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

European Data Protection Board. (2022, April 7). *EDPB adopts statement on the new Trans-Atlantic Data Privacy Framework, letter concerning independence of Belgian SA & discusses membership Spring Conference.* European Data Protection Board. https://edpb.europa.eu/news/news/2022/edpb-adopts-statement-new-trans-atlantic-data-privacy-framework-letter-concerning_en

Evans, J. (2018, April 13). Everything you need to know about Apple's GDPR privacy upgrade. *Computerworld*. https://www.computerworld.com/article/3268826/everything-you-need-to-know-about-apples-gdpr-privacy-upgrade.html

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Fried, I. (2021, October 8). *The long list of Facebook's insiders-turned-critics*. Axios. https://www.axios.com/2021/10/08/facebook-list-insiders-turned-critics

Friend, T. (2018, May 7). How frightened should we be of A.I.? *The New Yorker*. https://www.newyorker.com/magazine/2018/05/14/how-frightened-should-we-be-of-ai

Gordon, J.-S. (2020). Building moral robots: Ethical pitfalls and challenges. *Science and Engineering Ethics*, *26*(1), 141–157. https://doi.org/10.1007/s11948-019-00084-5

Greene, D., Hoffmann, A. L., & Stark, L. (2019). *Better, nicer, clearer, fairer: A critical assessment of the movement for ethical Artificial Intelligence and machine learning.* Hawaii International Conference on System Sciences. https://scholarspace.manoa.hawaii.edu/items/2c91a03d-caad-4c6f-9069-9bca4f393ce6 Gunkel, D. J. (2018). *Robot Rights*. MIT Press.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

Hanna, R., & Kazim, E. (2021). Philosophical foundations for digital ethics and AI ethics: A dignitarian approach. *AI and Ethics*, *1*, 405–423. https://doi.org/10.1007/s43681-021-00040-9

Heller, N. (2016, November 20). If animals have rights, should robots? *The New Yorker*. https://www.newyorker.com/magazine/2016/11/28/if-animals-have-rights-should-robots

Hickok, M. (2021). Lessons learned from AI ethics principles for future actions. *AI and Ethics*, *1*(1), 41–47. https://doi.org/10.1007/s43681-020-00008-1

Hodge, N. (2020, August 19). EU privacy advocate targets Facebook, Google in latest salvo. *Compliance Week.* https://www.complianceweek.com/data-privacy/eu-privacy-advocate-targets-facebook-google-in-latest-salvo/29335.article

Tucker, I. (2017, May 28). 'A white mask worked better': Why algorithms are not colour blind. *The Guardian*. https://www.theguardian.com/technology/2017/may/28/joy-buolamwini-when-algorithms-are-racist-facial-recognition-bias

Katyal, S. K. (2019). Private accountability in the age of Artificial Intelligence. *UCLA Law Review*, *54*, 89. https://www.uclalawreview.org/private-accountability-age-algorithm/

LaFrance, A. (2021, October 28). History will not judge US kindly. *The Atlantic*. https://www.theatlantic.com/ideas/archive/2021/10/facebook-papers-democracy-election-zuckerberg/620478/

Larsson, S. (2020). On the governance of Artificial Intelligence through ethics guidelines. *Asian Journal of Law and Society*, *7*(3), 437–451. https://doi.org/10.1017/als.2020.19

Latonero, M. (2018, May 11). Artificial Intelligence & Human Rights: A workshop at Data & Society. *Data & Society: Points*. https://points.datasociety.net/artificial-intelligence-human-rights-a-workshop-at-data-society-fd6358d72149

Lauer, D. (2021). You cannot have AI ethics without ethics. *AI and Ethics*, *1*(1), 21–25. https://doi.org/10.1007/s43681-020-00013-4

Lazzaro, S. (2021, September 30). Are AI ethics teams doomed to be a facade? Women who pioneered them weigh in. *Venture Beat*. https://venturebeat.com/2021/09/30/are-ai-ethics-teams-doomed-to-be-a-facade-the-women-who-pioneered-them-weigh-in/

Lyons, H., Velloso, E., & Miller, T. (2021). Conceptualising contestability: Perspectives on contesting algorithmic decisions. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–25. https://doi.org/10.1145/3449180

Mac, R., & Hill, K. (2022, May 9). Clearview AI settles suit and agrees to limit sales of facial recognition database. *The New York Times*. https://www.nytimes.com/2022/05/09/technology/clearview-ai-suit.html

Markoff, J. (2016). *Machines of loving grace: The quest for common ground between humans and robots*. HarperCollins Publishers.

Meta. (2020, August 17). *Updating our international data transfer mechanisms*. Meta. https://www.facebook.com/business/news/updating-our-international-data-transfer-mechanisms/

Metcalf, J., Moss, E., Boyd, D. (2019). Owning ethics: Corporate logics, Silicon Valley, and the institutionalization of ethics. *Social Research: An International Quarterly*, *86*(2), 449–476. https://muse.jhu.edu/article/732185

Milmo, D. (2021, December 6). Rohingya sue Facebook for £150bn over Myanmar genocide. *The Guardian*. https://www.theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence

Molnar, P. (2019). Technology on the margins: AI and global migration management from a human rights perspective. *Cambridge International Law Journal*, *8*(2), 305–330. https://doi.org/10.4337/cilj.2019.02.07

Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2021). Operationalising AI ethics: Barriers, enablers and next steps. *AI & Society*. https://doi.org/10.1007/s00146-021-01308-8

Nersessian, D., & Mancha, R. (2020). From automation to autonomy: Legal and ethical responsibility gaps in Artificial Intelligence innovation. *Michigan Technology Law Review*, *27, 43*. https://doi.org/10.36645/mtlr.27.1.from

Noble, S. U. (2018). *Algorithms of Oppression: How search engines reinforce racism*. NYU Press.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books.

Renieris, E. M., Kiron, D., & Mills, S. (2022, May 24). Should organizations link responsible AI and Corporate Social Responsibility? It's complicated. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/should-organizations-link-responsible-ai-and-corporate-social-responsibility-its-complicated/

Risse, M. (2019). Human rights and Artificial Intelligence: An urgently needed agenda. *Human Rights Quarterly*, *41*(1), 1–16. https://doi.org/10.1353/hrq.2019.0000

Rudschies, C., Schneider, I., & Simon, J. (2020). Value pluralism in the AI ethics debate – Different actors, different priorities. *The International Review of Information Ethics*, *29*. http://informationethics.ca/index.php/irie/article/view/419

Ryan, M. (2020). In AI we trust: Ethics, Artificial Intelligence, and reliability. *Science and Engineering Ethics*, *26*(5), 2749–2767. https://doi.org/10.1007/s11948-020-00228-y

Segun, S. T. (2021). From machine ethics to computational ethics. *AI & Society*, *36*(1), 263–276. https://doi.org/10.1007/s00146-020-01010-1

Selbst, A. D. (2020). Negligence and AI's human users. *Boston University Law Review*, *100*, 1315–1376. https://www.bu.edu/bulawreview/files/2020/09/SELBST.pdf

Stahl, B. C., Andreou, A., Brey, P., Hatzakis, T., Kirichenko, A., Macnish, K., Laulhe Shaelou, S., Patel, A., Ryan, M., & Wright, D. (2021). Artificial Intelligence for human flourishing –

Beyond principles for machine learning. *Journal of Business Research*, *124*, 374–388. https://doi.org/10.1016/j.jbusres.2020.11.030

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J. Tambe, M., & Teller, A. (2016). *Artificial Intelligence and life in 2030.* The one hundred year study on Artificial Intelligence: Report of the 2015 Study Panel, Stanford University, Stanford, CA. https://ai100.stanford.edu/2016-report

Snyder, B. (2022, April 17). *Here's how the U.S. may regulate artificial intelligence*. ScreenRant. https://screenrant.com/us-regulate-artificial-intelligence-how-advisory-committee/

State v. Loomis, 881 N.W.2d (Wis. 2016). https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690

van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, *1*(3), 213–218. https://doi.org/10.1007/s43681-021-00043-6

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Nerini, F. F. (2020). The role of Artificial Intelligence in achieving the sustainable development goals. *Nature Communications*, *11*(1), 233. https://doi.org/10.1038/s41467-019-14108-y

Wallach, W., & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. New York: Oxford University Press.

Walz, A., & Firth-Butterfield, K. (2018). Implementing ethics into Artificial Intelligence: A contribution, from a legal perspective, to the development of an AI governance regime. *Duke Law & Technology Review*, *17*, 180–231. https://scholarship.law.duke.edu/dltr/vol18/iss1/17/

Wheeler, T. (2022, March 29). U.S. regulatory inaction opened the doors for the EU to step up on Internet. *Brookings*. https://www.brookings.edu/blog/techtank/2022/03/29/u-s-regulatory-inaction-opened-the-doors-for-the-eu-to-step-up-on-internet/