Social worlds analysis of AI ethics

Alayna Anne Kennedy

School of Social and Political Science, University of Edinburgh, UK

Abstract

In this paper, I briefly summarize the emerging field of artificial intelligence (AI) ethics, focusing on work done in industry, academia, and government. I show that the majority of AI ethics work has been ineffective in impacting real world AI deployment and in mitigating harms of the technology. In order to understand why AI ethics has been so ineffective, I analyze each of these three actors with a social worlds approach, showing that the outsized power of the social world of industry has rendered government and academic actors ineffectual. By delving into issues of power and boundary setting between these three actors, I show that the social world of government has been employed by industry to do the boundary work of standards setting, while academics have been co-opted by industry to legitimate the industry paradigm of corporate social responsibility. In this way, both government and academic social worlds contribute to supporting the social world of industry and perpetuate its harms.

1. Background on AI ethics

1.1 AI technology has been developed within an industry setting

Artificial Intelligence (AI) technology has exploded in popularity over the last 10 years, with each wave of technical breakthroughs ushering in speculation about the potential impacts of AI on our society, businesses, and governments. First, the Big Data revolution promised to forever change the way we understood analytics, then Deep Learning promised human-level AI performance (Chen, 2018), and today AI seems to offer huge business returns to investors (Kennedy, 2020). AI has long been a buzzword in businesses across the world, but for many government agencies and larger organizations, earlier applications of commercial AI proved to be overhyped and underwhelming (Floridi, 2020). Only now are large-scale organizations, including governments, beginning to implement AI technology at scale, as the technology has moved from the research lab to the office (Kennedy, 2020).

In industry, research groups like OpenAI continually reach research goals previously considered unattainable, like with their enormous language model GPT-3 generating human-like text (Floridi, 2020). Deployment of AI in industry continues to scale wider and deeper, with most major tech companies releasing large models to perform automated tasks, like IBM's enormous Watson AI platform (High, 2012). Organizations are using the technology to optimize efficiency and profitability, which may be seen in metrics for operations, financial reporting, human resources, customer service, and a growing list of other aspects of running a business (Joyce, 2021).

1.2 AI ethics in government relies heavily on industry

In addition to being used within industry, AI systems are increasingly being used to make decisions within large scale government projects, including the deployment of humanitarian resources (Mitchell, 2019), who is granted bail (Chouldechova, 2017), and which citizens are subjected to increased police presence (Cobey, 2019. In the United States, every federal agency has been tasked with developing a plan for incorporating AI into their operations (OSTP, 2019). Through an executive order, an AI summit, and the creation of a website and a White House Select Committee on AI, the Office of Management and Budget and the Office of Science and Technology Policy are leading a governmentwide effort to maximize AI's benefit (Keegan, 2019).

Like the US, many other European, Australasian, and Middle Eastern countries seek to develop their AI capabilities for the public sector. All of these governments depend on strong relationships with the technology sector and public-private partnerships with big tech corporations - in other words, "public" AI is just private company's technology used for traditionally government-owned purposes (Filer, 2018).

1.2 AI ethics critiques within the context of academia

Although AI has been eagerly adopted by industry and government, it has its detractors. Each of the waves of AI development has been accompanied by a suite of ethical concerns and mitigation strategies (Kennedy, 2020).

Importantly, much AI ethics work originally came from academic criticism of industry, like the seminal gender shades paper from MIT researcher Joy Buolamwini (2018) that proved facial recognition technology is biased against people of color, and women of color specifically. Therefore, as AI has become a buzzword in every private company and public agency, academics and civil society researchers have begun to talk about the need for "AI ethics" to mitigate these harms (Ebell, 2021; Hagerty, 2019).

1.2.1 Industry has co-opted many of the academic critiques of AI ethics

While the field of AI ethics was borne from academics like Buolamwini who were concerned about racial and gender disparity, the idea of responsible AI quickly spread to industry and government. Between 2016 and 2019, 74 sets of ethical principles or guidelines for AI were published, focusing on high-level guidance like "creating transparent AI" (Hilligoss and Fjeld, 2021). In addition, a great deal of the research efforts in the field of AI ethics have focused on mitigating bias and unfair machine learning (ML) in the technical stages of development, such as detecting historically biased data and removing disparate impact from a model's output (Holstein, 2019; Mehrabi, 2021; Mitchell, 2019).

Although the field of AI ethics was borne out of academia's criticism about industry use of AI, industry has easily co-opted the language of AI ethics, to the point where the majority of AI ethics guidelines are published by private actors (Hagendorff, 2020). Most major tech companies have a group that seems to support AI ethics, like IBM's AI Ethics board and Microsoft's Fair, Accountability, Transparency, and Ethics (FATE) group. Even Facebook, notorious for ethical harms like using user's data without their consent (Lauer, 2021) and causing racially based violence in Myanmar (Stevenson, 2018) has a group focused on AI ethics and responsible technology (Hao, 2021).

However, critics have noted that Facebook's ethical failures are actually part of its business model – it is more profitable for them to sow division and spread misinformation because it leads to higher engagement and ad revenue (Lauer, 2021). Critical reporting on Facebook's responsible AI team shows that even their dedicated ethics teams are more focused on bringing the company good press and deflecting their harmful actions than on creating systemic change of Facebook's core business model (Hao, 2021). Arguably, these criticisms of Facebook hold true for other industry actors. The initial academic critiques of AI ethics cannot be incorporated into current industry business models, and current industry AI ethics is merely a way for companies to receive better press or higher profits (Morley, 2021; Wagner, 2018).

Therefore, a crucial question arises: do AI ethics guidelines have an actual impact in the field of AI and ML? After reviewing 22 major AI ethics guidelines, Hagendorff (2019) concludes: "No, most often not" (Hagendorff, 2019: 1). In order to understand why academic's critiques of AI ethics do not cause real changes in either industry or government, we must incorporate an understanding of power dynamics into our analysis of each of these three social worlds.

2. Social worlds analysis of AI ethics

The social worlds framework has been a useful tool in the history of science and technology studies (STS) in understanding controversies in new fields of technology; therefore, it feels like an applicable framework to examine the main actors in the new field of AI ethics (Star, 1990). "Social worlds" have been defined within the sociological literature as "amorphous and diffuse constellations of actors, organizations, events, and practices which have coalesced into spheres of interest and involvement" (Unruh 1980: 277).

While the conception of social worlds originally referred to cultural groups like opera, rock climbers, and jugglers, the idea has been extended to conceptualize the ways that different industries group together into their own social worlds, like the "computer world" (Kling and Gerson, 1978). Technologists within these industry-based social worlds all share conventions, language, and standards. Crucially, it is difficult and sometimes expensive for someone within a world to practice outside these sets of standards and norms (Star, 1990). Finally, a key component of social worlds is that the worlds can be broken down into subgroups (MacLean, 2020).

I argue that AI ethics represents an emerging social world, and that industry, academia, and government represent its three main subgroups. First, AI research represents a broad social world, a group of people consolidating around a

common interest in the emerging technology of Artificial Intelligence.

Although there has been a long-standing cultural fascination with creating intelligent machines, no cohesive definition of AI has yet emerged (Joyce, 2021). Since the technology was developed in the 1950s, it has been used in a wide array of applications, encompassing task automation, chess playing, algorithmic risk scoring, and robotics (Mittelstadt et al. 2016). AI developers also use an astonishingly broad array of computational techniques to develop what they all call "AI," including cybernetics, logic and rule-based systems, statistical and probabilistic AI, pattern recognition, adaptive behavioral models, and symbolic knowledge representation such as expert systems (Nilsson, 2009). Despite its broad and fluid definition, or perhaps because of it, enthusiasts of AI can make future-oriented, technologically utopian claims about its potential impact. This future-focused language and utopian vision of AI bring together many of its followers in a loose confederation, a social world.

However, the broad social world of AI is also easily divided into different subgroups through the process of drawing definitional boundaries and interpreting the context of AI (Forsythe, 2001). The very process of defining different subgroups of AI and AI ethics is an interpretive and political process (Star, 1990). Therefore, different subgroups emerge based on different interpretations of the context of AI, and the three different subgroups I examine in this paper – industry, government, and academia – all use different language, definitions, and standards when talking about AI and AI ethics (Joyce, 2021).

2.1 Social world of industry framed by corporate social responsibility

All social worlds are formed by a number of sociological phenomena – communication between actors, implicit behavior patterns, common activities, and shared language (Unruh, 1980). To examine the social world of AI ethics within industry, let us look at the shared language and culture of corporate social responsibility (CSR) that companies use when discussing and promoting AI ethics initiatives. Like AI itself, CSR eludes a single definition and encompasses many organizational strategies for improving profit by aligning with social values. A simple definition of CSR was provided by the European Commission in 2001:

> "A concept whereby companies integrate social and environmental concerns into their business operations and in their interactions with their stakeholders on a voluntary basis" (Dahlsrud, 2008).

A crucial element of this approach is that it is done "on a voluntary basis" (Dahlsrud, 2008). The voluntary nature of CSR initiatives undertaken by private companies contrasts with formal regulatory mechanisms like laws and standards that government have historically used to govern business and emerging technology (Blowfield and Frynas, 2005). Part of this regulatory shift is due to the shift in power over technological development away from government actors and toward corporations (Lim, 2011). For example, in America in the Cold War era, the development of technologies like nuclear missiles and ballistic defense systems was managed by the US government and department of defense. However, with the rise of technological corporations like Apple, Microsoft, and the more recent additions of Facebook and Google, the management of emerging technologies like AI is dominated by private actors (Whittaker, 2021).

The language of CSR appropriates the culture of benevolent governance, mimicking the idea of "serving citizens" and "working toward social good" that permeates the social world of government (Keegan, 2019). However, CSR is, most importantly, *profitable* for the companies that deploy it (McElhaney, 2009). Proven financial benefits of CSR can be found in the areas of human resources, reputation and branding, and operational cost savings, and now, arguably, in AI ethics (Blowfield, 2005; McElhaney, 2009).

We can see how companies approach AI ethics with the cultural and linguistic frame of CSR. Like Facebook, companies use their AI ethics teams as a way to deflect criticism and improve their shareholdings (Hao, 2021). Furthermore, companies are able to frame AI ethics as a regulatory framework in the absence of government laws and standards on AI. This legal conception of ethics also contributes to companies profits, since by setting the legal agenda companies avoid expensive lawsuits over AI misuse (Resseguier and Rodrigues, 2020). Companies are only able to frame AI ethics as a replacement for law because the social world of government has failed to set regulatory standards.

2.2 Social world of government does the invisible work of standards setting

While the social world of industry AI ethics can be characterized by CSR, the social world of government AI ethics can be distinguished by its shared discourse with industry. In the sociological study of social worlds, "shared discourses" were ways for social worlds and create "social wholes" – the interactions of collective actors and discourses (Clarke, 2008). Social wholes were a way to generate a shared perspective within a social world as the basis for collective action (Shibutani, 1955).

With AI implementation surging within private enterprise, public agencies have started to incorporate AI into their operations to increase efficiency and better serve citizens (Keegan, 2019). Most of these agencies propose a blended effort to tackle AI development, combining public and private capabilities to redress complex societal matters like public health and housing (Joyce, 2021; OSTP, 2019). Crucially, because governments do not have the technological capabilities or access to data that private actors have, they need to cooperate with industry actors to develop AI (Whittaker, 2021). Therefore, governments have tended to form a social whole with industry, often accepting their perspectives into the creation of AI standards. For example, the US National Institute for Standards and Technology (NIST) actively collaborates with tech companies in creating their standards for AI technology, and the National Security Commission on Artificial Intelligence (NSCAI) is led by a former CEO of Google (Dowd, 2021).

NIST represents one of the most important roles of the government sub-group within the larger social world of AI – the invisible work of standards setting. Within social worlds, classifications and standards serve as one of the main ways that boundaries are created. Specifically, standards serve as "sites for mediation between technical requirements…and political requirements" (Bowker, 1998: 232). Standards not only serve to shape the future development of technologies, but they are also implemented through a practical political process. In the case of AI ethics, industry opinions are welcomed, even celebrated, in setting standards. Therefore, in creating the important boundary objects of AI standards, regulations, and government agendas, government actors include and often defer to industry actors (Filer, 2018).

2.2 Social world of academia provides industry legitimization through critique

While government actors in AI ethics defer to industry actors in creating boundary objects, the social world of academia is characterized by its criticism of industry actors in the AI ethics space. A number of sociologists studying social worlds have emphasized a key component of the framework – cooperation within social worlds can proceed without consensus, that "individuals and collectives can 'set their differences aside'" (Clarke, 2008: 119) in order to promote the shared goals of the social world (Kling and Gerson, 1978).

For many academics within the AI ethics space, this "cooperation without consensus" has involved working with, and therefore legitimizing, AI ethics work within the CSR framework of industry (Clarke, 2008; MacLearn, 2020). Legitimation represents a core issue in social science, entering discussions about class, power, social control, and society itself (Berger and Luckman, 1966). As social worlds divide

into subworlds, this issue of legitimation becomes even more significant.

Subworlds legitimize themselves in several different ways, particularly through 1) distancing themselves from their other subworlds and 2) theorizing about the nature of their subworld and parent social world (Strauss, 1982). Within the realm of AI ethics, the subworld of academia serves to legitimate both itself and its parent subworld by providing criticism that distances it from its industry and government counterparts, and by publishing theories about the nature of AI ethics (Whittaker, 2021; Wagner, 2018). However, many academics' criticism of industry still fits within the CSR paradigm of AI ethics and does not fundamentally challenge the power industry actors have over both the subworld of government and the subworld of academia.

Academia's obedience to the subworld of industry can be seen through two phenomena: the focus on fairness and bias mitigation research, and industry's control of funding and data. First, efforts in the field of AI Ethics have overwhelmingly focused on mitigating bias and unfair ML in the technical stages of development, such as detecting historically biased data and removing disparate impact from a model's output (Holstein, 2019; Mitchell, 2019). This narrow lens on fairness lends itself to purely technical solutions and fails to challenge companies' core busines models. With the help of their academic counterparts, companies have developed technical "fixes" to algorithmic fairness that ultimately serve to add value to the company without challenging it (Ebell, 2021; Hagerty, 2019; Morley, 2021). Secondly, industry controls computing resources, data, and funding for AI development. Given the precipitous decline of academic funding from governments over the last few decades, many researchers interested in AI have no choice but to partner with industry actors to access the resources of data, compute power, and monetary funding (Whittaker, 2021). As former Google employee and AI critic Meredith Whittaker writes:

> "From industry-sponsored Ph.D. programs to initiatives that place tech-company offices literally in the middle of universities to the National Science Foundation partnering with Amazon to define the parameters of "fairness" in AI and awarding grants to those who meet their positivist criteria, we see myriad schemes to draw academia closer to tech companie". (2021).

3. Issues with the social worlds theory as a lens to examine AI ethics

After examining each of the main social sub worlds in the AI ethics field, we can begin to draw conclusions about why

the majority of current AI ethics work fails to have an impact in real-world AI applications. Government and academic sub-worlds use language, paradigms, and boundaries set by industry actors, who hold the most capital, data, and compute power (Hagerty, 2019). Embedded within AI socio-technical systems are deep inequalities between industry, academia, and government, as well as connections to profit and capitalism (Noble, 2019; Noble and Roberts; 2019).

3.1 Power dynamics within social worlds of AI ethics

The social worlds theory of AI ethics illuminates the disproportionate amount of power that industry actors hold in the space. For these social worlds to exist in tandem, they need to be balanced, and the very act of balancing them together is one of power relations and politics. As social world theorist Star writes:

"Power is about who brings worlds together and holds them there. It may be a power of the zero-point or a power of discipline; of enrolment or affinity; it may ... heal or create, erase or violate, impose a voice or embody more than one voice" (1990).

Industry AI ethics actors bring together both government and academic actors in a shared language of CSR and a focus on bias mitigation, which allows them to ignore core inequalities in the business model of AI technology.

4. Conclusion

We have seen that under the parent social world of AI ethics, the subworld of industry possesses the most power and influences the other two main subworlds of academia and government actors. Government standards are set with the help of industry professionals and a great deal of academic criticism of industry still fits within a profit-seeking CSR paradigm.

The social worlds framework is a helpful theoretical tool for identifying the power disparities in AI ethics, but offers fewer solutions for how to solve the power disparity issue. AI ethics researchers have called for more ethnographic research to understand the social impacts of AI (Hagerty, 2019). However, ultimately academics and government actors cannot effectively implement responsible and beneficial AI regulation while remaining in social worlds dominated by industry actors.

Recent developments suggest that academic critics of industry AI may be breaking out of the current social worlds equilibrium to form their own research community independent of, and opposed to, industry AI. Critical AI research has skyrocketed in the past year, and former Google employee Timnit Gebru just launched an independent research center "free from Big Tech's pervasive influence" (DAIR, 2021). As Gebru writes,

"In order to truly have checks and balances, we should not have the same people setting the agendas of big tech, research, government and the non-profit sector. We need alternatives" (2021).

Only by breaking out of the current social worlds dynamic, where academia is subservient to industry, can technology activists hope to create a truly beneficial AI ethics agenda.

References

Blowfield, M. and Frynas, J.G., 2005. Editorial Setting new agendas: critical perspectives on Corporate Social Responsibility in the developing world. *International affairs*, *81*(3), pp.499-513.

Berger Peter, L. and Luckmann, T., 1966. The social construction of reality. *A Treatise in the Sociology of Knowledge*.

Bowker, G.C. and Star, S.L., 1998. Building information infrastructures for social worlds—The role of classifications and standards. In *Community computing and support systems* (pp. 231-248). Springer, Berlin, Heidelberg.

Buolamwini, J. and Gebru, T., 2018, January. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.

Chen, A., 2018. *A pioneering scientist explains 'deep learning'*. [online] The Verge. Available at: <https://www.theverge.com/2018/10/16/17985168/deep-learning-revolution-terrence-sejnowski-artificial-intelligence-technology> [Accessed 8 December 2021].

Chouldechova, A., 2017. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, *5*(2), pp.153-163.

Clarke, A.E. and Star, S.L., 2008. The social worlds framework: A theory/methods package. *The handbook of science and technology studies*, *3*(0), pp.113-137.

Cobey, C., 2021. *Forbes Insights: AI Regulation: It's Time For Training Wheels*. [online] Forbes. Available at: <https://www.forbes.com/sites/insights-intelai/2019/03/27/ai-regulation-its-time-for-training-wheels/> [Accessed 8 December 2021].

Dahlsrud, A., 2008. How corporate social responsibility is defined: an analysis of 37 definitions. *Corporate social responsibility and environmental management*, *15*(1), pp.1

Dair-institute.org. 2021. THE DAIR Institute. [online] Available at: <https://www.dair-institute.org/> [Accessed 9 December 2021].

Dowd, M., 2021. *Opinion | A.I. Is Not A-OK*. [online] Nytimes.com. Available at: <https://www.nytimes.com/2021/10/30/opinion/eric-schmidt-ai.html> [Accessed 8 December 2021].

Ebell, C., Baeza-Yates, R., Benjamins, R., Cai, H., Coeckelbergh, M., Duarte, T., Hickok, M., Jacquet, A., Kim, A., Krijger, J. and MacIntyre, J., 2021. Towards intellectual freedom in an AI Ethics Global Community. *AI and Ethics*, *1*(2), pp.131-138.

Filer, T., 2018. *Developing AI for Government: What role and limits for the private sector?*. [online] Bennett Institute for Public Policy. Available at: <https://www.bennettinstitute.cam.ac.uk/blog/developing-ai-government-what-role-and-limits-priv/> [Accessed 8 December 2021].

Floridi, L. and Chiriatti, M., 2020. GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*(4), pp.681-694.

Forsythe, D., 2001. Studying those who study us: An anthropologist in the world of artificial intelligence. Stanford University Press.

Gebru, T., 2021. *For truly ethical AI, its research must be independent from big tech*. [online] The Guardian. Available at: <https://www.theguardian.com/commentisfree/2021/dec/06/google-silicon-valley-ai-timnit-gebruhttps://www.theguardian.com/commentisfree/dec/06/google-silicon-valley-ai-timnit-gebru> [Accessed 9 December 2021].

Hagendorff, T., 2020. The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), pp.99-120.

Hagerty, A. and Rubinov, I., 2019. Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence. *arXiv preprint arXiv:1907.07892*.

Hao, K., 2021. *How Facebook got addicted to spreading misinformation*. [online] MIT Technology Review. Available at: <https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/> [Accessed 8 December 2021].

High, R., 2012. The era of cognitive systems: An inside look at IBM Watson and how it works. *IBM Corporation, Redbooks*, *1*, p.16.

Hilligoss, H. and Fjeld, J., 2021. *Introducing the Principled Artificial Intelligence Project*. [online] Berkman Klein Center. Available at: <https://cyber.harvard.edu/story/2019-06/introducing-principled-artificial-intelligence-project> [Accessed 8 December 2021].

Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M. and Wallach, H., 2019, May. Improving fairness in machine learning systems: What do industry practitioners need?. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-16).

Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S.G., Noble, S.U. and Shestakofsky, B., 2021. Toward a Sociology of Artificial Intelligence: A Call for Research on Inequalities and Structural Change. *Socius*, *7*, p.237-281.

Keegan, M., 2019. *More Than Meets AI*. [online] Businessofgovernment.org. Available at: <https://www.businessofgovernment.org/sites/default/files/More%20Than%20Meets%20AI-Assessing%20the%20Impact%20of%20Artificial%20Intelligence%20on%20the%20Work%20of%20Government.pdf> [Accessed 8 December 2021].

Kennedy, A., 2020. *Why We Need to Audit Government AI | Montreal AI Ethics Institute*. [online] Montreal AI Ethics Institute. Available at: <https://montrealethics.ai/why-we-need-to-audit-government-ai/> [Accessed 8 December 2021].

Kling, R. and Gerson, E.M., 1978. Patterns of segmentation and intersection in the computing world. *Symbolic Interaction*, *1*(2), pp.24-43.

Lauer, D., 2021. Facebook's ethical failures are not accidental; they are part of the business model. *AI and Ethics*, pp.1-9.

Lauer, D., 2021. You cannot have AI ethics without ethics. *AI and Ethics*, *1*(1), pp.21-25.

Lim, W.K., 2011. Understanding risk governance: Introducing sociological neoinstitutionalism and foucauldian governmentality for further theorizing. *International Journal of Disaster Risk Science*, *2*(3), pp.11-20.

MacLean, S., Dwyer, R., Pennay, A., Savic, M., Wilkinson, C., Roberts, S., Turner, K., Saleeba, E. and Room, R., 2020. The 'social worlds' concept: a useful tool for public health-oriented studies of drinking cultures. *Addiction Research & Theory*, *29*(3), pp.231-238.

McElhaney, K., 2009. A strategic approach to corporate social responsibility. *Leader to leader*, *52*(1), pp.30-36.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A., 2021. A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), pp.1-35.

Mitchell, B., 2019. *Pentagon's jaic needs industry help for humanitarian assistance, disaster relief*. [online] https://www.fedscoop.com/pentagons-jaic-needs-industry-help-humanitarian-assistance-disaster-relief/. Available at: <https://www.fedscoop.com/pentagons-jaic-needs-industry-help-humanitarian-assistance-disaster-relief/> [Accessed 8 December 2021].

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D. and Gebru, T., 2019, January. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 220-229).

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L., 2016. The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), p.2053951716679679.

Morley, J., Floridi, L., Kinsey, L. and Elhalal, A., 2021. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles

into practices. In *Ethics, Governance, and Policies in Artificial Intelligence* (pp. 153-183). Springer, Cham.

Nilsson, N.J., 2009. *The quest for artificial intelligence*. Cambridge University Press.

Noble, S.U., 2018. *Algorithms of oppression*. New York University Press.

Noble, S.U. and Roberts, S.T., 2019. 6 Technological Elites, the Meritocracy, and Postracial Myths in Silicon Valley (pp. 113-130). Duke University Press.

OSTP (Office of Science and Technology Policy) . 2019. "Summary of AI Executive Order, February 11, 2019." White House Fact Sheets. Retrieved December 8, 2021. https://www.whitehouse.gov/ai/resources/white-house-fact-sheets/.

Resseguier, A. and Rodrigues, R., 2020. AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society*, *7*(2), p.205-41.

Star, S.L., 1990. Power, technology and the phenomenology of conventions: on being allergic to onions. *The Sociological Review*, *38*(1_suppl), pp.26-56.

Stevenson, A., 2018. Facebook admits it was used to incite violence in Myanmar. *The New York Times*, *6*.

Strauss, A., 1982. Social worlds and legitimation processes. *Studies in symbolic interaction*, *4*(17), pp.171-190.

Shibutani, T., 1955. Reference groups as perspectives. *American journal of Sociology*, *60*(6), pp.562-569.

Unruh, D.R., 1980. The nature of social worlds. *Pacific sociological review*, *23*(3), pp.271-296.

Wagner, B., 2018. Ethics as an escape from regulation. From "ethics-washing" to ethics-shopping?. In *Being Profiled* (pp. 84-89). Amsterdam University Press.

Whittaker, M., 2021. *The steep cost of capture | ACM Interactions*. [online] Interactions.acm.org. Available at: <https://interactions.acm.org/archive/view/november-december-2021/the-steep-cost-of-capture> [Accessed 8 December 2021].