

The Contestation of Tech Ethics: A Sociotechnical Approach to Ethics and Technology in Action

Ben Green
bzgreen@umich.edu
Michigan Society of Fellows
Gerald R. Ford School of Public Policy

Abstract

Recent controversies related to topics such as fake news, privacy, and algorithmic bias have prompted increased public scrutiny of digital technologies and soul-searching among many of the people associated with their development. In response, the tech industry, academia, civil society, and governments have rapidly increased their attention to “ethics” in the design and use of digital technologies (“tech ethics”). Yet almost as quickly as ethics discourse has proliferated across the world of digital technologies, the limitations of these approaches have also become apparent: tech ethics is vague and toothless, is subsumed into corporate logics and incentives, and has a myopic focus on individual engineers and technology design rather than on the structures and cultures of technology production. As a result of these limitations, many have grown skeptical of tech ethics and its proponents, charging them with “ethics-washing”: promoting ethics research and discourse to defuse criticism and government regulation without committing to ethical behavior. By looking at how ethics has been taken up in both science and business in superficial and depoliticizing ways, I recast tech ethics as a terrain of contestation where the central fault line is not *whether* it is desirable to be ethical, but *what* “ethics” entails and *who* gets to define it. This framing highlights the significant limits of current approaches to tech ethics and the importance of studying the formulation and real-world effects of tech ethics. In order to identify and develop more rigorous strategies for reforming digital technologies and the social relations that they mediate, I describe a sociotechnical approach to tech ethics, one that reflexively applies many of tech ethics’ own lessons regarding digital technologies to tech ethics itself.

Table of Contents

1	Introduction: The Crisis of Conscience	3
2	The Rise of Tech Ethics	6
2.1	Tech Industry	7
2.2	Academia	7
2.3	Civil Society.....	8
2.4	Government.....	9
3	The Limits of Tech Ethics.....	10
3.1	Tech ethics principles are abstract and toothless	10
3.2	Tech ethics is subsumed into corporate logics and incentives.....	11
3.3	Tech ethics has a myopic focus on individual engineers and technology design	14
3.4	Tech ethics has become an avenue for ethics-washing.....	16
4	The Contestation of Tech Ethics	18
4.1	Ethics in Science	20
4.2	Corporate Ethics and Co-optation.....	21
4.3	The Future of Tech Ethics.....	23
5	A Sociotechnical Approach to Tech Ethics	24
5.1	Determinism.....	25
5.2	Solutionism	25
5.3	Objectivity and Neutrality.....	26
5.4	Sociotechnical Systems.....	27

1 Introduction: The Crisis of Conscience

If digital technology production in the beginning of the 2010s was characterized by the brash spirit of Facebook's motto "move fast and break things" and the superficial assurances of Google's motto "don't be evil," digital technology toward the end of the decade was characterized by a "crisis of conscience" (Marantz, 2019) about these and other technologies' perils. While many of digital technology's harms were visible early in the decade (not to mention long before), it was not until stories of this technology causing harm reached a critical mass that they became salient in the public eye. Consider just a few of the cases that have prompted this crisis of conscience within tech and the associated "teclash"—the growing animosity of the public toward major technology companies—which in 2018 was deemed by both Oxford Dictionaries (Oxford Languages, 2018) and the Financial Times (Feroz, 2018) to be one of the words of the year.

Fake News: Throughout the 2016 U.S. presidential election between Donald Trump and Hillary Clinton, social media was plagued with fraudulent stories that went viral: "FBI confirms evidence of huge underground Clinton sex network," "Hillary sold weapons to ISIS," and more (Emery Jr., 2016; Ritchie, 2016). Social media platforms—Facebook in particular—came under scrutiny for their milquetoast response to fake news, allowing these stories to spread widely without providing any indication that they were not true. Given that fake news spreads further, faster, and more broadly than true news (Vosoughi, Roy, and Aral, 2018), social media business models of prioritizing user engagement appear to be partially responsible for the spread of fake news. In turn, numerous commentators—including Hillary Clinton herself—blamed Facebook for Donald Trump's presidential election victory (Blake, 2018; Graham, 2017; Read, 2016; Solon, 2016). Later reporting revealed that Facebook's leadership actively resisted sharing more information about Russian efforts to propagate fake news, instead prioritizing the company's business strategies (Perlroth, Frenkel, and Shane, 2018).

Cambridge Analytica: In 2018, *The New York Times* and *The Guardian* reported that the voter-profiling firm Cambridge Analytica had harvested information from the Facebook profiles of more than 50 million (later revealed to be 87 million) people, without their knowledge or permission, in order to target political ads to benefit Donald Trump's 2016 presidential campaign (Cadwalladr and Graham-Harrison, 2018; Lapowsky, 2018; Rosenberg, Confessore, and Cadwalladr, 2018).

Cambridge Analytica had acquired this data by exploiting the sieve-like nature of Facebook's privacy policy. These revelations raised new questions about how carefully Facebook protects user data and privacy: despite having learned about this data harvesting by 2015, the company did not alert users and took only cursory actions to protect the data from further misuse. After the Cambridge Analytica story was reported, Congress summoned Mark Zuckerberg to testify about Facebook's practices (Kang et al., 2018) and a concerted effort arose among Facebook users to delete their profiles (Hsu, 2018).

Military and ICE Contracts: In 2018, the technology website Gizmodo revealed that Google was working with the U.S. Department of Defense (DoD) to develop artificial intelligence software that could analyze drone footage (Cameron and Conger, 2018). This effort, known as Project Maven, was part of a \$7.4 billion investment in AI by the DoD in 2017 (Cameron and Conger, 2018) and represented an opportunity for Google to gain billions of dollars in future defense contracts (Tiku, 2019). The project set off intense debate within Google, as many engineers expressed concern about facilitating drone strikes and began organizing to assert that “Google should not be in the business of war” (Shane and Wakabayashi, 2018). Project Maven, along with similar stories about tech companies partnering with the Trump administration—such as reports that Palantir was developing software for Immigration and Customs Enforcement (ICE) to facilitate deportations (Woodman, 2017)—prompted new organizing among tech workers and computer science students in opposition to tech industry contracts with U.S. defense and intelligence agencies, often centered around the slogans #TechWontBuildIt and #NoTechForICE (Conger and Metz, 2018; Glaser and Oremus, 2018; Goldberg, 2020; Mijente, 2019).

Algorithmic Bias: In 2016, ProPublica revealed that an algorithm used in criminal courts was biased against Black defendants, mislabeling them as future criminals at twice the rates of white defendants (Angwin et al., 2016). The dispute about whether the algorithm in question was, in fact, biased (Dieterich, Mendoza, and Brennan, 2016; Kleinberg, Mullainathan, and Raghavan, 2016) helped build interest in the emerging field of “algorithmic fairness.” Through books such as Cathy O’Neil’s *Weapons of Math Destruction* (O’Neil, 2017), Virginia Eubanks’ *Automating Inequality* (Eubanks, 2018a), and Safiya Noble’s *Algorithms of Oppression* (Noble, 2018) as well as articles demonstrating the biases in algorithms used in contexts ranging from facial recognition

(Buolamwini and Gebru, 2018) to hiring (Dastin, 2018), the public began to recognize algorithms as both fallible and discriminatory—potentially the source of more harm than good.

These and other tech-related controversies were a shock to many both inside and outside the world of technology, as they arrived in an era of widespread (elite) optimism about the beneficence of technology. Yet these controversies also brought public attention to what scholars in fields such as STS¹, philosophy of science, critical data and algorithm studies, information science, human-computer interaction, and law have long argued: technology is shaped by social forces, technology structures society in often deleterious ways, and technology cannot solve every social problem. Broadly speaking, these fields bring a “sociotechnical” approach to studying technologies (digital and otherwise), analyzing how technologies shape, are shaped by, and interact with society. The sociotechnical frame emphasizes the ways in which social actors and technological artifacts are intertwined as part of unified—rather than discrete—systems and serve to simultaneously and mutually constitute one another (Bijker and Law, 1994; Jasanoff, 2004; Suchman et al., 1999). As tech scandals mounted, these fields’ critical and sociotechnical insights, long ignored by most technologists and technology journalists, were newly recognized (or in some form recreated).

Many in the tech sector and academia diagnosed these ills as the result of an inattention to ethics: a lack of training in ethical reasoning for engineers and a dearth of ethical principles in engineering practice, which in turn led to the development of unethical technologies (Fiesler, 2018b; Karoff, 2019; Marantz, 2019; Raicu, 2017; Zunger, 2018). In response, academics, technologists, companies, governments, and more have embraced a broad set of goals often characterized with the label “tech ethics”: the introduction of considerations around ethics and social responsibility into digital technology education, research, development, use, and governance. In the span of just a few years, tech ethics has become a dominant discourse discussed in technology companies, academia, civil society organizations, and governments. For those committed to combating an array of injustices connected to digital technologies, the rise of tech ethics has produced a range of responses: on the one hand, excitement that technologists are increasingly considering their social responsibilities and impacts; on the other hand, frustration regarding the limited scope and impacts of what tech ethics discourse and practice has thus far entailed.

¹ Referring to the related areas of a) science and technology studies and b) science, technology, and society.

This essay summarizes these developments and debates in tech ethics. I first describe the primary forms of tech ethics and summarize the central critiques made against these efforts, which focus on tech ethics' abstract and toothless nature, corporate logics, narrow and individualistic focus, and capacity for strategic use (“ethics-washing”). Against the backdrop of these debates, I then turn to describing tech ethics itself as a terrain of contestation, where the central debate is not over *whether* ethics is desirable but over *what* ethics entails and *who* obtains the authority to define it. These debates suggest the need for a sociotechnical approach to tech ethics that explores the social construction and real-world effects of tech ethics. I introduce such an approach through four frames: determinism, solutionism, objectivity and neutrality, and sociotechnical systems.

2 The Rise of Tech Ethics

Although scholars, activists, and others have long considered the ethics and social impacts of technology, attention to developing and promoting technology ethics has rapidly grown across the tech industry, academia, civil society, and government. Across these sectors, one common treatment of tech ethics is through statements of ethical principles. One analysis of 36 prominent AI principles documents showed the rapid rise in these statements, from 2 in 2014 to 16 in 2018 (Fjeld et al., 2020). These documents tend to cover very similar themes, particularly fairness and non-discrimination, privacy, accountability, and transparency and explainability (Fjeld et al., 2020). Many of the documents also reference human rights, with some taking international human rights as the framework for analyzing and promoting tech ethics (Fjeld et al., 2020).

The broad and diverse nature of the work connected with tech ethics can make it difficult to define where the boundaries of “tech ethics” begin and end. While many people and institutions directly embrace the label, others are pursuing related efforts without any direct reference to tech ethics. As tech ethics becomes a more widespread frame, however, it is often used as a catchall to refer to any effort to study or improve the social impacts of technology. And although sometimes directly tied to the philosophical discipline of ethics (i.e., moral philosophy), the “ethics” in tech ethics is more frequently tied to applied forms of ethics such as codes of ethics, research ethics, and the lived impacts of digital technologies. In order to evaluate the term and the debates that surround it, I will largely restrict my discussions of “tech ethics” to those people and organizations

that explicitly embrace the label (and its close analogues or derivatives, such as AI ethics and algorithmic fairness).

2.1 Tech Industry

The most pervasive treatment of tech ethics within tech companies has come in the form of ethics principles and ethics oversight bodies. Companies like Microsoft, Google, and IBM have developed and publicly shared AI ethics principles, which include statements such as “AI systems should treat all people fairly” and “AI should [...] Be socially beneficial” (IBM, 2018; Microsoft, 2018; Pichai, 2018). These principles are often supported through dedicated ethics teams and advisory boards within companies, with such bodies in place at companies including Microsoft, Google, Facebook, Alphabet subsidiary DeepMind, and policing technology company Axon (Legassick and Harding, 2017; Nadella, 2018; Novet, 2018; Vincent and Brandom, 2018; Walker, 2018). Companies such as Google, Accenture, and Clifford Chance have also begun offering tech ethics consulting services (Simonite, 2020; Accenture, n.d.; Clifford Chance, n.d.).

As part of these efforts, the tech industry has formed several coalitions aimed at advancing a common dialogue about safe and ethical artificial intelligence. In 2015, Elon Musk and Sam Altman (the then-president of the tech incubator Y Combinator) created OpenAI, an independent research organization that aims to develop socially beneficial artificial intelligence and mitigate the “existential threat” presented by AI, with more than \$1 billion in donations from major tech executives and companies (Dowd, 2017). A year later, Amazon, Facebook, DeepMind, IBM, and Microsoft founded the Partnership on AI (PAI), a nonprofit coalition to shape best practices in AI development, advance public understanding of AI, and support socially beneficial applications of AI (Finley, 2016; Hern, 2016).

2.2 Academia

Computer and information science programs at universities have rapidly increased their emphasis on ethics training in curricula. While some universities have taught computing ethics courses (within both computer science and other fields) for many years (Grosz et al., 2019; Reich et al., 2020; Shilton et al., 2017), the emphasis on ethics within computing education has increased dramatically in recent years (Fiesler, Garrett, and Beard, 2020). When information science

professor Casey Fiesler tweeted a link to an editable spreadsheet of tech ethics classes in November 2017, it quickly grew to a crowdsourced list of more than 200 courses (Fiesler, 2018a). This plethora of courses represents a dramatic shift in computer science training and culture, with ethics becoming a popular topic of discussion and study after being largely ignored by the mainstream of the field just a few years prior.

Research in computer science and related fields has also become increasingly focused on the ethics and social impacts of computing. This trend is observable in the recent explosion of conferences and workshops related to computing ethics. The ACM Conference on Fairness, Accountability, and Transparency (FAccT, formerly FAT*) and the AAAI/ACM Conference on AI, Ethics, and Society (AIES) both held their first annual meetings in February 2018 and have since grown rapidly. Through 2019, there had been more than 30 workshops related to fairness and ethics at major computer science conferences (ACM FAccT Conference, 2020). Many universities have supported these curricular and research efforts through the creation of new institutes focused on the social implications of technology. 2017 alone saw the launch of the AI Now Institute at NYU (AI Now Institute, 2017), the Princeton Dialogues on AI and Ethics (Sharlach, 2019), and the MIT/Harvard Ethics and Governance of Artificial Intelligence Initiative (MIT Media Lab, 2017). More recently formed centers include the MIT College of Computing (MIT News Office, 2018); the Stanford Institute for Human-Centered Artificial Intelligence (Adams, 2019); and the University of Michigan Center of Ethics, Society, and Computing (Marowski, 2020).

2.3 Civil Society

Numerous civil society organizations have also coalesced around tech ethics, with strategies that include grantmaking and developing principles and toolkits. Organizations such as the MacArthur and Ford Foundations have begun exploring and making grants in tech ethics (Robinson and Bogen, 2016). For instance, the Omidyar Network, Mozilla Foundation, Schmidt Futures, and Craig Newmark Philanthropies partnered on the Responsible Computer Science Challenge, which awarded \$3.5 million between 2018 and 2020 to support efforts to embed ethics into undergraduate computer science education (Mozilla, 2018). Many foundations also contribute to the research, conferences, and institutes that have emerged in recent years.

Other organizations have been created or have expanded their scope to consider the implications and governance of digital technologies. For example, the American Civil Liberties Union (ACLU) has begun hiring technologists and is increasingly engaged in debates and legislation related to new technology. Organizations such as Data & Society, Upturn, the Center for Humane Technology, and Tactical Tech study the social implications of technology and advocate for improved technology governance and design practices.

Many advocates call for engineers to follow an ethical oath modeled after the Hippocratic Oath, an ethical oath taken by physicians (Eubanks, 2018a; O'Neil, 2017; Patil, 2018; Simonite, 2018). In 2018, for instance, the organization Data for Democracy partnered with Bloomberg and the data platform provider BrightHive to develop a code of ethics for data scientists, developing 20 principles that include “I will respect human dignity” and “It’s my responsibility to increase social benefit while minimizing harm” (Data4Democracy, 2018). Former U.S. Chief Data Scientist DJ Patil described the event as the “Constitutional Convention” for data science (Eubanks, 2018b). A related effort, produced by the Institute for the Future and the Omidyar Network, is the Ethical OS Toolkit, a set of prompts and checklists to help technology developers “anticipat[e] the future impact of today’s technology” and “not [...] regret the things you will build” (The Institute for the Future and Omidyar Network, 2018).

2.4 Government

Many governments have also taken up the mantle of tech ethics, developing commissions and principles dedicated to the topic. In the United States, for example, the National Science Foundation formed a Council for Big Data, Ethics, and Society (Council for Big Data, 2014), the National Science and Technology Council published a report about AI that emphasized ethics (National Science and Technology Council, 2018), and the Department of Defense adopted ethical principles for AI (U.S. Department of Defense, 2020). Elsewhere, governing bodies in Dubai (Smart Dubai, 2018), Europe (European Commission High-Level Expert Group on Artificial Intelligence, 2019), Japan (Integrated Innovation Strategy Promotion Council, 2019), and Mexico (Martinho-Truswell et al., 2018), as well as international organizations such as the OECD (Organisation for Economic Co-operation and Development, 2019) have all put forward documents stating principles for ethical AI development. A 2019 analysis of global AI ethics

guidelines found 84 such documents (with more than a third from the U.S. and U.K. and none from Africa or South America) espousing a common set of principles: transparency, justice and fairness, non-maleficence, responsibility, and privacy (Jobin, Ienca, and Vayena, 2019).

3 The Limits of Tech Ethics

Despite the rapid adoption of “ethics” as an analytic frame for digital technologies, critical analyses of these early efforts have indicated that tech ethics suffers from several core limitations. First, the actual principles espoused by tech ethics statements are too abstract and lacking in enforcement mechanisms to reliably spur ethical behavior in practice. Second, as ethics is incorporated into tech companies, ethical ideals are subsumed into corporate logics and incentives. Third, by emphasizing the design decisions of individual engineers, tech ethics overlooks the structural forces that shape technology’s harmful social impacts. Collectively, these issues suggest that the emphasis on ethics represents a strategy of technology companies “ethics-washing” their behavior with a façade of ethics while largely continuing with business-as-usual.

3.1 Tech ethics principles are abstract and toothless

Tech ethics codes deal in universal principles (Greene, Hoffmann, and Stark, 2019). In 2016, for example, Accenture published a report explicitly outlining “a universal code of data ethics” (Accenture, 2016). Professional computing societies also present ethical commitments in a highly abstract form, encouraging computing professionals, for instance, “to be ever aware of the social, economic, cultural, and political impacts of their actions” and to “contribute to society and human well-being” (Stark and Hoffmann, 2019). Ethics codes in computing and information science are notably lacking in explicit commitments to normative principles (Stark and Hoffmann, 2019).

The emphasis on universal principles papers over the fault lines of debate and disagreement that spurred the emergence of tech ethics as a widespread discourse in the first place. The recent upsurge in tech ethics was prompted by a spate of tech scandals and ensuing critiques of digital technologies. Yet the principles that have been developed embody an almost eerie level of agreement: two 2019 reports on global AI ethics guidelines similarly emphasized a “global convergence” (Jobin, Ienca, and Vayena, 2019) and a “consensus” (Fjeld et al., 2020) in the principles espoused. Although these documents tend to reflect a common set of global principles,

the actual interpretation and implementation of these principles raise substantive conflicts (Jobin, Ienca, and Vayena, 2019). The superficial consensus around abstract ideals may thus be hindering substantive deliberation regarding whether the chosen values are appropriate, how those values should be balanced in different contexts, and what those values actually entail in practice.

This level of abstraction and consensus is particularly troubling due to a lack of mechanisms to enact or enforce the principles embodied in tech ethics principles. When framed at such a high level of abstraction, values such as fairness and respect are unable to guide specific actions (Mittelstadt, 2019). Nor are these principles binding. The ethics oversight boards in companies such as Google and Axon and the ethics principles in companies and governments around the world lack the independent authority to veto projects or require certain behaviors (Harwell, 2018; Jobin, Ienca, and Vayena, 2019; Knight, 2019). Similarly, professional computing organizations such as the IEEE and ACM lack the power to meaningfully sanction individuals who violate their codes of ethics (Mittelstadt, 2019). Moreover, unlike fields such as medicine, which has a strong and established emphasis on professional ethics, computing lacks a common aim or fiduciary duty to unify disparate actors around shared ethical practices (Mittelstadt, 2019). All told, “Principles alone cannot guarantee ethical AI” (Mittelstadt, 2019).

3.2 Tech ethics is subsumed into corporate logics and incentives

Digital technology companies have embraced ethics as a matter of corporate concern, stemming in large part from the desire to avoid reputational or financial harm. In recent SEC filings, both Alphabet and Microsoft noted the potential harms they could face for products that are deemed unethical (Simonite, 2019). Corporate tech ethics thus emphasizes the presentation of ethical behavior for scrutinizing audiences. An ethnography of ethics work in Silicon Valley found that “[p]erforming, or even showing off, the seriousness with which a company takes ethics becomes a more important sign of ethical practices than real changes to a product” (Metcalf, Moss, and boyd, 2019). For instance, after an effort at Twitter to reduce online harassment stalled, an external researcher involved in the effort noted, “The impression I came away with from this experience is that [Twitter was] more sensitive to deflecting criticism than in solving the problem of harassment” (Seetharaman, 2020).

Ethics is also framed by companies in terms of its direct alignment with business strategy. A software engineer at LinkedIn described algorithmic fairness as being profitable for companies, arguing, “If you’re very biased, you might only cater to one population, and eventually that limits the growth of your user base, so from a business perspective you actually want to have everyone come on board, so it’s actually a good business decision in the long run” (Johnson, 2019). Similarly, one of the people behind the Ethical OS toolkit described being motivated to produce “a tool that helps you think through [societal] consequences and makes sure what you’re designing is good for the world and good for your longer-term bottom line” (Pardes, 2018).

Finding this alignment between ethics and business is an important task for those charged with promoting ethics in tech companies. Recognizing that “[market] success trumps ethics,” individuals focused on ethics in Silicon Valley feel pressure to align ethical principles with corporate revenue sources (Metcalf, Moss, and boyd, 2019). As one senior researcher in a tech company notes, “the [ethics] system that you create has to be something that people feel adds value and is not a massive roadblock that adds no value, because if it is a roadblock that has no value, people literally won’t do it, because they don’t have to” (Metcalf, Moss, and boyd, 2019). When ethical ideals are at odds with a company’s bottom line, they are met with resistance (Marantz, 2019).

The emphasis on business strategy creates significant conflicts with ethics in technology companies, whose business models often rely on extractive and exploitative practices that were partially responsible for the scandals that prompted the “techlash” in the first place. Indeed, announcements by Facebook and Twitter that they would invest in privacy, security, and combating misinformation and abusive behavior led to rapid declines in the companies’ stock values (Neate, 2018; Phillips, 2018). Moreover, even as tech companies espouse a devotion to ethics, they continue to develop products and services that raise ethical red flags but promise significant profits. For example, even after releasing AI ethics principles that include safety, privacy, and inclusiveness (Microsoft, 2018) and committing not to “deploy facial recognition technology in scenarios that we believe will put these freedoms at risk” (Smith, 2018), Microsoft invested in AnyVision, an Israeli facial recognition company that supports military surveillance of Palestinians in the West Bank (Solon, 2019). Microsoft claimed that AnyVision complied with its

ethics principles (even while former AnyVision employees admitted that the company did not meet Microsoft's ethical standards), suggesting either that these principles are so flexible as to allow such applications or that they are followed only when doing so is convenient.

These examples indicate that tech ethics is being subsumed into existing tech company logics and business practices rather than meaningfully challenging or changing those logics and practices (even if some individuals within companies do want to create meaningful change). This absorption allows companies to take up the mantle of caring about ethics without making substantive changes to their processes or business strategies. The goal in companies is to find practices “which the organization is *not yet doing* but is *capable of doing*” (Metcalf, Moss, and boyd, 2019), indicating an effort to find relatively costless reforms that provide the veneer of ethical behavior. Ethics vision statements “co-opt the language of some critics,” taking critiques grounded in a devotion to equity and social justice and turning them into ethical discussions that are akin to “conventional business ethics” (Greene, Hoffmann, and Stark, 2019). As they integrate these principles into new practices and products, tech companies “are learning to speak and perform ethics rather than make the structural changes necessary to achieve the social values underpinning the ethical fault lines that exist” (Metcalf, Moss, and boyd, 2019).

These limits to corporate tech ethics are exemplified by Google's firing of Timnit Gebru (and later Meg Mitchell) following the company's concerns about a paper examining the limitations and harms of large language models, which are central to Google's business (Hao, 2020). Despite Gebru's and Mitchell's supposed charge as co-leads of Google's ethical AI team, Google objected to a paper they had written with several internal and external colleagues about ethical concerns related to these AI models, suggesting that the authors were insufficiently attentive to recent technical advances that mitigate these concerns (Hao, 2020). Soon after, journalists revealed that Google had expanded its review of papers that discuss “sensitive topics,” telling researchers, for instance, to “take great care to strike a positive tone” regarding Google's technologies and products (Dave and Dastin, 2020). Thus, even as Google publicly advertised its care for the ethics of its technologies, internally the company was acting strongly to curtail criticisms that it deemed threatening to its core business interests.

3.3 *Tech ethics has a myopic focus on individual engineers and technology design*

Tech ethics typically emphasizes the roles and responsibilities of engineers, paying relatively little attention to the broader environments in which these individuals work. Although professional codes in computing and related fields state general commitments to the public, profession, and one's employer, "the morality of a profession's or an employer's motives are not scrutinized" (Stark and Hoffmann, 2019). Similarly, ethics within computer science curricula tends to focus on ethical decision making for individual engineers (Silbey, 2018).

From this individualistic frame comes an emphasis on appealing to the good intentions of engineers, with the assumption that better design practices and procedures will lead to better technology. Ethics becomes a matter of individual engineers and managers "doing the right thing" "for the right reasons" (Metcalf, Moss, and boyd, 2019). Efforts to provide ethical guidance for tech CEOs rest on a similar logic: "if a handful of people have this much power—if they can, simply by making more ethical decisions, cause billions of users to be less addicted and isolated and confused and miserable—then, isn't that worth a shot?" (Marantz, 2019). The broader public beyond technical experts is not seen as having a role in defining ethical concerns or shaping the responses to these concerns (Greene, Hoffmann, and Stark, 2019).

Tech ethics therefore centers debates about *how* to build better technology rather than *whether* or *in what form* to build technology (let alone who gets to make such decisions). Underlying ethics work across academia, civil society, and tech companies is an assumption that artificial intelligence and machine learning are "inevitable," such that "'better building' is the only ethical path forward" (Greene, Hoffmann, and Stark, 2019). In turn, tech ethics diagnoses the harmful social consequences of technology as treatable through technical and procedural solutions (Metcalf, Moss, and boyd, 2019). Ethics teams within tech companies have developed and shared numerous ethics and fairness toolkits, including Datasheets at Microsoft (Gebu et al., 2018), Model Cards at Google (Mitchell et al., 2019), AI Fairness 360 at IBM (Varshney, 2018), the Fairness Tool at Accenture (Peters, 2018), and Fairness Flow at Facebook (Gershgorn, 2018).

Although efforts such as these stand to remedy certain harms that result from digital technology, approaches to tech ethics that focus on the design decisions of engineers omit much of what

scholarship in STS, law, and other fields has diagnosed regarding the source of social harms connected to digital technologies. For example, scholars and journalists have articulated the social harms associated with business models that rely on collecting massive amounts of data about the public (Schneier, 2016; Zuboff, 2018), companies that wield monopolistic power (Khan, 2017; Wu, 2018), technologies that are built through the extraction of natural resources and the abuse of workers (Crawford and Joler, 2018; Dobbe and Whittaker, 2019; Evans, 2019; Gray and Suri, 2019), and the exclusion of women, minorities, and non-technical experts from decisions with significant social impacts (Jasanoff, 2006; West, Whittaker, and Crawford, 2019). To the extent that efforts based in individual design decisions are taken as the heart of what it means to “do” ethics, they bear the “risk of a premature foreclosure of the fundamentally open-ended and irresolvable questions that underlie human value commitments” into a set of procedures that must be followed (Metcalf, Moss, and boyd, 2019).

The focus on improving design also relies on a narrow theory of change for how to reform technology. Regardless of their intentions and the design frameworks at their disposal, individual engineers typically have little power to shift corporate strategy. Executives can limit which teams know about a project and what they know about it, enforcing strict secrecy and segmenting project teams to prevent knowledge and internal dissent about controversial projects (Conger and Metz, 2018; Gallagher, 2018). Even when engineers do know about and protest projects, the result is often them resigning or being replaced rather than the company changing course (Conger and Metz, 2018; Simonite, 2018). Instead of efforts focused explicitly on design and ethics, many improvements in tech systems and new regulations have been the result of collective action among tech workers as well as external pressure and organizing from activists, journalists, workers, and scholars (Crawford et al., 2019; Haskins, 2020).

These structural conditions place significant barriers on the benefits that design-oriented tech ethics will be able to achieve. As MIT anthropologist Susan Silbey notes in regard to teaching engineering ethics, “while we might want to acknowledge human agency and decision-making at the heart of ethical action, [...] we blind ourselves to the structure of those choices—incentives, content, and pattern—if we focus too closely on the individual and ignore the larger pattern of opportunities and motives that channel the actions we call ethics” (Silbey, 2018). By ignoring the

contexts of technology production and how technology interacts with society, we risk tinkering on the margins of technology, developing a lingo and practices that come to define ethical behavior while leaving in place the structures that generated the controversies spurring the rise of tech ethics.

3.4 Tech ethics has become an avenue for ethics-washing

As evidence of tech ethics' limitations has grown, many have critiqued tech ethics as a strategic effort among technology companies to quell public scrutiny rather than as a noble effort to take responsibility for technology's social impacts. This strategy has been labeled "ethics-washing" (i.e., "ethical white-washing"): using the language of ethics to paint a superficial portrait of ethical behavior in order to avoid heightened public backlash and the introduction of regulations that would require substantive concessions (Metzinger, 2019; Nemitz, 2018; Wagner, 2018).

In other words, ethics discourse has become a convenient way for tech companies to defuse criticism and regulation by creating structures for self-governance without any commitment to meaningfully altering their behavior. As an ethnography of ethics in Silicon Valley found, "It is a routine experience at 'ethics' events and workshops in Silicon Valley to hear ethics framed as a form of self-regulation necessary to stave off increased governmental regulation" (Metcalf, Moss, and boyd, 2019). Recognizing this strategy casts important "flaws" of tech ethics as features rather than bugs: by focusing public attention on the actions of individual engineers and particular technical limitations (such as algorithmic bias), companies perform a sleight-of-hand that shifts structural questions about power and profit out of view.

Thomas Metzinger, a philosopher who served on the European Commission's High-Level Expert Group (HLEG) on Artificial Intelligence to develop AI ethics guidelines (European Commission High-Level Expert Group on Artificial Intelligence, 2019), provides a particularly striking account of ethics-washing in action (Metzinger, 2019). The HLEG on AI contained only four ethicists out of 52 total people and was dominated by representatives from industry. Metzinger's own work to develop "Red Lines" that AI applications should not cross was "watered down" by industry representatives eager for a "positive vision" for AI. All told, Metzinger describes the HLEG's guidelines as "lukewarm, short-sighted and deliberately vague" and concludes that the tech

industry is “using ethics debates as elegant public decorations for a large-scale investment strategy” (Metzinger, 2019).

Tech companies have further advanced this “ethics-washing” agenda through funding academic research and events. Many of the scholars writing about tech policy and ethics are funded by Google, Microsoft, and others, yet often do not disclose this funding (Google Transparency Project, 2017; Williams, 2019). Tech companies also provide funding for prominent academic conferences, including the ACM Conference on Fairness, Accountability, and Transparency (FAccT); and the AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES); and the Privacy Law Scholars Conference (PLSC). Even if these funding practices do not directly influence the research output of individual scholars, they allow tech companies to shape the broader academic and public discourse regarding tech ethics, raising up certain voices and conversations at the expense of others.²

Further debate about the value and impacts of tech ethics erupted in December 2019, spurred by an article written by MIT graduate student Rodrigo Ochigame (Ochigame, 2019) in the wake of the revelations that the MIT Media Lab—a center that conducts research and provides grants related to tech ethics—had received secret funding from Jeffrey Epstein, the financier charged with sex trafficking of minors (Farrow, 2019; Tracy and Hsu, 2019). Describing his experiences working in Joi Ito’s AI ethics group at the Media Lab and collaborating with the Partnership on AI, Ochigame articulated how “the discourse of ‘ethical AI’ [...] was aligned strategically with a Silicon Valley effort seeking to avoid legally enforceable restrictions of controversial technologies” (Ochigame, 2019). Ochigame described witnessing firsthand how the Partnership on AI made recommendations that “aligned consistently with the corporate agenda” by reducing political questions about the criminal justice system to matters of technical consideration. A central part of this effort was tech companies strategically funding researchers and conferences in order to generate a widespread discourse about “ethical” technology. Finding that “the corporate lobby’s effort to shape academic research was extremely successful,” Ochigame concluded that “[b]ig tech money and direction proved incompatible with an honest exploration of ethics.”

² The integrity of academic tech ethics has been further called into question due to funding from other sources beyond tech companies (Domínguez et al., 2019; Farrow, 2019; Mboya, 2019).

Some believed that Ochigame oversimplified the story, failing to fully credit the many people behind tech ethics and the reforms that this movement have prompted (Darling, 2019; Epstein, 2019; Sindere, 2019). On this view, all of the work by scholars and activists increasing attention to technological harms and pushing for legislation was worthy of praise as more than just corporate capture. Yet many of the people (often activists and scholars of color) centrally involved in efforts to expose and resist the harms of digital technology see their work as distinct from tech ethics, which represents the narrow domain of efforts typically promulgated by tech companies. Safiya Noble described Ochigame's article as "All the way correct and worth the time to read" (Noble, 2019). Lilly Irani and Ruha Benjamin expressed similar sentiments, noting that "AI ethics is not a m[o]v[e]m[e]nt" (Irani, 2019) and that "many of us don't frame our work as 'ethical AI'" (Benjamin, 2019).

This debate in response to Ochigame's article exposed the fault lines at the heart of tech ethics. While tech ethics is often framed as encompassing the broad societal debates about the social impacts of technology, many of the people at the forefront of those efforts see tech ethics as the narrower (typically industry-led) efforts to promote "ethics" in technology. Where, then, do the bounds of tech ethics lie? Who is behind tech ethics: tech companies, or the activists and scholars who are highlighting harmful technologies and advocating for reforms? What are the appropriate strategies for pursuing ethical technology: through reforming design practices, or through reforming the political and economic forces that generate and govern technology? And, perhaps most fundamentally, what is the value of tech ethics: does it provide valuable reforms (even if incremental ones), or is it an active hindrance to achieving more just technology? The answers to these contested questions will shape the future of efforts to pursue more just digital infrastructures and technologies.

4 The Contestation of Tech Ethics

The debates described in the previous section reveal that the central question regarding tech ethics is not over *whether* it is desirable to be ethical, but over *what* "ethics" entails and *who* gets to define it. In this sense, ethics is an "essentially contested concept" (Gallie, 1955), where the central debates regard the meaning—rather than desirability—of ethics. Akin to the debates described

above, the “ethics” in tech ethics tends to take on four overlapping yet often conflicting meanings: moral justice, corporate values, legal risk, and compliance (Moss and Metcalf, 2020).

Such definitional debates have significant stakes. In the context of antidiscrimination law, Kimberlé Crenshaw describes the “definitional tension” between an “expansive view” of antidiscrimination that strives to eradicate the oppression of Blacks and a “restrictive view” of antidiscrimination that aims only to prevent discrimination based explicitly on race (Crenshaw, 1988). In any given instance, “specific interpretations proceed largely from the world view of the interpreter” (Crenshaw, 1988). Similarly, different interpretations of “tech ethics” suggest drastically different paths forward: tech ethics embodies a conflict between an expansive view that aims to remedy a broad range of injustices associated with digital technologies and a restrictive view that aims to limit a specific set of tangible harms caused by these technologies. Because of these conflicting accounts co-existing within the same term, to call for tech ethics is, in Crenshaw’s terms, “to demand nothing specific” (Crenshaw, 1988).

In this sense, tech ethics is itself a terrain of contestation whose impacts hinge on who is able to demarcate the bounds of legitimate authority regarding what ethics entails in relation to technology and who is responsible for technologies’ social impacts. Whether it be technology companies projecting procedural toolkits as solutions to ethical dilemmas, computer scientists reducing normative questions into mathematical metrics, academic tech ethics institutes being funded by billionaires and led primarily by white men (Gershgorn, 2019), or tech ethics principles being disseminated predominantly by the U.S. and Western Europe, the contestation of tech ethics centers on certain actors attempting to acquire or maintain authority over what it means for technology to be “ethical,” at the expense of others. These efforts embody practices of “boundary-work,” which involves drawing boundaries between legitimate and illegitimate sources of authority (Gieryn, 1983). Drawing these boundaries can help scientists acquire intellectual authority and maintain professional autonomy, often in ways that exclude women, minorities, the Global South, and other publics (Collins, 2000; Haraway, 1988; Visvanathan, 2005).

Examples of how ethics has been implemented in two other domains—science and business—shed light on the nature and stakes of present debates about tech ethics.

4.1 *Ethics in Science*

Many areas of science have embraced ethics in recent decades in response to debates about the social implications of emerging research and applications. Despite the seeming promise of ethics in science, however, existing approaches to promoting ethics in science are limited in their ability both to raise debates about the structure and values of science and to promote democratic governance of science. Science ethics bodies suffer from limited “ethical imaginations” and are often primarily concerned with “keep[ing] the wheels of research turning while satisfying publics that ethical standards are being met” (Jasanoff, 2016). “Ethical analysis that does not advance such instrumental purposes tends to be downgraded as not worthy of public support” (Jasanoff, 2016).

Rather than interrogating fundamental questions about the purposes of research or who gets to shape that research, ethics has become increasingly institutionalized, instrumentalized, and professionalized, with an emphasis on filling out forms and checking off boxes (Jasanoff, 2016; Reardon, 2013). In turn, “systems of ethics [...] play key roles in eliding fundamental social and political issues” that inhere in the conception and development of research (Reardon, 2011). For instance, efforts to introduce ethics into genetic research throughout the 1990s and 2000s treated ethics “as something that could be added onto science—and not something that was unavoidably implicit in it,” thus obscuring the “bioconstitutional” questions about racial groups at the heart of genetic testing projects (Reardon, 2011). Because “ethical choices inhered in efforts to study human genetic variation, regardless of any explicit effort to *practice ethics*,” these research projects “bypass[ed] responsibility for their roles in co-constituting natural and moral orderings of human difference, despite efforts to address ethics at the earliest stages of research design” (Reardon, 2011).

The turn to ethics can also entail an explicit effort among scientists to defuse external scrutiny and to develop a regime of self-governance. For example, biologists in the 1970s, frightened by calls for greater public participation in genetic engineering, organized a conference at the Asilomar Conference Center in California. The scientific community at Asilomar pursued two, intertwined goals. In order to present a unified and responsible public image, the Asilomar organizers restricted the agenda to eschew discussions of the most controversial applications of genetic engineering

(biological warfare and human genetic engineering). And in order to convince the American public and politicians to allow biologists to self-govern their pursuit of genetic engineering, the Asilomar attendees “redefined the genetic engineering problem as a technical one” that only biologists could credibly discuss (Wright, 2001). Although often hailed as a remarkable occasion of scientific self-sacrifice for the greater good, accounts from the conference itself demonstrate that “[s]elf-interest, not altruism, was most evident at Asilomar,” as not making any sacrifices and appearing self-serving would have only invited stringent, external regulation (Wright, 2001).

Tech ethics mirrors many of these attributes in scientific ethics. As with ethics in other fields of science, tech ethics involves a significant emphasis on institutionalized design practices, often entailing checklists and worksheets. Mirroring ethics in genetic research, the emphasis on ethical design in computer science departments and tech companies treats ethics as something that can be *added on* to digital technologies by individual engineers, overlooking the epistemologies and economic structures that underlie these technologies and their harms. Furthermore, tech companies and computer scientists focusing on technical challenges such as algorithmic fairness mirror the strategic efforts of molecular biologists at Asilomar to reframe moral questions as technical questions in order to retain self-regulation.³ The removal of red lines in the European Commission’s High-Level Expert Group on AI bears a striking resemblance to the exclusion of controversial topics from the agenda at Asilomar.

4.2 *Corporate Ethics and Co-optation*

Codes of ethics have long been employed by groups of experts (e.g., doctors and lawyers) to codify a profession’s expectations regarding culture and behavior and to shore up the profession’s public reputation (Abbott, 1983; Metcalf, 2014). Similarly, companies across a wide range of sectors have embraced ethics codes, particularly over the past half century and typically in response to public perceptions of unethical behavior (Wood and Rimmer, 2003).

Yet it has long been clear that the public benefits of corporate ethics codes are minimal. While ethics codes can help make a group appear ethical, on their own they do little to promote a culture

³ In an ironic parallel, the Future of Life Institute organized an Asilomar Conference on Beneficial AI in 2017, leading to the development of 23 “Asilomar AI Principles” (Future of Life Institute, 2017).

of ethical behavior (Wood and Rimmer, 2003). The primary goal of business ethics has instead been the “inherently unethical” motivation of corporate self-preservation: to reduce public and regulatory scrutiny by promoting a visible appearance of ethical behavior (Cressey and Moore, 1983; Wood and Rimmer, 2003). This emphasis on corporate reputation and profit is facilitated by ethics codes making universal moral claims that “are extremely important as claims but extremely vague as rules” and emphasizing individual actors and behaviors, leading to a narrow, “one-case-at-a-time approach to control and discipline” (Abbott, 1983). Ethics codes in the field of information systems have long exhibited a notable lack of explicit moral obligations for computing professionals (Oz, 1992; Stark and Hoffmann, 2019).

Business ethics is indicative of the broader phenomenon of co-optation: an institution incorporating elements of external critiques from groups such as social movements—often gaining the group’s support and improving the institution’s image—without meaningfully acting on that group’s demands or providing that group with decision-making authority (Gamson, 1975; Selznick, 1948; Trumpy, 2014). The increasing centrality of companies as the target of social movements has led to a particular form of co-optation called “corporatization,” in which “corporate interests come to engage with ideas and practices initiated by a social movement and, ultimately, to significantly shape discourses and practices initiated by the movement” (King and Busa, 2017). Through this process, large corporate actors in the United States have embraced “diluted and deradicalized” elements of social movements “that could be scaled up and adapted for mass markets” (King and Busa, 2017). Two factors make movements particularly susceptible to corporatization: heterogeneity—movement factions that are willing to work with companies gain influence through access to funding—and materiality—structural changes get overlooked in favor of easily commodifiable technological “fixes.” By participating in movement-initiated discourses, companies are able to present themselves as part of the solution rather than part of the problem, and in doing so can avoid more restrictive government regulations.

Tech ethics embodies many of the attributes of corporate ethics, particularly the significant efforts to shore up legitimacy and avoid external regulation. Abstract and individualized tech ethics codes reproduce the virtue signaling and self-preservation behind traditional business ethics. And in a notable example of co-optation and corporatization, technology companies have promoted tech

ethics as a diluted and commoditized version of tech-critical discourses that originated largely from outside of technology circles. Because societal efforts to improve technology are often aimed at companies and include both heterogeneity and materiality, it is particularly vulnerable to corporatization. Through this process of corporatization, tech companies are using tech ethics to present themselves as part of the solution rather than part of the problem and are using funding to empower the voices of certain scholars and academic communities. The power of this influence can be seen in the expanding scope of work that is published and discussed at conferences, workshops, and other events under the banner of “tech ethics.” Even scholars who do not embrace the tech ethics label are increasingly subsumed into this category, either lumped into it by others or compelled into it as opportunities to publish research, impact policymakers, and receive grants are increasingly shifting to the terrain of “ethics.”

4.3 The Future of Tech Ethics

Drawing on these two examples of ethics in action leads to two conclusions about tech ethics. First, the parallels to ethics in science and business indicate that current tech ethics discourse is likely to enable technologists and technology companies to label themselves as “ethical” without substantively altering their practices. Although there are multiple notions of tech ethics currently coexisting—many individuals and groups, including some within tech companies, are pursuing expansive forms of tech ethics—the influence of tech companies makes it likely that their narrow vision of “tech ethics” will subsume any others. Furthermore, many of the most prominent voices regarding tech ethics are white men who want to claim expertise and thought leadership while ignoring the work of established fields and scholars, many of whom are women and people of color (Irani and Chowdhury, 2019; Mozilla, 2020). To the extent that tech ethics continues to follow the path of science ethics, business ethics, and corporatization, tech companies and others will be able to define ethics in such a way that technologies and technology companies are deemed “ethical” even while continuing to produce significant social harm. For the tech companies that dominate tech ethics discourse, such narrowing is precisely the point.

Second, rather than treating “ethics” as well-defined and inherently desirable, those striving for substantive and structural improvements in digital technologies must look to the formulation and real-world effects of “tech ethics.” The examples of science and business ethics indicate that ethics

in practical contexts can be quite distinct from the normative demands and moral inquiry that gave rise to the embrace of ethics in the first place. Just as digital technology is often applied as a solution to social problems, tech ethics is today being applied as a solution to sociotechnical problems. As with technology itself, tech ethics is a tool with particular affordances that is being developed to serve social purposes. If “technologies can be assessed only in their relations to the sites of their production and use” (Suchman et al., 1999), then so too, we might say, tech ethics can be assessed only with regard to how it is conceived and how it affects the world in practice.

Thus, rather than presenting a unifying and beneficent set of principles and practices, tech ethics has emerged as a central site of struggle regarding the future of digital architectures, governance, and economies. As was hinted at in the varied responses to Ochigame’s critique of ethics-washing, many of the more radical critics of technology and inequality see themselves as outside of (if not in opposition to) the dominant strains of tech ethics. Discourses and practices of resistance (Anti-Eviction Mapping Project, 2020), defense (Lewis et al., 2018), abolition (Hamid, 2020; Stop LAPD Spying Coalition and Free Radicals, 2020), and decentering technology (Gangadharan and Niklas, 2019) have also emerged (often from activists and communities rather than academics) in response to the social injustices meted out through digital technologies. Although some may see all of these efforts as falling under the broad umbrella of an expansive tech ethics, these nascent movements embody distinct aspirations from the narrow mainstream of tech ethics. Conflating them under a common label risks giving tech ethics the imprimatur of radical, justice-oriented work even as its core tenets and practices eschew such commitments. Efforts to resist oppressive technological architectures and to study or support such efforts must be attentive to these porous and slippery boundaries.

5 A Sociotechnical Approach to Tech Ethics

Given these dynamics of contestation surrounding tech ethics, ethics will not, on its own, provide a salve for technology’s social harms. Nonetheless, taking a reflexive (Bloor, 1991) approach that applies core elements of tech ethics’ analyses of digital technology to tech ethics itself contains the seeds of a more robust approach to pursuing a more just society, digital and otherwise. One dimension of tech ethics (and critiques of tech ethics) is a focus on technology’s social impacts and on how technology interacts with society—in other words, approaching digital technologies

through a sociotechnical analysis. This sociotechnical approach, drawing on STS, has much to offer engineering ethics, shedding light on the responsibilities of engineers and how artifacts shape and are shaped by the social world (Johnson and Wetmore, 2007). This suggests the value of an approach to tech ethics that mirrors the sociotechnical approach to technology—one that can inform our understanding of the responsibilities of those behind tech ethics and how tech ethics shapes and is shaped by the social world. With this aim in mind, it is fruitful to consider tech ethics through the lens of four sociotechnical frames: determinism, solutionism, objectivity and neutrality, and sociotechnical systems.

5.1 Determinism

A central component of a sociotechnical approach to technology is rejecting technological determinism: the belief that technology evolves autonomously and determines social outcomes (Dafoe, 2015; Marx and Smith, 1994). Such an approach has been rejected through scholarship demonstrating that even as technology plays some role in shaping society, so too, simultaneously, does society shape technology (Bijker and Law, 1994; Jasanoff, 2004; Pinch and Bijker, 1987; Winner, 1986). Ethics in digital technology is today being treated through a similar sort of “ethical determinism,” with an underlying assumption that adopting “ethics” will lead to ethical technologies. Yet in both science and business there is a long history demonstrating that such an embrace of ethics is not sufficient to prompt substantive changes in behavior, and there is emerging evidence of the same limitation with regard to tech ethics. As with technology, ethics does not emerge in a vacuum and does not on its own determine sociotechnical outcomes. What is needed, then, is a new approach that looks to the sociotechnical complexities of tech ethics: how it shapes the development and governance of technologies (and collective understandings of technologies) as well as how ethical discourses and practices are themselves shaped by a variety of social forces.

5.2 Solutionism

Closely intertwined with a belief in technological determinism is the practice of technological solutionism: the expectation that technology can and will solve all social problems (Morozov, 2014). A great deal of scholarship has demonstrated how technological solutions not only typically fail to provide the intended solutions, but also can exacerbate the problems they are intended to solve (Ames, 2019; Green, 2019; Morozov, 2014; Toyama, 2015). Yet even as tech ethics debates

have highlighted how technology is not always the answer to social problems, a common response has been to embrace an “ethical solutionism”: promoting ethics principles and practices as the solution to these sociotechnical problems. A notable example (at the heart of many tech ethics agendas) is the response to algorithmic discrimination through algorithmic fairness, which often centers narrow mathematical definitions of fairness but leaves in place the structural and systemic conditions that generate a great deal of algorithmic harms (Green, 2020; Hoffmann, 2019). Efforts to introduce ethics across science, business, and technology function similarly, providing an addendum of ethical language and practices on top of existing structures and epistemologies which themselves go largely uninterrogated. Thus, just as technical specifications of algorithmic fairness are insufficient to guarantee fair algorithms, tech ethics principles are insufficient to guarantee ethical technologies. Ethics principles, toolkits, and other mechanisms are just one component of what must be a broad array of approaches to improve technology.

5.3 Objectivity and Neutrality

Debates about biased and harmful technologies have led to an increasingly widespread recognition in computer science and the tech industry of notable insights from STS: engineers are not objective and technology is not neutral. It is clear that improving digital technologies requires grappling with the normative commitments of engineers and incorporating more voices into the design of technology (Costanza-Chock, 2020; Green and Viljoen, 2020). Yet even as tech ethics emphasizes principles such as fairness and inclusiveness, the range of perspectives remains quite narrow and ethics is treated as an objective, universal body of principles (Fjeld et al., 2020; Greene, Hoffmann, and Stark, 2019; Jobin, Ienca, and Vayena, 2019). The consensus around particular ethical principles may therefore say less about the objective universality of these ideals than about the narrow range of perspectives that have been given voice regarding tech ethics. In many cases, white and male former technology company employees are cast to the front lines of public influence regarding tech ethics (Irani and Chowdhury, 2019; Mozilla, 2020). Rather than treating tech ethics as the search for objective and universal moral principles, it is necessary to grapple with the standpoints and power of different actors, the normative principles embodied in different ethical frameworks, and potential mechanisms for adjudicating between conflicting ethical commitments.

5.4 *Sociotechnical Systems*

A key result of treating technologies as embedded within sociotechnical systems is expanding the frame of analysis beyond the technical artifact itself: rather than technical features fully determining an artifact's social impacts, the artifact and the social world “co-produce” social outcomes (Jasanoff, 2004). Technologies are not discrete objects that can be properly evaluated in the abstract from the context of their use (Suchman et al., 1999). Indeed, many of the animating concerns and critiques of tech ethics connect to harms that arise when digital technologies are integrated into society without proper attention to the social context and human interactions that will shape its impacts (Green and Chen, 2019; Green and Viljoen, 2020; Rose, 2019; Vincent, 2016; Vosoughi, Roy, and Aral, 2018). Nonetheless, efforts to promote ethical technology typically focus on the internal characteristics of tech ethics—which principles to promote, for instance—with little attention to the impacts that these efforts will or will not have when integrated into larger settings such as a tech company or computer science curriculum. In turn, tech ethics has had limited effects on technology production and has played a sometimes-legitimizing role for technology companies. But just as the “unintended consequences” of technology often represents a failure to consider how artifacts might be used or abused in practice (Jasanoff, 2016), these limited impacts of tech ethics must not be seen as having been impossible to predict, particularly given the precedents of science ethics, business ethics, and corporate co-optation. Rather than treating tech ethics as being defined by its formal characteristics such as the principles espoused, any attempts to promote more ethical technology must operate with an eye toward the many factors that will shape the real-world impacts of tech ethics efforts. This includes considering how central framings (such as ethics) could be redefined and wielded by different actors as well as how to robustly embody moral principles in the procedures involved in developing, evaluating, and deploying technology.

Acknowledgments

I thank Elettra Bietti, Anna Lauren Hoffmann, Jenny Korn, Kathy Pham, and Luke Stark for their comments on this essay. I also thank the Harvard STS community, particularly Sam Weiss Evans, for feedback and direction on an earlier iteration of this chapter.

References

- Abbott, Andrew. 1983. "Professional Ethics." *American Journal of Sociology* 88 (5):855-885.
- Accenture. 2016. Universal principles of data ethics. <https://www.accenture.com/acnmedia/pdf-24/accenture-universal-principles-data-ethics.pdf>.
- Accenture. N.d. AI Ethics & Governance. <https://www.accenture.com/us-en/services/applied-intelligence/ai-ethics-governance>.
- ACM FAccT Conference. 2020. ACM FAccT Network. <https://facctconference.org/network/>.
- Adams, Amy. 2019. Stanford University launches the Institute for Human-Centered Artificial Intelligence. *Stanford News*. https://news.stanford.edu/2019/03/18/stanford_university_launches_human-centered_ai/.
- AI Now Institute. 2017. The AI Now Institute Launches at NYU to Examine the Social Effects of Artificial Intelligence. <https://ainowinstitute.org/press-release-ai-now-launch>.
- Ames, Morgan G. 2019. *The Charisma Machine: The Life, Death, and Legacy of One Laptop per Child*: MIT Press.
- Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias. *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Anti-Eviction Mapping Project. 2020. *Counterpoints: A San Francisco Bay Area Atlas of Displacement & Resistance*. PM Press.
- Benjamin, Ruha. 2019. <https://twitter.com/ruha9/status/1208831999940714496>.
- Bijker, Wiebe E., and John Law, eds. *Shaping Technology/Building Society: Studies in Sociotechnical Change*. MIT press, 1994.
- Blake, Aaron. 2018. A new study suggests fake news might have won Donald Trump the 2016 election. *The Washington Post*. <https://www.washingtonpost.com/news/the-fix/wp/2018/04/03/a-new-study-suggests-fake-news-might-have-won-donald-trump-the-2016-election/>.

- Bloor, David. 1991. *Knowledge and Social Imagery*: University of Chicago Press.
- Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research.
- Cadwalladr, Carole, and Emma Graham-Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.
- Cameron, Dell, and Kate Conger. 2018. Google Is Helping the Pentagon Build AI for Drones. *Gizmodo*. <https://gizmodo.com/google-is-helping-the-pentagon-build-ai-for-drones-1823464533>.
- Clifford Chance. N.d. Tech Group. <https://www.cliffordchance.com/hubs/tech-group-hub/tech-group.html>.
- Collins, Patricia Hill. 2000. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*: Routledge.
- Conger, Kate, and Cade Metz. 2018. Tech Workers Now Want to Know: What Are We Building This For? *The New York Times*. <https://www.nytimes.com/2018/10/07/technology/tech-workers-ask-censorship-surveillance.html>.
- Costanza-Chock, Sasha. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need*: MIT Press.
- Council for Big Data, Ethics, and Society. 2014. Council for Big Data, Ethics, and Society. <https://bdes.datasociety.net>.
- Crawford, Kate, Roel Dobbe, Theodora Dryer, Genevieve Fried, Ben Green, Elizabeth Kazianas, Amba Kak, Varoon Mathur, Erin McElroy, Andrea Nill Sánchez, Deborah Raji, Joy Lisi Rankin, Rashida Richardson, Jason Schultz, Sarah Myers West, and Meredith Whittaker. 2019. AI Now 2019 Report. [https://ainowinstitute.org/AI Now 2019 Report.pdf](https://ainowinstitute.org/AI_Now_2019_Report.pdf).
- Crawford, Kate, and Vladan Joler. 2018. Anatomy of an AI System: The Amazon Echo as an anatomical map of human labor, data and planetary resources. <https://anatomyof.ai>.
- Crenshaw, Kimberlé Williams. 1988. "Race, Reform, and Retrenchment: Transformation and Legitimation in Antidiscrimination Law." *Harvard Law Review* 101 (7):1331-1387.

- Cressey, Donald R., and Charles A. Moore. 1983. "Managerial Values and Corporate Codes of Ethics." *California Management Review* 25 (4):53-77. doi: 10.2307/41165032.
- Dafoe, Allan. 2015. "On Technological Determinism: A Typology, Scope Conditions, and a Mechanism." *Science, Technology, & Human Values* 40 (6):1047-1076. doi: 10.1177/0162243915579283.
- Darling, Kate. 2019. https://twitter.com/grok_/status/1208434972564037633.
- Dastin, Jeffrey. 2018. 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>.
- Data4Democracy. 2018. Ethics Resources. <https://github.com/Data4Democracy/ethics-resources>.
- Dave, Paresh, and Jeffrey Dastin. 2020. Google told its scientists to 'strike a positive tone' in AI research - documents. *Reuters*. <https://www.reuters.com/article/us-alphabet-google-research-focus/google-told-its-scientists-to-strike-a-positive-tone-in-ai-research-documents-idUSKBN28X1CB>.
- Dieterich, William, Christina Mendoza, and Tim Brennan. 2016. COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity. *Northpoint Inc. Research Department*. http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf.
- Dobbe, Roel, and Meredith Whittaker. 2019. AI and Climate Change: How they're connected, and what we can do about it. *AI Now Institute*. <https://medium.com/@AINowInstitute/ai-and-climate-change-how-theyre-connected-and-what-we-can-do-about-it-6aa8d0f5b32c>.
- Domínguez, Alonso Espinosa, Remy Bassett-Audain, Husayn Karimi, Berenice Estrada, Claire Isabel Webb, Ruth Perry, Sally Haslanger, Jonathan King, Kevin Leonardo, Sarah Aladetan, Agnes Fury Cameron, Yarden Katz, Andrew Bolton, Lauren Surface, Kade Crockford, Katherine McConachie, Subrata Ghoshroy, and Alice Pote. 2019. Celebrating war criminals at MIT's 'ethical' College of Computing. *The Tech*. <https://thetech.com/2019/02/14/celebrating-war-criminals>.
- Dowd, Maureen. 2017. Elon Musk's Billion-Dollar Crusade to Stop the A.I. Apocalypse. *Vanity Fair*. <https://www.vanityfair.com/news/2017/03/elon-musk-billion-dollar-crusade-to-stop-ai-space-x>.

- Emery Jr., C. Eugene. 2016. Evidence ridiculously thin for sensational claim of huge underground Clinton sex network. *PolitiFact*. <https://www.politifact.com/factchecks/2016/nov/04/conservative-daily-post/evidence-ridiculously-thin-sensational-claim-huge-/>.
- Epstein, Greg. 2019. <https://twitter.com/gregmepstein/status/1208798637221974016>.
- Eubanks, Virginia. 2018a. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*: St. Martin's Press.
- Eubanks, Virginia. 2018b. A Hippocratic Oath for Data Science. <https://virginia-eubanks.com/2018/02/21/a-hippocratic-oath-for-data-science/>.
- European Commission High-Level Expert Group on Artificial Intelligence. 2019. Ethics Guidelines for Trustworthy AI. <https://ec.europa.eu/futurium/en/ai-alliance-consultation>.
- Evans, Will. 2019. Ruthless Quotas at Amazon Are Maiming Employees. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2019/11/amazon-warehouse-reports-show-worker-injuries/602530/>.
- Farrow, Ronan. 2019. How an Élite University Research Center Concealed Its Relationship with Jeffrey Epstein. *The New Yorker*. <https://www.newyorker.com/news/news-desk/how-an-elite-university-research-center-concealed-its-relationship-with-jeffrey-epstein>.
- Fiesler, Casey. 2018a. Tech Ethics Curricula: A Collection of Syllabi. <https://medium.com/@cfiesler/tech-ethics-curricula-a-collection-of-syllabi-3eedfb76be18>.
- Fiesler, Casey. 2018b. What Our Tech Ethics Crisis Says About the State of Computer Science Education. *How We Get To Next*. <https://howwegettonext.com/what-our-tech-ethics-crisis-says-about-the-state-of-computer-science-education-a6a5544e1da6>.
- Fiesler, Casey, Natalie Garrett, and Nathan Beard. 2020. "What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis." The 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20).
- Finley, Klint. 2016. Tech Giants Team Up to Keep AI From Getting Out of Hand. *Wired*. <https://www.wired.com/2016/09/google-facebook-microsoft-tackle-ethics-ai/>.
- Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Christopher Nagy, and Madhulika Srikumar. 2020. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based

- Approaches to Principles for AI. *Berkman Klein Center Research Publication No. 2020-1*.
<https://cyber.harvard.edu/publication/2020/principled-ai>.
- Foroohar, Rana. 2018. Year in a Word: Techlash. *Financial Times*.
<https://www.ft.com/content/76578fba-fca1-11e8-ac00-57a2a826423e>.
- Future of Life Institute. 2017. Beneficial AI 2017. <https://futureoflife.org/bai-2017/>.
- Gallagher, Ryan. 2018. Google Shut Out Privacy and Security Teams From Secret China Project. *The Intercept*. <https://theintercept.com/2018/11/29/google-china-censored-search/>.
- Gallie, Walter B. 1955. "Essentially Contested Concepts." *Proceedings of the Aristotelian Society* 56:167-198.
- Gamson, Howard. 1975. *The Strategy of Social Protest*: The Dorsey Press.
- Seeta Peña Gangadharan & Jędrzej Niklas. 2019. "Decentering technology in discourse on discrimination." *Information, Communication & Society*, 22:7, 882-899. doi: 10.1080/1369118X.2019.1593484.
- Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumeé III, and Kate Crawford. 2018. "Datasheets for Datasets." *arXiv preprint arXiv:1803.09010*.
- Gershgorn, Dave. 2018. Facebook says it has a tool to detect bias in its artificial intelligence. *Quartz*. <https://qz.com/1268520/facebook-says-it-has-a-tool-to-detect-bias-in-its-artificial-intelligence/>.
- Gershgorn, Dave. 2019. Stanford's new AI institute is inadvertently showcasing one of tech's biggest problems. *Quartz*. <https://qz.com/1578617/stanfords-new-diverse-ai-institute-is-overwhelmingly-white-and-male/>.
- Gieryn, Thomas F. 1983. "Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists." *American Sociological Review* 48 (6):781-795. doi: 10.2307/2095325.
- Glaser, April, and Will Oremus. 2018. "A Collective Aghastness": Why Silicon Valley workers are demanding their employers stop doing business with the Trump administration. *Slate*. <https://slate.com/technology/2018/06/the-tech-workers-coalition-explains-how-silicon-valley-employees-are-forcing-companies-to-stop-doing-business-with-trump.html>.
- Goldberg, Emma. 2020. 'Techlash' Hits College Campuses. *The New York Times*. <https://www.nytimes.com/2020/01/11/style/college-tech-recruiting.html>.

- Google Transparency Project. 2017. Google Academics Inc. <https://www.techtransparencyproject.org/sites/default/files/Google-Academics-Inc.pdf>.
- Graham, Jefferson. 2017. Hillary Clinton — tech has to fix fake news. *USA Today*. <https://www.usatoday.com/story/tech/talkingtech/2017/05/31/hrc-tech-has-fix-fake-news/102357904/>.
- Gray, Mary L., and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*: Houghton Mifflin Harcourt.
- Green, Ben. 2019. *The Smart Enough City: Putting Technology in Its Place to Reclaim Our Urban Future*: MIT Press.
- Green, Ben. 2020. "The False Promise of Risk Assessments: Epistemic Reform and the Limits of Fairness." Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain.
- Green, Ben, and Yiling Chen. 2019. "Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments." Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA.
- Green, Ben, and Salomé Viljoen. 2020. "Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought." Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain.
- Greene, Daniel, Anna Lauren Hoffmann, and Luke Stark. 2019. "Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning." Proceedings of the 52nd Hawaii International Conference on System Sciences.
- Grosz, Barbara J., David Gray Grant, Kate Vredenburg, Jeff Behrends, Lily Hu, Alison Simmons, and Jim Waldo. 2019. "Embedded EthiCS: Integrating Ethics Broadly Across Computer Science Education." *Communications of the ACM* 62 (8):54-61.
- Hamid, Sarah. 2020. "Community Defense: Sarah T. Hamid on Abolishing Carceral Technologies." *Logic Magazine*. <https://logicmag.io/care/community-defense-sarah-t-hamid-on-abolishing-carceral-technologies/>.
- Hao, Karen. 2020. We read the paper that forced Timnit Gebru out of Google. Here's what it says. *MIT Technology Review*. <https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/>.

- Haraway, Donna. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14 (3):575-599.
- Harwell, Drew. 2018. Facial recognition may be coming to a police body camera near you. *The Washington Post*. <https://www.washingtonpost.com/news/the-switch/wp/2018/04/26/facial-recognition-may-be-coming-to-a-police-body-camera-near-you/>.
- Haskins, Caroline. 2020. The Los Angeles Police Department Says It Is Dumping A Controversial Predictive Policing Tool. *BuzzFeed News*. <https://www.buzzfeednews.com/article/carolinehaskins1/los-angeles-police-department-dumping-predpol-predictive>.
- Hern, Alex. 2016. 'Partnership on AI' formed by Google, Facebook, Amazon, IBM and Microsoft. *The Guardian*. <https://www.theguardian.com/technology/2016/sep/28/google-facebook-amazon-ibm-microsoft-partnership-on-ai-tech-firms>.
- Hoffmann, Anna Lauren. 2019. "Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse." *Information, Communication & Society* 22 (7):900-915. doi: 10.1080/1369118X.2019.1573912.
- Hsu, Tiffany. 2018. For Many Facebook Users, a 'Last Straw' That Led Them to Quit. *The New York Times*. <https://www.nytimes.com/2018/03/21/technology/users-abandon-facebook.html>.
- IBM. 2018. Everyday Ethics for Artificial Intelligence. <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>.
- Integrated Innovation Strategy Promotion Council. 2019. AI for Everyone: People, Industries, Regions and Governments. <https://www8.cao.go.jp/cstp/english/humancentricai.pdf>.
- Irani, Lilly. 2019. <https://twitter.com/gleemie/status/1208793442509152258>.
- Irani, Lilly, and Rumman Chowdhury. 2019. "To Really 'Disrupt,' Tech Needs to Listen to Actual Researchers." *Wired*. <https://www.wired.com/story/tech-needs-to-listen-to-actual-researchers/>.
- Jasanoff, Sheila. 2004. "The idiom of co-production." In *States of Knowledge: The Co-Production of Science and the Social Order*, edited by Sheila Jasanoff, 1-12. Routledge.
- Jasanoff, Sheila. 2006. "Technology as a Site and Object of Politics." In *The Oxford Handbook of Contextual Political Analysis*, edited by Robert E. Goodin and Charles Tilly.

- Jasanoff, Sheila. 2016. *The Ethics of Invention: Technology and the Human Future*: W. W. Norton & Company.
- Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. "The global landscape of AI ethics guidelines." *Nature Machine Intelligence* 1 (9):389-399. doi: 10.1038/s42256-019-0088-2.
- Johnson, Deborah G., and Jameson M. Wetmore. 2007. "STS and Ethics: Implications for Engineering Ethics." In *The Handbook of Science and Technology Studies, Third Edition*, edited by Edward J. Hackett, Olga Amsterdamska, Michael E. Lynch and Judy Wajcman. MIT Press.
- Johnson, Khari. 2019. How to operationalize AI ethics. *VentureBeat*. <https://venturebeat.com/2019/10/07/how-to-operationalize-ai-ethics/>.
- Kang, Cecilia, Tiffany Hsu, Kevin Roose, Natasha Singer, and Matthew Rosenberg. 2018. Mark Zuckerberg Testimony: Day 2 Brings Tougher Questioning. *The New York Times*.
- Karoff, Paul. 2019. Embedding ethics in computer science curriculum. *The Harvard Gazette*. <https://news.harvard.edu/gazette/story/2019/01/harvard-works-to-embed-ethics-in-computer-science-curriculum/>.
- Khan, Lina M. 2017. "Amazon's Antitrust Paradox." *The Yale Law Journal* 126 (3):564-907.
- King, Leslie, and Julianne Busa. 2017. "When corporate actors take over the game: the corporatization of organic, recycling and breast cancer activism." *Social Movement Studies* 16 (5):549-563. doi: 10.1080/14742837.2017.1345304.
- Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. 2016. "Inherent trade-offs in the fair determination of risk scores." *arXiv preprint arXiv:1609.05807*.
- Knight, Will. 2019. Google appoints an "AI council" to head off controversy, but it proves controversial. *MIT Technology Review*. <https://www.technologyreview.com/2019/03/26/136376/google-appoints-an-ai-council-to-head-off-controversy-but-it-proves-controversial>.
- Lapowsky, Iessie. 2018. Facebook Exposed 87 Million Users to Cambridge Analytica. *Wired*. <https://www.wired.com/story/facebook-exposed-87-million-users-to-cambridge-analytica/>.

- Legassick, Sean, and Verity Harding. 2017. Why we launched DeepMind Ethics & Society. *DeepMind Blog*. <https://deepmind.com/blog/announcements/why-we-launched-deepmind-ethics-society>.
- Lewis, Tamika, Seeta Peña Gangadharan, Mariella Saba, and Tawanna Petty. (2018). *Digital Defense Playbook: Community Power Tools for Reclaiming Data*. Our Data Bodies.
- Marantz, Andrew. 2019. Silicon Valley's Crisis of Conscience. *The New Yorker*. <https://www.newyorker.com/magazine/2019/08/26/silicon-valleys-crisis-of-conscience>.
- Marowski, Steve. 2020. Artificial intelligence researchers create ethics center at University of Michigan. *MLive*. <https://www.mlive.com/news/ann-arbor/2020/01/artificial-intelligence-researchers-create-ethics-center-at-university-of-michigan.html>.
- Martinho-Truswell, Emma, Hannah Miller, Isak Nti Asare, André Petheram, Richard Stirling, Constanza Gómez Mont, and Cristina Martínez. 2018. Hacia una Estrategia de IA en México: Aprovechando la Revolución de la IA (Towards an AI Strategy in Mexico: Leveraging the AI Revolution). https://docs.wixstatic.com/ugd/7be025_ba24a518a53a4275af4d7ff63b4cf594.pdf.
- Marx, Leo, and Merritt Roe Smith. 1994. "Introduction." In *Does Technology Drive History?: The Dilemma of Technological Determinism*, edited by Merritt Roe Smith and Leo Marx. MIT Press.
- Mboya, Arwa. 2019. Why Joi Ito needs to resign. *The Tech*. <https://thetech.com/2019/08/29/joi-ito-needs-to-resign>.
- Metcalf, Jacob. 2014. Ethics Codes: History, Context, and Challenges. <https://bdes.datasociety.net/wp-content/uploads/2016/10/EthicsCodes.pdf>.
- Metcalf, Jacob, Emanuel Moss, and danah boyd. 2019. "Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics." *Social Research* 86 (2):449-476.
- Metzinger, Thomas. 2019. Ethics washing made in Europe. *Der Tagesspiegel*. <https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html>.
- Microsoft. 2018. Microsoft AI principles. <https://www.microsoft.com/en-us/ai/responsible-ai>.
- Mijente. 2019. 1,200+ Students at 17 Universities Launch Campaign Targeting Palantir. <https://notechforice.com/20190916-2/>.

- MIT Media Lab. 2017. MIT Media Lab to participate in \$27 million initiative on AI ethics and governance. *MIT News*. <https://news.mit.edu/2017/mit-media-lab-to-participate-in-ai-ethics-and-governance-initiative-0110>.
- MIT News Office. 2018. MIT reshapes itself to shape the future. *MIT News*. <http://news.mit.edu/2018/mit-reshapes-itself-stephen-schwarzman-college-of-computing-1015>.
- Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. "Model Cards for Model Reporting." Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA.
- Mittelstadt, Brent. 2019. "Principles alone cannot guarantee ethical AI." *Nature Machine Intelligence* 1 (11):501-507. doi: 10.1038/s42256-019-0114-4.
- Morozov, Evgeny. 2014. *To Save Everything, Click Here: The Folly of Technological Solutionism*. PublicAffairs.
- Moss, Emanuel, and Jacob Metcalf. 2020. Too Big a Word. *Data & Society: Points*. <https://points.datasociety.net/too-big-a-word-13e66e62a5bf>.
- Mozilla. 2018. Announcing a Competition for Ethics in Computer Science, with up to \$3.5 Million in Prizes. *The Mozilla Blog*. <https://blog.mozilla.org/blog/2018/10/10/announcing-a-competition-for-ethics-in-computer-science-with-up-to-3-5-million-in-prizes/>.
- Mozilla. 2020. <https://twitter.com/mozilla/status/1308542908291661824>.
- Nadella, Satya. 2018. Embracing our future: Intelligent Cloud and Intelligent Edge. *Microsoft News Center*. <https://news.microsoft.com/2018/03/29/satya-nadella-email-to-employees-embracing-our-future-intelligent-cloud-and-intelligent-edge/>.
- National Science and Technology Council. 2018. Preparing for the Future of Artificial Intelligence. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.
- Neate, Rupert. 2018. Twitter stock plunges 20% in wake of 1m user decline. *The Guardian*. <https://www.theguardian.com/technology/2018/jul/27/twitter-share-price-tumbles-after-it-loses-1m-users-in-three-months>.

- Nemitz, Paul. 2018. "Constitutional democracy and technology in the age of artificial intelligence." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376 (2133). doi: 10.1098/rsta.2018.0089.
- Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*: NYU Press.
- Noble, Safiya Umoja. 2019. <https://twitter.com/safiyanoble/status/1208812440403660800>.
- Novet, Jordan. 2018. Facebook forms a special ethics team to prevent bias in its A.I. software. *CNBC*. <https://www.cnbc.com/2018/05/03/facebook-ethics-team-prevents-bias-in-ai-software.html>.
- O'Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*: Broadway Books.
- Ochigame, Rodrigo. 2019. The Invention of "Ethical AI": How Big Tech Manipulates Academia to Avoid Regulation. *The Intercept*. <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/>.
- Organisation for Economic Co-operation and Development. 2019. Recommendation of the Council on Artificial Intelligence. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.
- Oxford Languages. 2018. Word of the Year 2018: Shortlist. *Oxford Languages*. <https://languages.oup.com/word-of-the-year/2018-shortlist/>.
- Oz, Effy. 1992. "Ethical Standards for Information Systems Professionals: A Case for a Unified Code." *MIS Quarterly*:423-433.
- Pardes, Arielle. 2018. Silicon Valley Writes a Playbook to Help Avert Ethical Disasters. *Wired*. <https://www.wired.com/story/ethical-os/>.
- Patil, D.J. 2018. A Code of Ethics for Data Science. <https://medium.com/@dpatil/a-code-of-ethics-for-data-science-cda27d1fac1>.
- Perloth, Nicole, Sheera Frenkel, and Scott Shane. 2018. Facebook Exit Hints at Dissent on Handling of Russian Trolls. *The New York Times*. <https://www.nytimes.com/2018/03/19/technology/facebook-alex-stamos.html?mtrref=undefined>.

- Peters, Adele. 2018. This tool lets you see—and correct—the bias in an algorithm. *Fast Company*. <https://www.fastcompany.com/40583554/this-tool-lets-you-see-and-correct-the-bias-in-an-algorithm>.
- Phillips, Matt. 2018. Facebook's Stock Plunge Shatters Faith in Tech Companies' Invincibility. *The New York Times*. <https://www.nytimes.com/2018/07/26/business/facebook-stock-earnings-call.html>.
- Pichai, Sundar. 2018. AI at Google: our principles. <https://www.blog.google/technology/ai/ai-principles/>.
- Pinch, Trevor J., and Wiebe E. Bijker. 1987. "The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other." In *The Social Construction of Technological Systems*, edited by Wiebe E. Bijker, Thomas P. Hughes and Trevor Pinch. MIT Press.
- Raicu, Irina. 2017. Rethinking Ethics Training in Silicon Valley. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2017/05/rethinking-ethics-training-in-silicon-valley/525456/>.
- Read, Max. 2016. Donald Trump Won Because of Facebook. *New York Magazine*. <https://nymag.com/intelligencer/2016/11/donald-trump-won-because-of-facebook.html>.
- Reardon, Jenny. 2011. "Human Population Genomics and the Dilemma of Difference." In *Reframing Rights: Bioconstitutionalism in the Genetic Age*, edited by Sheila Jasanoff, 217-238.
- Reardon, Jenny. 2013. "On the Emergence of Science and Justice." *Science, Technology, & Human Values* 38 (2):176-200. doi: 10.1177/0162243912473161.
- Reich, Rob, Mehran Sahami, Jeremy M. Weinstein, and Hilary Cohen. 2020. "Teaching Computer Ethics: A Deeply Multidisciplinary Approach." Proceedings of the 51st ACM Technical Symposium on Computer Science Education, Portland, OR, USA.
- Ritchie, Hannah. 2016. Read all about it: The biggest fake news stories of 2016. *CNBC*. <https://www.cnbc.com/2016/12/30/read-all-about-it-the-biggest-fake-news-stories-of-2016.html>.
- Robinson, David, and Miranda Bogen. 2016. Data Ethics: Investing Wisely in Data at Scale. *Upturn*. https://www.upturn.org/static/reports/2016/data-ethics/files/Upturn_-_Data%20Ethics_v.1.0.pdf.

- Rose, Kevin. 2019. The Making of a YouTube Radical. *The New York Times*. <https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html>.
- Rosenberg, Matthew, Nicholas Confessore, and Carole Cadwalladr. 2018. How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times*. <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>.
- Schneier, Bruce. 2016. *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*: W. W. Norton & Company.
- Seetharaman, Deepa. 2020. Jack Dorsey's Push to Clean Up Twitter Stalls, Researchers Say. *The Wall Street Journal*. <https://www.wsj.com/articles/jack-dorseys-push-to-clean-up-twitter-stalls-researchers-say-11584264600>.
- Selznick, Philip. 1948. "Foundations of the Theory of Organization." *American Sociological Review* 13 (1):25-35. doi: 10.2307/2086752.
- Shane, Scott, and Daisuke Wakabayashi. 2018. 'The Business of War': Google Employees Protest Work for the Pentagon. *The New York Times*. <https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html>.
- Sharlach, Molly. 2019. Princeton collaboration brings new insights to the ethics of artificial intelligence. <https://www.princeton.edu/news/2019/01/14/princeton-collaboration-brings-new-insights-ethics-artificial-intelligence>.
- Shilton, Katie, Michael Zimmer, Casey Fiesler, Arvind Narayanan, Jake Metcalf, Matthew Bietz, and Jessica Vitak. 2017. We're Awake — But We're Not At the Wheel. *PERVADE: Pervasive Data Ethics*. <https://medium.com/pervade-team/were-awake-but-we-re-not-at-the-wheel-7f0a7193e9d5>.
- Silbey, Susan S. 2018. How Not to Teach Ethics. *MIT Faculty Newsletter*. <https://web.mit.edu/fnl/volume/311/silbey.html>.
- Simonite, Tom. 2018. Should Data Scientists Adhere to a Hippocratic Oath? *Wired*. <https://www.wired.com/story/should-data-scientists-adhere-to-a-hippocratic-oath/>.
- Simonite, Tom. 2019. Google and Microsoft Warn That AI May Do Dumb Things. *Wired*. <https://www.wired.com/story/google-microsoft-warn-ai-may-do-dumb-things/>.

- Simonite, Tom. 2020. Google Offers to Help Others With the Tricky Ethics of AI. *Wired*. <https://www.wired.com/story/google-help-others-tricky-ethics-ai/>.
- Sinders, Caroline. 2019. <https://twitter.com/carolinesinders/status/1208443559998873601>.
- Smart Dubai. 2018. AI Ethics Principles & Guidelines. https://www.smartdubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf?sfvrsn=d4184f8d_6.
- Smith, Brad. 2018. Facial recognition: It's time for action. *Microsoft On The Issues*. <https://blogs.microsoft.com/on-the-issues/2018/12/06/facial-recognition-its-time-for-action/>.
- Solon, Olivia. 2016. Facebook's failure: did fake news and polarized politics get Trump elected? *The Guardian*. <https://www.theguardian.com/technology/2016/nov/10/facebook-fake-news-election-conspiracy-theories>.
- Solon, Olivia. 2019. Why did Microsoft fund an Israeli firm that surveils West Bank Palestinians? *NBC News*. <https://www.nbcnews.com/news/all/why-did-microsoft-fund-israeli-firm-surveils-west-bank-palestinians-n1072116>.
- Stark, Luke, and Anna Lauren Hoffmann. 2019. "Data Is The New What?: Popular Metaphors & Professional Ethics in Emerging Data Cultures." *Journal of Cultural Analytics*. doi: 10.22148/16.036.
- Stop LAPD Spying Coalition and Free Radicals. 2020. The Algorithmic Ecology: An Abolitionist Tool for Organizing Against Algorithms. <https://medium.com/@stoplapdspying/the-algorithmic-ecology-an-abolitionist-tool-for-organizing-against-algorithms-14fcbd0e64d0>.
- Suchman, Lucy, Jeanette Blomberg, Julian E. Orr, and Randall Trigg. 1999. "Reconstructing Technologies as Social Practice." *American Behavioral Scientist* 43 (3):392-408. doi: 10.1177/00027649921955335.
- The Institute for the Future, and Omidyar Network. 2018. Ethical OS Toolkit. <https://ethicalos.org>.
- Tiku, Nitasha. 2019. Three Years of Misery Inside Google, the Happiest Company in Tech. *Wired*. <https://www.wired.com/story/inside-google-three-years-misery-happiest-company-tech/>.
- Toyama, Kentaro. 2015. *Geek Heresy: Rescuing Social Change from the Cult of Technology*: PublicAffairs.
- Tracy, Marc, and Tiffany Hsu. 2019. Director of M.I.T.'s Media Lab Resigns After Taking Money From Jeffrey Epstein. *The New York Times*.

- <https://www.nytimes.com/2019/09/07/business/mit-media-lab-jeffrey-epstein-joichi-ito.html>.
- Trumpy, Alexa J. 2014. "Subject to Negotiation: The Mechanisms Behind Co-Optation and Corporate Reform." *Social Problems* 55 (4):480-500. doi: 10.1525/sp.2008.55.4.480.
- U.S. Department of Defense. 2020. DOD Adopts Ethical Principles for Artificial Intelligence. <https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/>.
- Varshney, Kush R. 2018. Introducing AI Fairness 360. *IBM Research Blog*. <https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/>.
- Vincent, James. 2016. Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge*. <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>.
- Vincent, James, and Russell Brandom. 2018. Axon launches AI ethics board to study the dangers of facial recognition. *The Verge*. <https://www.theverge.com/2018/4/26/17285034/axon-ai-ethics-board-facial-recognition-racial-bias>.
- Visvanathan, Shiv. 2005. "Knowledge, justice and democracy." In *Science and Citizens: Globalization and the Challenge of Engagement.*, edited by Melissa Leach, Ian Scoones and Brian Wynne. Zed Books.
- Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The spread of true and false news online." *Science* 359 (6380):1146-1151. doi: 10.1126/science.aap9559.
- Wagner, Ben. 2018. "Ethics as Escape From Regulation: From Ethics-Washing to Ethics-Shopping?" In *Being Profiling. Cogitas Ergo Sum*, edited by Emre Bayamlioglu, Irina Baraliuc, Liisa Albertha Wilhelmina Janssens and Mireille Hildebrandt. Amsterdam University Press.
- Walker, Kent. 2018. Google AI Principles updates, six months in. *The Keyword*. <https://www.blog.google/technology/ai/google-ai-principles-updates-six-months/>.
- West, Sarah Myers, Meredith Whittaker, and Kate Crawford. 2019. Discriminating Systems: Gender, Race, and Power in AI. <https://ainowinstitute.org/discriminatingystems.pdf>.
- Williams, Oscar. 2019. How Big Tech funds the debate on AI ethics. *New Statesman*. <https://www.newstatesman.com/science-tech/technology/2019/06/how-big-tech-funds-debate-ai-ethics>.

- Winner, Langdon. 1986. *The Whale and the Reactor: A Search for Limits in an Age of High Technology*: University of Chicago Press.
- Wood, Greg, and Malcolm Rimmer. 2003. "Codes of Ethics: What Are They Really and What Should They Be?" *International Journal of Value-Based Management* 16 (2):181-195.
- Woodman, Spencer. 2017. Palantir Provides the Engine for Donald Trump's Deportation Machine. *The Intercept*. <https://theintercept.com/2017/03/02/palantir-provides-the-engine-for-donald-trumps-deportation-machine/>.
- Wright, Susan. 2001. "Legitimizing Genetic Engineering." *Perspectives in Biology and Medicine* 44 (2):235-247.
- Wu, Tim. 2018. *The Curse of Bigness: Antitrust in the New Gilded Age*: Columbia Global Reports.
- Zuboff, Shoshana. 2018. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*: PublicAffairs.
- Zunger, Yonatan. 2018. Computer science faces an ethics crisis. The Cambridge Analytica scandal proves it. *The Boston Globe*. <https://www.bostonglobe.com/ideas/2018/03/22/computer-science-faces-ethics-crisis-the-cambridge-analytica-scandal-proves/IzaXxl2BsYBtwM4nxezgcP/story.html>.