ETHICS OWNERS

A New Model of Organizational Responsibility in Data-Driven Technology Companies

DATA&
SOCIETY

Emanuel Moss
Jacob Metcalf

September 2020

Author: **Emanuel Moss,**
doctoral candidate in Anthropology,
CUNY Graduate Center, New York.

Author: **Jacob Metcalf,**
PhD, 2010, Philosophy,
University of California, Santa Cruz.

PREFACE

The world has moved in remarkable ways since this research was first initiated in late 2018 and the analysis was finalized in early-2020. Since then, we have seen the COVID-19 pandemic and the resurgence of the #BlackLivesMatter movement following the police killings of George Floyd, Breonna Taylor, and all too many others. Given these critical events, we as researchers and as members of a research organization paused to reflect on how our work speaks to the urgent injustices laid bare today.

These events have dramatically foregrounded existing calls for justice at the intersections of technology, unjust social and political structures, digital privacy, surveillance, and the values and purpose of technology corporations. Many of the defining events that have shaped ethics in the tech sector in recent years—including tech worker organization, walkouts and resignations over military contracts, continued contestations over racial and sexual harassment inside of tech companies, legislation and regulations, and critical audits—have been and continue to be the contentious background to the research conducted for this report. Our research takes place amidst, but is not primarily about, these tectonic changes that have repeatedly reframed the broader tech ethics conversation. It is instead about the ways those inside of tech companies have begun reshaping corporate practices against this background—how they understand the problems their industry faces, the means at their disposal to address these problems, and the lines of thought that shape those understandings.

Public debates about social justice remain in the background of this research. Ethics practices, as they have been institutionalized in Silicon Valley tech companies, are often focused on addressing corporate values, legal risk, and compliance while bracketing out broader understandings of what is meant by "ethics." These broader understandings include questions of moral, social, and racial justice that corporate ethics practices are more likely to address in incremental, and not structural or systematic, ways. As we have

discussed elsewhere[1], questions of justice have long been pushed out of the corporate tech ethics conversation for complex reasons that are not reducible to a simple focus on profit and market dominance.

The exclusion of justice from these conversations has much to do with the social and political position of who gets to define what ethics *is*. As we note in the report, our interviewees were almost exclusively white, and the rooms in which "ethics owners" held the workshops and informal conversations we researched were similarly white—and as white, male researchers we benefited from privileged access to these rooms, too. The topic of race came up only in passing. The absence of discourse about race and racial equity is itself a theme that runs through our corpus of interviews and observations. Recent work has cast a spotlight on the ways Silicon Valley technologies embed anti-Blackness through products that extend carceral logics that surveil and police Black people. At the same time, these technologies operate within parameters that assume a universal, raceless default subject position synonymous with whiteness. This creates a feedback loop that undermines efforts to center Black and other vulnerable communities in the tech ethics discourse.[2]

The lack of voices from Black, Indigenous, and people of color leads us to emphasize that the essential critical voices on justice and technology hold different, and often more marginalized, forms of social power than those who have to date been tasked with implementing changes in the industry. This must be understood as an untenable state, as it is precisely these voices that are best situated to speak to the risks technology products pose. While our data provides evidence of the absence of these voices, it primarily focuses on the ways ethics owners have adapted to their task, as they understand it, which heightens the importance of what is missing. Building from this, we believe that there are important

1 https://points.datasociety.net/too-big-a-word-13e66e62a5bf

2 Benjamin, Ruha, ed. *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life.* Durham, NC: Duke University Press, 2019.; Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code.* Medford, MA: Polity Press, 2019.; Browne, Simone. *Dark Matters: On the Surveillance of Blackness*. Durham: Duke University Press, 2015.; Tu, Thuy Linh N. and Nelson, Alondra, eds. *Technicolor: Race, Technology, and Everyday Life*. New York: NYU Press, 2001.

lessons and methods for tech ethicists seeking opportunities to be anti-racist in their work to be found in this research (which is discussed in more depth in a blog post published alongside this report).

No matter what form the lurching, fractional changes the tech industry takes toward ethical and anti-racist technology development, they will only be sustained if they become embedded in organizational practices. This research points to how the much-needed changes have been, and currently are being, embedded in such practices. From this point forward, there will inevitably be new roles, new infrastructures, and new types of governance that emerge, but the measure of their success will be the extent to which they dismantle injustice and oppression, particularly inside the very rooms where critical voices are most needed.[3]

3 https://points.datasociety.net/looking-for-race-in-tech-company-ethics-956919fe48ee

CONTENTS

EXECUTIVE SUMMARY

This report presents findings from interviews and ethnographic observation of an emerging field of Silicon Valley professionals tasked with managing the ethical footprints of tech companies. We call these roles "**ethics owners**," because they handle challenging ethical dilemmas with the tools of tech management and within familiar tech company structures, translating public pressure into new corporate practices. Owning ethics for an entire company is a challenging job, and understanding the precursors, assumptions, and organizational limitations that ethics owners face is key to making real changes to tech's impact on the world.

Ultimately, we find that ethics owners' primary work is navigating rather than resolving tensions. That is, there will always be conflicts between, say, individual and corporate responsibilities, or quantitative and qualitative "measures" of ethics. And across each of these tensions, the common challenge reported by ethics owners is *scale*. Whether balancing the scale of an entire, multi-department corporation, or wrestling with the scale of a product with millions of users, ethics owners report again and again that older tools for ethical oversight are blunted by just how big the big questions can get. Therefore, what is crucial are ways of expanding the industry's capacity to recognize problems—both long-standing and emerging—through new forms of cross-company collaboration, increased capacity to include perspectives from beyond the current configuration of Silicon Valley employees, and a commitment to well-documented, context-specific case studies.

Prior Frameworks and Practices

The ethics owners we spoke with identified a host of existing approaches to evaluating and operationalizing "ethics," each with its own strengths and weaknesses. Current ethics owners draw upon earlier ethical frameworks, including medical research ethics, business ethics, and professional ethics. While each of these are

valuable models, none fully address the dilemma currently facing ethics owners. Similarly, ethics owners draw upon existing practices for organizational ethics within Silicon Valley, such as corporate statements of principle, ethical review boards, or software development lifecycle tools. Still, none of these are sufficient to the nature and scale of the problem.

Navigating Tensions

By extrapolating from our fieldwork, we identify six fundamental tensions in the work of ethics owners:

- **Personal ethics, corporate duties.** As employees, ethics owners always have both personal ethics and corporate duties, which often conflict.
- **Upside benefits, downside risks.** Ethics owners find that it is far easier to advocate for changes that limit potentially harmful effects of a product than it is to advocate for changes that may prove beneficial for users despite agreement that such upsides exist.
- **Direct-to-consumer, business-to-business.** For ethics owners, how they approach their work will be deeply affected by their employer's business model, and standard practices might not work across different business models.
- **Measurable and nonmeasurable impact.** Developing metrics for measuring the ethical impact of tech products, and tracking the success of interventions, remains one of the greatest challenges for ethics owners.
- **Users, nonusers.** Ethics owners are given the task of understanding broader social impact of their companies' products, but have little insight into how nonusers are affected by those products.
- **Specificity, generalizability.** Like the rest of Silicon Valley, ethics owners face significant challenges when attempting to scale their methods across contexts.

Opportunities for Ethics Owners

Despite the above challenges, we identify several strategies that expand the capacity and depth of corporate practices around the issues with which ethics owners are tasked. Some of these issues will require shifting who has power within these settings and who is represented on teams lead by ethics owners. At the same time, because of how ethics owners are currently positioned within their organizations, and within the industry, they have a significant amount of leverage to:

- **Publish public case studies.** This report recommends that ethics owners develop a robust set of useful case studies. This would challenge existing ways of "doing" ethics inside of tech companies. These case studies should be grounded in particularity but offer lessons that can be used across the industry.

- **Talk about failures.** Additionally, the case studies should include instances of failure, as well as success. Because corporate communications are far more likely to relate successful ethics initiatives to the public, ethics owners should convene meetings where they can discuss less successful initiatives, as well as instances when they have forestalled serious ethical harm that otherwise does not enter into public discourse.

- **Support colleagues in civil society.** Ethics owners should also try to support the work of those in civil society, outside the corporate world, who bear much of the burden for identifying and mitigating the unanticipated harm produced by technology. Ethics owners can support this work by attending to the types of information that outside groups need to be effective advocates for the public.

- **Highlight possibilities of social benefit and justice.** Lastly, ethics owners can be points of leverage within companies for the forms of accountability that go beyond the metric-driven evaluation of risk and harm that characterizes much of the tech ethics conversation. They can do this by foregrounding the possibilities of social benefit and a broader understanding of justice in their work.

INTRODUCTION

"[Ethics is] fundamentally a risk management process."

In the past few years, Silicon Valley companies have become increasingly concerned about a public that is growing wary of their impact on the world, despite the otherwise benevolent cloak the tech industry has drawn around itself. At the same time, the tech industry workforce has increasingly begun to feel its own power as (informally, so far) organized, highly paid, relatively scarce labor to steer the direction of some of the companies they work within.[1]

As a result, many of the corporate giants of Silicon Valley have created new, amorphous roles inside their ranks tasked with responsibility over the ethical footprint of these companies. These are professionals from a range of backgrounds, placed at various levels, across myriad teams, and tasked with everything from crafting policy to managing products. We call these people **ethics owners**. They go by different job titles and occupy a range of positions, and there is little consistency across the industry for these relatively new roles. None refer to themselves as "ethics owners," but most claim expertise in AI ethics, data ethics, research ethics, or organizational ethics.[2] In corporate parlance, someone who has responsibility for a product, process, or initiative *across an organization* is typically said to "own" that responsibility. This is particularly true of the flat organizational structures that pervade Silicon Valley, where even as companies grow from startup to mature organization, there remain few layers to the organizational chart, and where one employee might hold responsibilities across several departments at once.[3] Tech companies therefore look to ethics owners, whatever their formal title, to coordinate the way practices are distributed across the organization from a discrete, legible position.

1 Nataliya Nedzhvetskaya and J. S. Tan, "What We Learned from over a Decade of Tech Activism," *The Guardian*, December 22, 2019, https://www.theguardian.com/commentisfree/2019/dec/22/tech-worker-activism-2019-what-we-learned and RK Upadhya, "Tech Workers Against Imperialism," *Tech Workers Coalition*, November 21, 2018, https://medium.com/tech-workers-coalition/tech-workers-against-imperialism-2d8024e461a7.

2 These domains are related to, but distinct from, existing ethical domains like business ethics and compliance, corporate social responsibility, or research ethics.

3 Homa Bahrami and Stuart Evans, "Super-Flexibility for Real-Time Adaptation: Perspectives from Silicon Valley," *California Management Review* 53, 3 (2011): 21–39.

From our research, conducted with ethics owners in Silicon Valley, we identify how they are tasked with turning "external" pressure—from the public, from advocacy groups and critics, and from their own activist employees—into practices compatible with day-to-day operations. In translating outside pressure into internal practices, some ethics owners use existing organizational values to intervene in product design. They may also leverage external sources (e.g., human rights frameworks) or internal guidelines (e.g., corporate statements of principles or codes of conduct).[4] They might "officially" work in one division like "legal" or "policy" while being tasked with shifting practices in "product" or "quality assurance." Therefore, ethics owners predominantly use the existing tools of those divisions—policy documents for policy teams, product requirements documents for products teams, metadata tracking tools for data teams, and corporate communications for marketing teams—in order to coordinate ethics capacity across those divisions.

This report draws on two dozen interviews with tech industry employees to identify how ethics owners and their practices have emerged within Silicon Valley. Between December 2018 and April 2019 we interviewed team leaders and other principals within large (1,000+ employees) companies with headquarters or significant footprints in Silicon Valley. We also interviewed those at smaller companies, as well as outside activists from advocacy organizations, and several outspoken tech workers who have been critics of Silicon Valley. The titles of our interviewees reflect the lack of consistency in the field. These titles were often variations of "ethics officer," "privacy and rights lead," or "responsible AI lead," but we also spoke with a "director of data science research" who had become an ethics owner, as well as an "AI for good lead" and an "executive director." Among the ethics owners we interviewed, many have technical backgrounds in engineering, computer science, and user experience/user interface design, although several also hold law degrees and MBAs.

4 Christiaan van Veen, "Artifical Intelligence: What's human rights got to do with it?" *Data & Society Points (blog)*, May 18, 2018. https://points.datasociety.net/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d5; The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems," "Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems, First Edition," IEEE, 2019.

Others have backgrounds in political science, physics, communications, or philosophy. Depending on how their role is designed within their particular company, they may report directly to their CEO or their board, but most ethics owners report to heads of legal, product, policy, or marketing.

Our interviews were informed by embedded ethnographic work inside tech companies, attending many of the informal meetings, workshops, and other gatherings for ethics professionals working inside tech companies. We have also closely tracked and analyzed the extensive public documents discussing tech ethics, consisting of corporate communications explaining ethical decision-making about particular products, journalistic accounts of the unfolding practices of tech ethics, and public debates about the ethical implications of technology and data-driven development.

The emergence of this new role for technology ethics practitioners is happening amid a ferment of conversations about what "ethics" means in this space. In our interviews and observations, we regularly heard the demands of the current moment compared to familiar practical ethics resources and histories from both corporate and academic spaces. We heard repeated mentions of the need for institutional review boards (IRBs) for corporate research, but also disdain for the IRBs found in universities. Multiple interview subjects asserted their work did *not* duplicate already-existing corporate roles which contain the title "ethics," particularly compliance and ethics officers located in human resources or legal departments whose roles address issues like whistleblowing and harassment. We heard multiple interviewees state, as if it were a consensus, that the challenges under consideration could not be appropriately addressed by existing corporate social responsibility (CSR) offices, which manage the corporation's contributions to the public good through philanthropy and community outreach. Thus, at least in the eyes of those tasked with marshalling a response, there is clearly a dynamic at play in the tech industry—as demonstrated by the relentless headlines about "ethics" and the material investments in ethics resources in response—that is not directly addressable using established "ethics" resources.

"If you don't establish the lines and the frame in which you feel you wanna operate, you can always manipulate a process to bend one way or another, so you have to sort of establish that ahead of time. It's like a precursor."

However they are positioned, the ethics owners we spoke with are meaningfully distinct from the traditional corporate positions that handle ethical issues. The ethics owner differs from the familiar corporate roles that *deflect pressure* (public relations), *demonstrate compliance* with laws (business ethics, compliance and whistleblowing), and attempt to *distribute social goods* (CSR). Rather, ethics owners *adapt to pressure*, *facilitate compliance* with informal and evolving standards, *prepare for potential future regulations*, and attempt to *prevent social harm*.

This distinct role also means that ethics owners are now much more closely involved in the production of technology products and services than were those in previous ethics roles. And yet their role is characterized by a set of tensions between the business of generating profitable products and ethical responsibilities to the public. As we demonstrate below, understanding how the people occupying these new roles navigate the tensions they face—between the claims they make about the public interest and the existing organizational and cultural norms of Silicon Valley—is critical to identifying tractable forms of governance over these increasingly powerful technologies and the companies that build and deploy them.

A Reader's Guide

This report was written with a wide-ranging audience in mind. If you are just interested in what ethics owners can do *today*, flip to "Opportunities for Ethics Owners." If you are interested in the specific challenges facing ethics owners, and what they can do about those challenges in their current roles, read "The Challenges of Owning Ethics" and then feel free to skip ahead to the sections on "Navigating Tensions" and "Opportunities for Ethics Owners." If you are particularly interested in prior approaches to organizational ethics, you may want to focus on "Ethics in Industry: Existing Frameworks and Methods." And if you have a general interest in the topic of ethics in the tech industry, you will find it worthwhile to read this report from top to bottom, including the (admittedly verbose) footnotes.

THE CHALLENGES OF OWNING ETHICS

"As we start to talk about ethics in companies, they start to look more like government."

Across Silicon Valley tech companies, upper management is struggling to fit the responsibility to oversee "ethics" into a job description. These new ethics owner roles are being shaped based on existing ways of organizing workers inside Silicon Valley companies. In their work, ethics owners use their companies' existing tools and practices, and often draw haltingly on existing corporate approaches to addressing ethical concerns. Despite seemingly common goals, each company has come up with its own approach to "owning ethics," and the result is a bespoke set of duties for every ethics owner we interviewed. Each has a unique position and source of authority within their organization with little standardization across the industry. As a consequence, the responsibilities of an ethics owner do not fit neatly into the space where industry and ethics have intersected in the past, although they do draw on historical precedents.

Nor does the overall mission of ethics owners fit neatly into the standard practices of Silicon Valley tech companies, even if many of their day-to-day activities would seem familiar to almost anyone working in the Valley. Ethics owners we spoke with will, on any given day, find themselves holding open office hours for engineers, giving TED-style talks at an industry conference, attending a Chatham House Rules meeting with ethics owners from other companies, running a design-thinking event and wrangling a whiteboard full of sticky-notes, meeting with the general counsel, drafting an academic paper on tech ethics, or being pulled into a product team meeting. The same might be said of anyone in a managerial role inside these companies.

But for ethics owners, their core role is to develop processes, policies, and tools for addressing ethical concerns within the everyday structures of corporate work. Ethics owners interrogate product designs by asking: Will a product exclude a particular racial group from services provided on their platform? Could a product be used to deprive citizens of their human rights under an authoritarian regime? How should a private company balance the liberatory potential of their communications technology with the public understanding of free speech or the possibility of harassment at scale?

Ethics owners interrogate their own work by asking logistical questions such as: Should the software development life cycle management tools be updated? Ought a "red team" be convened to test the ethical implications of certain products, similarly to how security risks are evaluated?[5] How can their organizations incentivize ethical practices by modifying personnel evaluations and bonuses? Does a product need an ethical review before it goes into production? When in the development process should that review happen? And if so, who should be involved?

However, there are several challenges to this approach. One problem is simply deciding where their work should be positioned, organizationally. Many ethics owners we interviewed described the complications of interacting with roles that are similar to, but not the same as, ethics work. They listed roles in legal, safety, data or AI "for good," content moderation, CSR, policy, product design, and research all as having overlapping duties and responsibilities to their own. These overlapping responsibilities highlight the importance of the coordinating role ethics owners are meant to play. In addition to translating outside pressure demanding that companies "do the right thing" into concrete practices, ethics owners must also translate the various forms of expertise that sit adjacent to their own into practices that are legible across departments. They might find themselves translating between a privacy team and a trust and safety team, or addressing an ethical challenge by updating a content moderation policy.

Most of the ethics owners we spoke with were considered members of either the product or legal teams. However, such teams work very differently. Engineers on product teams often struggle to operationalize ethics problems as technical solutions in ways that satisfy the relevant parties. The kinds of problems that are classified as AI/ML ethics problems are often inextricably social, and therefore

5 Jeyavijayan Rajendran, Vinayaka Jyothi, and Ramesh Karri, "Blue team red team approach to hardward trust assessment," *2011 IEEE 29th International Conference on Computer Design (ICCD)*, 285–88, https://ieeexplore.ieee.org/document/6081410/; Andy Greenberg, "Security Isn't Enough. Silicon Valley Needs 'Abusability' Testing," *Wired*, January 28, 2019, https://www.wired.com/story/abusability-testing-ashkan-soltani/.

"Ethics can't live on paper. They live inside people."

may not be "solved" like a traditional software engineering or design problem. Engineers can see such issues as tangential to their core responsibilities. This perception is also reinforced by the use of objective and key results (OKRs), a ubiquitous corporate technique for setting and monitoring productivity goals of individuals or teams, which are difficult to apply to nuanced ethical concerns.[6]

In his book "Measure What Matters," John Doerr gives the example of a key result for a software engineer tasked with developing "three new features" with "fewer than five bugs per feature in quality assurance testing."[7] OKRs, crucially, are intended to be tools to evaluate performance against a measurable result. In many ways, the managerial toolkit of Silicon Valley is focused on tangible outcomes that can be tracked and evaluated at scale, across a business enterprise that may quite literally span the globe. Ethics owners consistently described the challenges they face in trying to integrate ethical objectives into OKRs, largely due to the difficulty of *measuring* ethics, a problem that increases as companies scale their products and services.

On the flip side, many legal teams may feel that the interventions proposed by ethics owners do not meet strict legal tests (such as liability), and so may see them as irrelevant to their own work. Ethics problems are fundamentally about human values: What kind of world do we prefer? Whose interests are served when there is

6 OKRs are a management tool that has gained broad adoption across Silicon Valley for setting goals and evaluating performance across an organization, and tend to be focused on tangible outcomes.

7 John Doerr, *Measure What Matters: How Google, Bono, and the Gates Foundation Rock the World with OKRs* (New York: Portfolio/Penguin, 2018).

a conflict? How should we balance competing rights claims? These types of questions are not directly translatable to the compliance frameworks that are predominant in corporate law. In some cases, the ethical outcome may not be a legally viable one, such as where meeting a demonstrable measure of algorithmic fairness would require collecting sensitive personal data without user consent. Despite the challenges of seating ethics owners in legal roles, we also heard that their internal initiatives often piggyback on the technical and organizational changes necessary to comply with major regulatory requirements, particularly the *General Data Protection Regulation (GDPR)* in Europe and consent decrees with the Federal Trade Commission in the U.S. For example, the metadata tagging processes required for GDPR compliance have dramatically reduced the engineering workload for developing ethics-related metadata tagging in machine learning technical systems.

The ethics owners we spoke with explained the challenges of how their placement is determined within their companies' organizational structure. Placement within a corporate hierarchy implies the relative status their role holds, but also has practical consequences. For instance, ethics owners and their supervisors have to figure out, without precedents, if they should report directly to a C-suite executive. And if so, should they report to an executive with internally facing responsibilities, like a chief technology officer or a chief information officer, to someone with external responsibilities, like a CEO or even the chief legal counsel. Yet again, the decision between product and legal teams is fraught. In the former, they might have a more active role in making ethical design choices but be just one voice among many, subject to the same kinds of performance evaluations as engineers and product managers. In the latter, they might be able to make clear policy decisions, but run the risk of turning into, as one ethics owner put it, a "Dr. No" who only stands in the way of ideas and whom potential collaborators within the company tactically avoid. As another ethics owner observed, "As soon as you're a compliance function, you're a second-class citizen and you're annoying."

A complementary problem to the matter of where the ethics owner should sit is that "ethics" is not any one thing. What types of issues count as "ethics" can include very different things to different people,

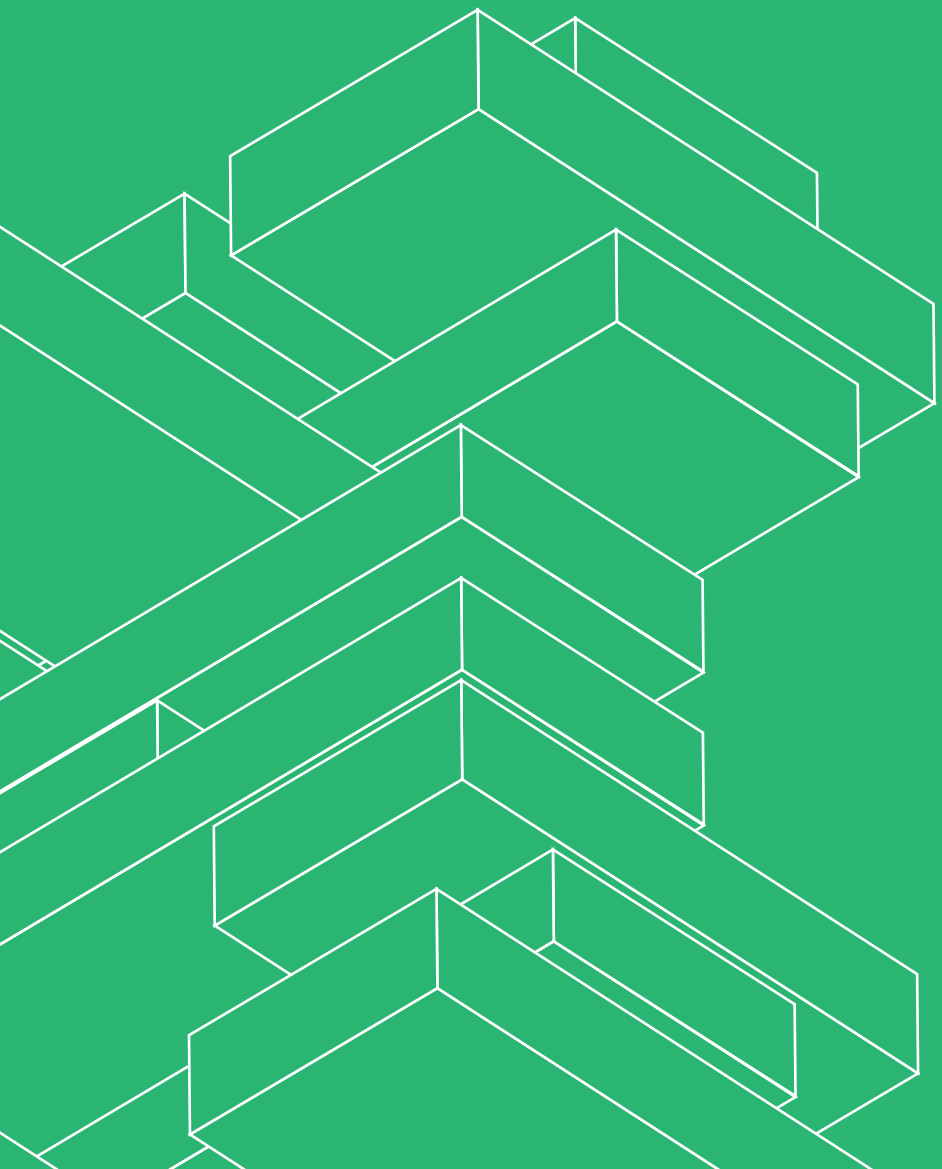
and thereby invoke a range of organizational resources. One ethics owner described a discussion about client work, and whether or not it was aligned with their organization's goals. In deliberating about whether to take on a controversial client, the decision came down to whether the company could build systems for the client in ways that guaranteed their client couldn't use their products for contentious purposes. Such a question invokes matters typically handled by technical teams, marketing and sales, legal and contracting, policy, and public relations. Beyond that, the ethics owner had to weigh the risk of their organization being seen as working with such a client. They explained that this particular conflict was not really about "the ethical—it was more about the mechanics and the politics." That "the political" can be neatly separated from "the ethical" for one ethics owner is indicative of how contested the application of ethics is for all ethics owners—what is and what is not "ethics" remains an active debate. Because of this, it can be difficult for ethics owners to synthesize a consistent set of organizational values, let alone to place their work into predictable, replicable institutional practices.

The positioning of ethics owners is further complicated by the corporate growth patterns in tech companies. Because of scrappy origin stories and pressure from outside investors for exponential, "hockey-curve" growth and an eventual public offering, tech companies tend to focus on market strategy and minimum viable products before they address questions of ethics. When younger companies are going through their initial growth phases, they emphasize versatility; stories abound of startups in which the CEO also had to schedule all the meetings, or in which the most senior engineers were also pressed into duty as salespeople. It was not uncommon for us to talk to ethics owners who had previously been coders or played quality assurance roles, sometimes simultaneously, before their company was mature enough to carve out a specific set of responsibilities for them around an ethics portfolio. Ethics owners we interviewed often told us that it is often only after a company's long-term viability is somewhat assured, and pressure from investors shifts in its tenor and intensity toward long-term reputational risk and addressing the consequences of scale, that their company began to think about ethics. Once companies are profitable enough to face real financial losses, working on ethics becomes a more tractable value

proposition—one ethics owner explained, "ethics … never makes you money, but ethics can save you a lot of money." By then, other roles are well defined and ethics owners must yet find a place within the organizational chart.

In developing these strategies within mature companies, ethics owners are attempting to institute changes across an entire organization. The ethics owners we spoke with emphasized how important it is that senior leadership is fully bought-into their work and willing to make organizational commitments to support that work. "There has to be real and genuine interest on the part of the leadership. If there's not, it's like taking a gashing wound and putting a few Band-Aids on it," explained one ethics owner.

ETHICS IN INDUSTRY: EXISTING FRAMEWORKS AND METHODS



The core work of ethics owners is to transform outside pressure on Silicon Valley companies to act ethically into internal practices that comport with how business is done inside the industry. For ethics owners, this transformation involves translating a broad set of ethical frameworks and statements about corporate values and principles into concrete professional practices. Ethical and values frameworks might be drawn from various sources, including human rights frameworks, privacy principles, biomedical research guidelines, and corporate principles statements. In doing this translational work, ethics owners draw on several prior models for engaging with ethics within professional settings. From our interviews, we identify several key existing frameworks for how ethics owners think about their roles. Similarly, we identify a set of existing organizational methods that characterize much of the product development work done inside Silicon Valley, which ethics owners are adapting to their work.

Existing Ethics Frameworks

Ethics owners, as part of their translation work, are modifying existing practices inside of Silicon Valley in the service of ethics. But they are also drawing on existing models for "doing ethics" in institutional and organizational contexts. The history of these approaches—most prominently, research ethics for human subjects, business ethics and CSR, and professional ethics—has a wide range of implications for what constitutes ethics inside of the technology industry today, including a focus on matters of informed consent, expert review, and documentation. Still, none of these precedents are entirely adequate for the challenges of "owning" ethics today.

Medical Research Ethics

Medical and research ethics are some of the oldest and most standardized forms of professional ethics, and were regularly invoked by the ethics owners whom we interviewed. However, while there are some important resonances between medical and research ethics and the contemporary needs of the tech industry, there are also significant disjunctions. For example, the Hippocratic Oath, most often glossed as "First, do no harm," is commonly administered to new physicians

"Codes of ethics are helpful guidance for individual practitioners, but do not directly speak to the needs of organizations."

as part of their graduation and licensing requirements, requiring that physicians place the moral interests of patients ahead of their own.[8] In our research, we have regularly observed calls for the adoption of a "Hippocratic Oath for data scientists."[9] However, the Hippocratic Oath is focused on the intimate relationship between a physician and an individual patient, something that is not analogous to the relationship between product developers and many millions or billions of users.[10] As is so often the case for ethics inside Silicon Valley, this difference in *scale* is the defining problem: our most prominent examples of applied professional ethics don't speak to how we can apply ethical reasoning to the interests of millions or billions of people.

Although medical ethics covers a huge swath of topics, most regulatory attention is found where medicine overlaps with scientific research practices. Research ethics addresses the potential moral risks involved in conducting scientific research on human beings. It has spawned one

8 Raphael Hulkower, "The history of the Hippocratic Oath: outdated, inauthentic, and yet still relevant," *Einstein Journal of Biology and Medicine* 25.1 (2016): 41-44.

9 Tom Simonite, "Should data scientists adhere to a Hippocratic Oath ?" *Wired,* February 8, 2018, https://www.wired.com/story/should-data-scientists-adhere-to-a-hippocratic-oath/; Ian Sample, "Maths and tech specialists need Hippocratic oath, says academic," *The Guardian*, August 16, 2019, https://www.theguardian.com/science/2019/aug/16/mathematicians-need-doctor-style-hippocratic-oath-says-academic-hannah-fry; National Academies of Sciences, Engineering, and Medicine; Division of Behavioral and Social Sciences and Education; Board on Science Education; Division on Engineering and Physical Sciences; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Analytics; Computer Science and Telecommunications Board; Committee on Envisioning the Data Science Discipline: The Undergraduate Perspective, *Data science for undergraduates: Opportunities and options* (Washington, DC: National Academics Press, 2018), https://www.ncbi.nlm.nih.gov/books/NBK532770/.

10 Indeed, privacy rules and technical access controls likely forbid those working with data from knowing much about any individual user and intervening directly in their lives as individuals.

of the most procedurally developed methods of organizational ethics, and has an outsized influence on how ethics owners imagine their jobs today.[11,12]

In the U.S., current human subject research rules were first codified in 1981, through the adoption of the Common Rule, which mandates the respect for persons, justice, and beneficence of new research.[13] In addition, the Common Rule is responsible for the creation of IRBs inside of U.S. universities (and to a lesser extent, government agencies, research hospitals and drug manufacturers). IRBs are tasked with reviewing research proposals that involve human beings and determining whether the proposals rightfully fall under the purview of the Common Rule and, if so, whether they meet the norms of respect for persons, justice, and beneficence. Where excessive risks to the human subjects are found, then IRBs work with researchers to reduce those risks and sometimes prevent a proposal from moving forward.

11 Double blind drug trials are the clearest example of this conflict. For example, a physician may be caring for a cancer patient, and therefore required to pursue their patient's best interests, while also running a controlled trial on a new chemotherapy drug, and therefore obligated to not favor their patient in a manner that would disrupt the scientific validity of the drug trial. Knowing whether the patient is receiving the experimental drug or the control drug/placebo is certainly the best way to support the well-being of the patient, but the only way to control for the bias in the results is to ensure that neither the physician nor the patient knows which is being administered. Therefore, the norms of informed consent are meant as a mediating factor between these conflicting duties—the patient acknowledges that a potential trade-off between their health and the betterment of scientific knowledge is acceptable to them as long as they are treated with dignity and kept as safe as possible.

12 Historically, the answers to these questions have been driven in large part by judicial and legislative responses to abuses of human research subjects. The Nuremberg Code (1947) was written in response to horrific abuses in Nazi Germany, and established the contemporary hallmarks of ethical research practices, such as requiring the informed consent of the patient, ensuring that the research subject could withdraw at any time without penalty, justifying human experimentation with prior research, and balancing the potential risk to the individual with potential benefit to humanity. The Nuremberg Code, and its descendants in research ethics, is therefore primarily focused on resolving the conflict between the duties of the physician-researcher to the well-being of individual patients who have become their research subjects and duties to further scientific knowledge for the benefit of all people. The question at hand for research ethics is always what principles and practices are available to scientists and physicians to not trample the rights and well-being of individuals along the way to those hypothetical collective benefits.

13 It is known as the Common Rule because it is administered by Health and Human Services but subscribed to in common by nearly all other federal agencies that fund experimentation.

Today, ethics owners and others often invoke IRBs as a model, or a foil, when discussing how the tech industry might get a handle on potential harm caused by their products.[14] However, the ways in which IRBs are structured, and the Common Rule has been interpreted, make simply porting the standards of research ethics over to product development at least cumbersome, if not mistaken.[15] Most notably, the central purpose of IRBs as defined by the Common Rule is to protect individual human subjects from harms caused by virtue of the research methods, and not to track aggregate and downstream consequences over populations or society that characterize concerns over automated decision-making systems. Furthermore, IRBs are only structured to consider proposed research prior to contact with human subjects, and not the iterative product development and deployment cycles that are used in technology companies.

Business Ethics

Perhaps the most obvious precedent for ethics owners to draw from is existing work on business ethics. While research and medical ethics are focused on protecting the moral interests of vulnerable individuals, business ethics is focused on the use of human values to coordinate activity inside a business operation. For example, should a business primarily serve the financial interests of its shareholders/owners, or be guided by a broader model of stakeholder interests?[16] This underlying

14 Molly Jackman and Lauri Kanerva, "Evolving the IRB: Building robust review for industry research," *Washington and Lee Law Review Online* 72, 3 (June 14, 2016): 442–457, https://scholarlycommons.law.wlu.edu/wlulr-online/vol72/iss3/8; Mitchell L. Stevens, "An Ethically Ambitious Higher Education Data Science," *Research & Practice in Assessment* 9 (Winter 2014): 96–97; Cinnamon Bloss, Camille Nebeker, Matthew Bietz, Deborah Bae, Barbara Bigby, Mary Devereaux, James Fowler, Ann Waldo, Nadir Weibel, Kevin Patrick, Scott Klemmer, and Lori Melichar, "Reimagining human research protections for 21st century science," *Journal of Medical Internet Research* 18, 12 (December 2016): e329, https://doi.org/10.2196/jmir.6634.

15 Jacob Metcalf and Kate Crawford, "Where are human subjects in big data research? The emerging ethics divide," *Big Data & Society*, Spring 2016, http://papers.ssrn.com/abstract=2779647.

16 Anant K. Sundaram and Andrew C. Inkpen, "The corporate objective revisited," *Organization science* 15, no. 3 (2004): 350-363; R. Edward Freeman, "Managing for stakeholders: Trade-offs or value creation," *Journal of Business Ethics* 96, no. 1 (2010): 7-9; Jacob M. Rose, "Corporate directors and social responsibility: Ethics versus shareholder value," *Journal of Business Ethics* 73, no. 3 (2007): 319-331; Karen Ho, *Liquidated: an ethnography of Wall Street* (Durham, NC: Duke University Press, 2009).

conflict remains central to ethics owners: One of our interviewees stated that "It's really about metrics going back to … how much value are we adding to our stock and to the stockholder? And [stockholder value isn't] based on what are we doing that's positive in the world despite whatever values or mottos the individual companies may have."

The debate about whose values are paramount is not simply a matter of moral consideration, but also of organizational imperatives, such as how managers should interpret the mandate to maximize either type of value, how employees can be motivated to fulfill the organization's model of value, and how success can be measured and thereby maximized.[17] For example, this conflict between stakeholder and shareholder value (a relatively recent orthodoxy of economic reasoning) is prominent in Silicon Valley's use of app-based gig work, where the well-being of gig workers in terms of wages and working conditions is directly pitted against the financial gains of venture capitalists and public investors.[18] Similarly, business ethics literature has addressed questions about how to attend to the well-being of employees in terms of the dignity and purpose of their work,[19] which is exemplified by the recent merging of broad "ethics" concerns and labor activism in the tech sector discussed above.

We observed ethics practitioners invoke engaging a broad range of stakeholders as a proxy for "ethics," but they also struggled with the challenges of identifying stakeholders and accounting for their diverse interests at scale. "Stakeholder value" is already challenging to measure when discussing, for example, the obligations a manufacturer may have to support a hometown economy rather than searching for cheaper labor overseas. But it is clearly even more amorphous when discussing a social media product that may be utilized in the overthrow

17 Michael C. Jensen, "Value maximization, stakeholder theory, and the corporate objective function," *Journal of applied corporate finance* 14, no. 3 (2001): 8-21.

18 Sheelah Kolhatkar, "The economist who put stock buybacks back in Washington's crosshairs," *The New Yorker*, June 20, 2019, https://www.newyorker.com/business/currency/the-economist-who-put-stock-buybacks-in-washingtons-crosshairs.

19 Denis G. Arnold and Norman E. Bowie, "Respect for Workers in Global Supply Chains: Advancing the Debate Over Sweatshops," *Business Ethics Quarterly* 17 (1): 135–45; Andrew Sayer, "Dignity at work: Broadening the agenda," *Organization* 14 (4): 565–81; John Hooker and Tae Wan Kim, "Ethical implications of the fourth industrial revolution for business and society," *Business Ethics: Volume 3* edited by David M. Wasielski and James Weber (Emerald Publishing Limited, 2019).

of a government halfway around the world from Silicon Valley. One of our interviewees who works for an enterprise software vendor pointed out that stakeholder engagement requires a major investment of time and resources: "I think it depends on the context of the company, but it is really just like a massive swath of stakeholders, right? And it's not—it's more of like who do they engage with rather than who needs to be under them. But in order to have that level of engagement, it takes a, you know, it's a fair amount of work. I could see two to three people, full-time, dedicated to just doing stakeholder engagement, right, holding events, interviews, the whole deal, right?"

Professional Ethics

In addition to the traditions of medical, research and business ethics, the tech industry also has precedents from engineering and computing professional ethics to draw on. These precedents include the recent addition of or updates to ethical guidelines and principles from a wide range of technology organizations. These include the *2012 Menlo Report* outlining responsible research practices for informational technologies that was sponsored by the U.S. Department of Homeland Security Science & Technology Directorate, the Association of Internet Researchers (AoIR), the Association of Computing Machinery (ACM), and the Institute of Electrical and Electronics Engineers' (IEEE) current code of ethics, which has been updated to include a discussion of computing's effects on society that was absent from the previous version.[20, 21]

20 US Department of Homeland Security, "The Menlo report: Ethical principles guiding information and communication technology research," August, 2012, https://www.dhs.gov/publication/st-menlo-report; "Ethics," AOIR, accessed February 26, 2020, https://aoir.org/ethics/; "ACM Code of Ethics and Professional Conduct," ACM Ethics, June 22, 2018, https://ethics.acm.org/; "ACM Updates Code of Ethics," Association for Computing Machinery, July 17, 2018, https://www.acm.org/articles/bulletins/2018/july/new-code-of-ethics-released. Jacob Metcalf, "Big data analytics and revision of the common rule," *Communications of the ACM* 59, 7 (2016): 31–33, https://doi.org/10.1145/2935882. "7.8 IEEE code of Ethics," IEEE, accessed February 26, 2020, https://www.ieee.org/about/corporate/governance/p7-8.html.

21 IEEE has invested in developing hybrid ethical-technical standards for autonomous computing systems, which provide more direct guidance than codes of ethics. See: https://standards.ieee.org/industry-connections/ec/autonomous-systems.html.

Codes of ethics such as these speak to the obligations of professionals.[22] They lay out what practitioners of a set of technical skills ought to do and ought not to do to be considered responsible members of a profession, which is a long-standing feature of trade organizations.[23] What such codes of ethics do not do, however, is offer guidance on how to resolve the concerns of ethics owners (who may not be computer or data scientists), which are fundamentally about how to understand the effects of technology products as they are integrated with society, at scale, how to track and measure the potential impact of these products, and how to ameliorate harm. Codes of ethics are helpful guidance for individual practitioners, but do not directly speak to the needs of organizations.[24]

When tech company executives have issued statements of ethical principles as a method for signaling the company's priorities to their staff, they are drawing on precedents that derive from business ethics and science and engineering professional ethics. As described by Adams et al., documenting and communicating ethical principles can be used to improve coordination within an organization because those principles can signal group cohesion and empower individual employees to make decisions that reflect a shared direction for the company.[25] As seen when Google CEO Sundar Pichai released a set of principles for ethical AI design in the midst of employee objections to the Maven project,[26] ethics can also be used as a relief valve for internal pressure. And, as argued by Greene et al. in their analysis

22 Dave West, "Why tech companies need a code of ethics for software development," *Entrepreneur*, April 19, 2018, https://www.entrepreneur.com/article/311410; McGreal, "A code of ethics for software," Scrum.org, October 31, 2016, https://www.scrum.org/resources/blog/code-ethics-software.

23 Fernand Braudel, *Afterthoughts on Material Civilization and Capitalism*, translated by Patricia Ranum (Baltimore: The Johns Hopkins University Press, 1977).

24 How these codes do potentially benefit ethics efforts in tech corporations is to provide leverage for technical workers to make demands upon their employers. See: https://www.informationweek.com/strategic-cio/it-strategy/acm-updates-code-of-ethics-for-computer-pros/d/d-id/1332362.

25 Janet S. Adams, Armen Tashchian, and Ted H. Shore, "Codes of ethics as signals for ethical behavior," *Journal of Business Ethics* 29 (2001), 199–211. See also: Heather E. Canary and Marianne M. Jennings, "Principles and influence in codes of ethics: A centering resonance analysis comparing pre- and post-Sarbanes-Oxley codes of ethics." *Journal of Business Ethics* 80, 2 (2008): 263–78.

26 Sundar Pichai, "AI at Google: Our principles," *The Keyword*, June 7, 2018, https://www.blog.google/technology/ai/ai-principles/.

of statements of ethical principles from industry, businesses also use ethical commitments to signal excellence and competitive advantage by drawing attention to their ostensible competence.[27]

Given these disjunctures between pre-existing ethics frameworks and the conditions of the technology industry, there are reasons to believe that the data-driven technologies of Silicon Valley pose distinct ethical challenges that require distinct approaches. While medical, business, and professional ethics frameworks often serve as inspirational (or limiting) precedents for ethics owners, we argue that the inability to address ongoing harms, to identify those harms outside of corporate limits, and evaluate organizational rather than individual behaviors means that there remains no sufficient means of owning ethics today. As we state above, none of these frameworks have the means to address the fundamental issue of scale in Silicon Valley.

Still, in our interviews, we often heard ethics owners invoking these existing approaches, largely echoing the public discourse about ethics in Silicon Valley. But, given the specific histories, purposes, and trajectories along which these forms of ethical reasoning have developed, these prior approaches seem largely orthogonal to the particular goals of ethics owners. In order to understand why ethics owners cannot simply pick up these existing approaches to ethics, we need to look much more closely at the actual conditions and practices that they engage in. Doing so, we find that (regardless of the ethical precursors they draw from) the work of ethics owners is largely constrained by the conventional methods of work typical of tech development in Silicon Valley.

Existing Ethics Methods

Silicon Valley, like other industries, relies on replicable routines—email trees, reporting requirements, quarterly evaluations—to coordinate

27 Daniel Greene, Anna Lauren Hoffmann, and Luke Stark, "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*, 2019, https://scholarspace.manoa.hawaii.edu/bitstream/10125/59651/0211.pdf.

activities across the organization. These routines can be a source of stability across large organizations, fostering predictability and accountability. Routines can also drive change across an organization, as they offer opportunities to model novel, desirable behavior through existing pathways.[28] Ethics owners, in seeking to routinize behaviors that attend to ethical issues across an organization, are therefore working to leverage existing routines. We see this with the privacy practices developed for compliance with GDPR and various consent decrees (for example, metadata tagging and privacy review documentation) being adapted for the more amorphous purposes of "ethics." But because most ethics owners' roles are new, there is no standard list of best practices. Instead, we identify some of the most common methods we observed: statements of principles and codes of ethics, ethics review boards, development of lifecycle tools, and other approaches like red teams and checklists. As we show, operating within these existing methods of "business as usual" both makes their work tractable and creates some challenging pitfalls.

Statements of Principles and Codes of Ethics

Some of the most public approaches available to ethics owners are statements of principles (sometimes statements of purpose) and codes of ethics.[29] There is a subtle distinction between the two: statements of principles guide an organization, and codes of ethics guide an individual or members of professional organizations. For example, one would expect Facebook and Google to have distinct statements of principles simply because they are distinct organizations with different business models, but it is another matter entirely to expect a single company like Google or Facebook to undertake developing codes of ethics for all data scientists or all UX researchers across the industry.

28 Martha S. Feldman and Brian T. Pentland, "Reconceptualizing Organizational Routines as a Source of Flexibility and Change," *Administrative Science Quarterly* 48, 1 (2003): 94.

29 Brent Mittelstadt, "Principles Alone Cannot Guarantee Ethical AI" *Nature Machine Intelligence* (May 20, 2019), https://papers.ssrn.com/abstract_id=3391293; Anna Jobin, Marcello Ienca and Effy Vayena, "The global landscape of AI ethics guidelines," *Nature Machine Intelligence* 1 (2019): 389–99; Greene, Hoffmann, and Stark, "Better, Nicer, Clearer, Fairer."

An organization's statement of principles acknowledges the responsibilities and commitments of the organization, and helps set its agendas and priorities. And organizations can have multiple statements, on multiple topics. In 2018, Google laid out a statement of AI principles, codifying their aspirations in developing AI products and laying out what applications of AI they will not pursue.[30] Statements of principle, according to ethics owners we spoke with, serve multiple important functions. Not only do they serve as a public articulation of a company's stances on controversial issues, they also serve as a point of leverage; employees inside the company might use such a document to justify a specific design choice, or as the basis for deliberations that might otherwise be reduced to more simplistic cost-benefit-risk analyses. This can figure in both subtle design choices and large decisions about whether to accept a particular client—one ethics owner described a case in which a potential client's business model contradicted a principle their company held that commits them to uphold the UN Universal Declaration of Human Rights.

Statements of principle tend to symbolize the end result of a process in which a company has invested time (and perhaps other resources) in identifying "for themselves, given their core values, their mission as a company," as one ethics owner put it. Going through this exercise can be clarifying, we were told, by bringing specificity to the areas of concern for a particular company. Many ethics owners we spoke with, however, pointed out that statements of principle cannot accomplish much if they are not deliberately linked to corporate practices. At minimum, statements of principle should be incorporated into onboarding and employee training materials, according to several participants. While such trainings can be superficial—who among us hasn't clicked through to get to the inevitable multiple-choice quiz at the end?—ethics owners note that they draw attention to the underlying issues, and provide a vocabulary for naming and discussing problems. Ethics owners pointed out that statements of principle also need to be "consistently lived" as concrete practices, or else they remain lofty public relations documents. "Principles are great, but I need practices," one ethics owner told us. Another told us that "until you add accountability and specific institutions … in the organization

30 Pichai, "AI at Google: Our principles."

… it's almost impossible to realize the real value or outcome." It is, for ethics owners, in the translations of principles into practices that their work becomes clear, and there are significant drawbacks to an over-reliance on statements of principle.

Codes of ethics similarly articulate expectations, but for the behavior of individuals within an organization or a profession. In Silicon Valley, codes of ethics may apply to individual employees of a company or individual members of a professional community, like software engineers or lawyers. For example, software engineers, many of whom are members of the IEEE, might commit themselves to abide by the IEEE Code of Ethics.[31] There has been much support for codes of ethics specific to data technology engineers in Silicon Valley, but no single code has gained broad acceptance. And with their application to individuals, not organizations, they are unlikely candidates for articulating models of institutional practices.

Ethics Review Boards

Many larger companies have ethics review boards, often explicitly modeled off of IRBs, that evaluate their products' potential ethical implications. These review boards vary widely in their size and composition, although most include senior leadership, specialists from various product divisions, legal and policy personnel, and, recently, ethics owners. Some ethics review boards are static committees, in which every member of the board participates in the review of every product that comes across its transom. Others are organized as a standing reserve committee, out of which a review board is empaneled with subject-matter experts who have relevant experience with a particular type of product or decision. Ethics review boards also vary widely in their power to affect product design or launch decisions—some ethics boards make non-binding recommendations to senior leadership, others are empowered to greenlight or halt projects or to return a project to an earlier development phase with specific recommendations.

Some ethics owners we spoke with had been involved in creating or working with these boards, and others have considered the usefulness of such a board at their own companies. And while some boards convene

31 "7.8 IEEE code of Ethics," IEEE.

"incredibly senior folks [to] look at these massive questions," more day-to-day decisions often require drawing together temporary groups from across multiple teams. The problem with the latter approach, one ethics owner pointed out, was that it's "just unfortunately not able to scale…. There's hundreds of these small, little decisions that get made, and so if you have a board that needs to evaluate all of them … that's just not scalable." Because of this, many ethics owners are working on how to categorize and triage the problems an ethics review board might need to address. Ensuring that the risks associated with a product or design get the requisite degree of scrutiny, and that the relevant information about those risks is available for review, remains an active area of concern for ethics owners.

Ethical review boards can also feel like punishment, we were told, and teams may go to great lengths to game the process through which their projects might receive such scrutiny. To circumvent this, one ethics owner told us, "What we used to do in my old job is we would—it was a great way to get people to adopt—is they had to apply to work with us. Human psychology is fascinating. Anyway … that really worked. It drove adoption. And as part of deciding who we would work with and accept with, we would go through a number of things [that aided the review process]."

Development Lifecycle Tools

Software development lifecycle tools like Jira or Azure DevOps are an integral part of software engineers' daily workflows. These tools track which work has been completed and which still needs to be done. They provide an overview of how the many pieces of a complex project being built by a multi-person team come together. Software requirements, bug reports, documentation, progress toward completion, and attribution of work can all be tracked with these tools, and because of their ubiquity, ethics owners see them as a way of embedding ethics into the daily work of software engineers and data scientists. For instance, several ethics owners we spoke with described efforts to use development lifecycle tools to identify potential sources of bias or

unfairness in datasets and algorithms.[32] Other ethics owners are looking to these tools as a way of tracking product decisions in ways that make the fact-finding task of ethics review boards less onerous; rather than having to track down every decision that was made about a product, a properly configured development lifecycle tool could generate a report about those decisions on demand.

When discussing lifecycle development tools as part of a strategy for managing the ethical implications of products, ethics owners highlighted the challenge of distributing the responsibility of evaluating ethical implications across an entire organization or team. For these tools to work as designed, each team member must use them consistently; incomplete usage or conflicting definitions for what might be tracked inside such a tool make a comprehensive review difficult, if not impossible. These tools, whether used as an instrument of ethics owners or not, rely on the diligence and discernment of individual engineers. While ethics owners might trust everyone in their company to "do the right thing," lifecycle development tools embed that trust in technical systems.

Other Approaches

Ethics owners have been adapting numerous other existing corporate methods for organizing work. For instance, some ethics owners are looking at popular development methodologies as places where ethical concerns can be addressed. "Agile"—a software development methodology that emphasizes an iterative process utilizing in-person "stand-up" meetings and a series of development "sprints" towards clearly defined tasks—is one of the most common development cycle methodologies in Silicon Valley.[33] Some ethics owners discussed an adaptation of Agile called "consequence scanning," in which an "event" can be inserted into a standard Agile methodology that prompts developers to consider the ethical implications of their

32 Kush R. Varshney, "Introducing AI fairness 360." *IBM Research Blog*. https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/.

33 Kent Beck, Mike Beedle, Arie van Bennekum, Alistair Cockburn, Ward Cunningham, Martin Fowler, James Grenning, Jim Highsmith, Andrew Hunt, Ron Jeffries, Jon Kern, Brian Marick, Robert C. Martin, Steve Mellor, Ken Schwaber, Jeff Sutherland, Dave Thomas, "Manifesto for agile software development," 2001, https://agilemanifesto.org.

products in a structured way.[34] Similar such adaptations of Agile abound, reproducing the Agile methodology in miniature, taking a series of steps familiar to product development teams and orienting them toward the issues that ethics owners see for their work. For example: adapting "user stories" to include a racially or gender diverse population, adapting product requirements to include a consideration of unforeseen consequences.[35]

Another common product development technique in Silicon Valley is the use of a "red team" to act as an adversary that tries to breach the security of a system. One ethics owner we spoke with suggested adapting the "mentality of a red team" to look at a product to determine how a product might be abused or used unethically. They pointed out that "it's so easy for people who are in that invention business to only look at the positive and say 'This is going to be so awesome. This is going to be so great.' And then either forget or ignore that hey, somebody's going to potentially take this product and use it for something that's really not good." Ethics owners described red teams as useful for understanding how unanticipated harms might occur upon the release of a product, as well as how bad actors might co-opt a product for nefarious purposes.

Ethics owners have also looked to other industries for inspiration. We heard several ethics owners reference pre-flight checklists and pre-surgery checklists as models that might be emulated in tech ethics; by requiring engineers and product managers to complete a checklist prior to pushing a feature or launching a product, it could

34 "Consequence scanning—an agile practice for responsible innovators," doteveryone, accessed February 26, 2020, https://www.doteveryone.org.uk/project/consequence-scanning/.

35 "Ethical OS: A guide to anticipating the future impact of today's technology," Institute for the Future and Omidyar Network, August, 2018, https://ethicalos.org/wp-content/uploads/2018/08/Ethical-OS-Toolkit-2.pdf; Alix, "Agile Ethics: Managing Ethical Complexity in Technology," *Medium*, March 14, 2019, https://medium.com/@alixtrot/agile-ethics-managing-ethical-complexity-in-technology-63131f7bab95.

help them check common mistakes.[36] Such checklists have even been transformed into python packages so coders can call up a checklist and mark items complete, directly in the programming environment.[37] A similar approach can be found in the "nutrition labels" for data sets and algorithmic models, developed by Google and Microsoft researchers, which encourage developers to document their models and data.[38] This metadata, we were told, can then be used for evaluating the ethical implication of the components of tech products. Like all systems relying on metadata or checklists, they are only as robust as those using them, and therefore, in the absence of an audit mechanism, rely on the diligence and integrity of individuals and teams. Pre-flight and surgical checklists also have a clear binary outcome to avoid (crashing and patient death, respectively), whereas ethics in technology requires address multi-factorial scenarios spread across populations.

Lastly, given the challenges of doing the work of ethics inside corporations, ethics owners are keenly aware of the power of metrics to achieve goals, particularly when tied to incentives. One ethics owner described what happened when "something became so important to us that we materially changed how people were paid … when that happened we saw people do things that I thought were outside of their skill set." By this logic, if more ethical product development is to truly become an organizational goal, metrics will have to be implemented that reflect this goal. Developing metrics, particularly around measures like OKRs, is already recognized as highly desirable by ethics owners. We heard a steady refrain during fieldwork: "If we can measure it, we

36 Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes, "Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing," In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)* (New York: Association for Computing Machinery, 2020); Atul Gawande, *The Checklist Manifesto: How to Get Things Right* (New York: Henry Holt, 2011).

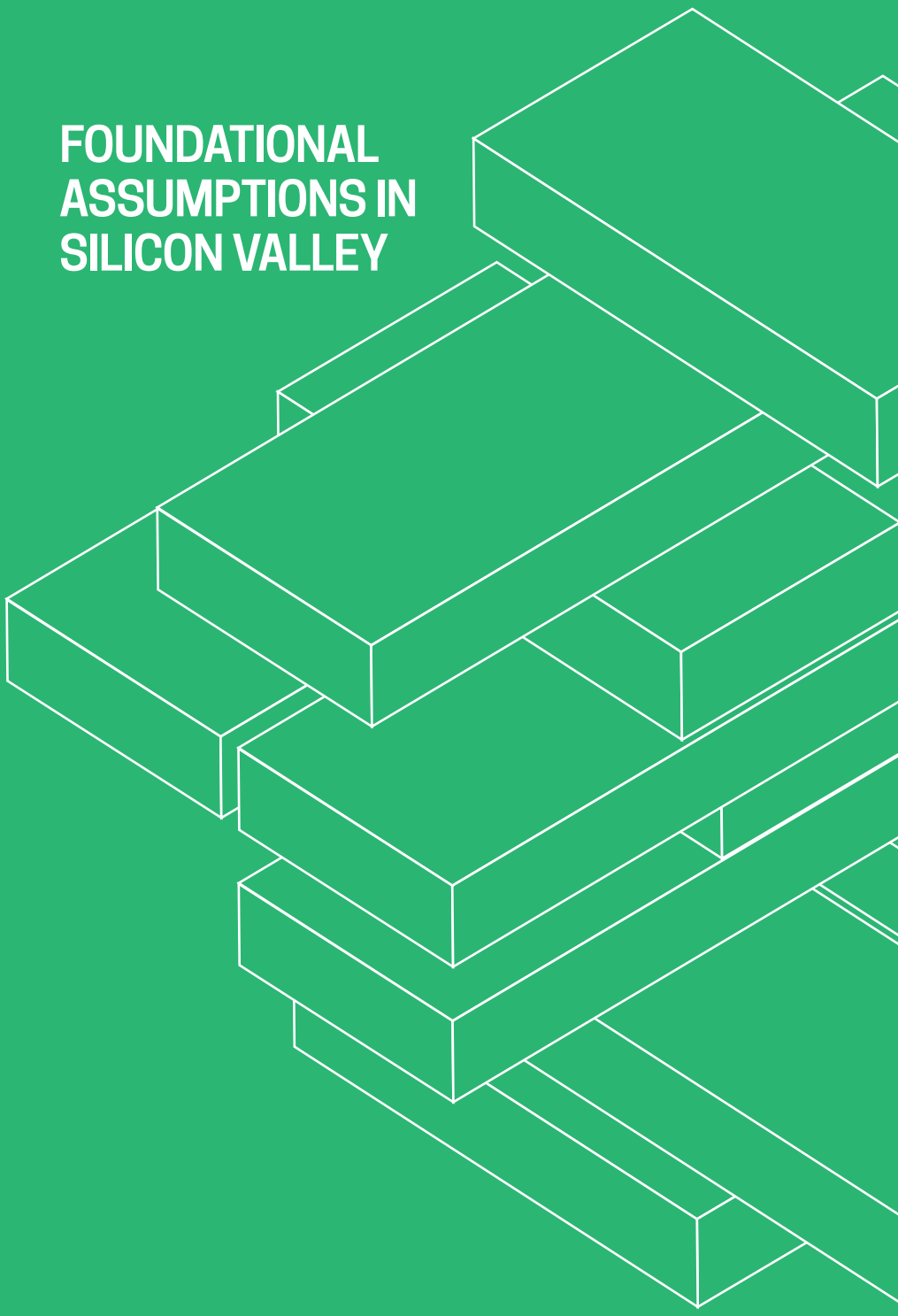
37 drivendataorg/deon, Github, accessed February 26, 2020, https://github.com/drivendataorg/deon.

38 Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru, "Model Cards for Model Reporting," In *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT** (Atlanta, GA: ACM Press, 2019). https://doi.org/10.1145/3287560.3287596; Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumeé III, and Kate Crawford, "Datasheets for datasets," arXiv preprint arXiv:1803.09010 (2018).

can fix it." Establishing clear metrics may indeed provide leverage for prioritizing work and supporting ethical practices; such metrics can address the most clearly defined issues, or the most pressing issues that affect the largest number of people, first, and work down the stack in order of descending importance and immediateness. However, this approach risks ignoring important aspects of ethics work that remain unmeasured or unmeasurable.

Like software development lifecycle tools, these other approaches—development methodologies, red teams, checklists, and metrics—all rely on developers to evaluate the potential harms of their products. While ethics owners are optimistic about approaches that encourage developers to anticipate harms, these existing approaches all are necessarily constrained by the scope of developers' imaginations about how products they build might affect the lives of others whose lived experience may not resemble their own. And, of course, ethics owners are also concerned about addressing *un*anticipated harms that their products might produce in the world. But in our interviews, ethics owners largely indicated that they rely on individual users bringing these harms to their attention through existing customer service portals. They also find themselves responding to investigative reporting about their products. Both these sources lead to what ethics owners have described as one-off or *ad hoc* responses that, while perhaps leading to substantial modifications to products and services, are not easily incorporated into a systematic approach for addressing future unanticipated harms.

FOUNDATIONAL ASSUMPTIONS IN SILICON VALLEY



Most ethics owners have worked inside Silicon Valley companies for significant parts of their careers before being hired into their current roles. Thus these ethics owners are likely to craft interventions, and work through the tensions they face, in ways consistent with the standard way of doing business in Silicon Valley. Whether they change what it means for companies to do "business as usual" or whether they fit in seamlessly with existing business processes, ethics owners operate within a set of underlying assumptions about the interactions of culture, technology, and economic markets. These assumptions help position Silicon Valley tech companies as both the cause of, and solution to, the ethical issues at the center of data-driven development.

Because these assumptions are so fundamental to Silicon Valley's practices, they remain subtle, at times invisible, but nevertheless relentlessly powerful forces that shape how work is done. These assumptions are by no means limited to Silicon Valley and are deeply implicated in how economic and political power is reproduced along class, race, and gender lines. Broadly speaking, operating within these assumptions leaves in place the patterns of racial and gendered power that persist across society, and in Silicon Valley in particular.[39] By virtue of their subtlety and pervasiveness, these assumptions may be replicated by ethics owners in ways that imperil the viability of their project.

We identify three key assumptions that carry over from Silicon Valley, writ large to constrain the work of ethics owners: meritocracy, technological solutionism, and market fundamentalism. As we have stated elsewhere, these assumptions operate whenever the idea is put forward that "trenchant social problems can be addressed through innovative technical solutions developed by those with the most aptitude and creative energy, and that an unencumbered market will recognize, reward, and disseminate the best solutions."[40] Across our

39 Ruha Benjamin, *Race after Technology: Abolitionist Tools for the New Jim Code* (New York John Wiley & Sons, 2019); Catherine D'Ignazo, and Lauren F. Klein, *Data Feminism* (Cambridge, MA: MIT Press, 2020); Jessie Daniels, Mutale Nkonde, and Darakhshan Mir, "Advancing Racial Literacy in Tech: Why Ethics, Diversity in Hiring & Implicit Bias Trainings Aren't Enough," Data & Society Research Institute, 2019.

40 Jake Metcalf, Emanuel Moss, danah boyd, "Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics," *Social Research* 86, 2 (2019):449–76.

interviews, we consistently heard ethics owners struggle (implicitly or explicitly) with these assumptions, and often re-embed them into their own approaches.

Meritocracy

In Silicon Valley, ethics owners are often constrained by organizational pressure to justify their practices on meritocratic grounds. "Meritocracy," as a term, originated in Michael Young's 1958 speculative fiction, *The Rise of the Meritocracy 1870-2033*, which lampooned British society for accepting rampant inequality as it retreated from liberal democratic values.[41] But it has since been deployed in all seriousness as a way to rationalize inequality by offering an ideological explanation that links the unequal distribution of wealth (and power) to natural differences in individuals' abilities.[42] Meritocracy is visible in calls for perpetual self-promotion and self-improvement,[43] but also in the influence those who have demonstrated "merit" are given in public affairs. This remains the case even as the central claim of meritocracy is debunked by evidence of contesting "race- and gender-blind" opportunities for advancement in the industry. Meritocracy props up the ability of those who have achieved success in one domain to be seen as authoritative in another domain.[44]

41 Michael Young, *The Rise of the Meritocracy* (Bristol, UK: Thames and Hudson, 1958).

42 Yascha Mounk, 2017. *The Age of Responsibility: Luck, Choice and the Welfare State* (Cambridge, MA: Harvard University Press, 2017); Emilio J. Castilla and Stephen Benard, "The Paradox of Meritocarcy in Organizations," *Administrative Science Quarterly 55, 4 (2010)*: 543–676; Ajantha Subramanian, *The Caste of Merit: Engineering Education in India*. (Cambridge, MA: Harvard University Press, 2019).

43 Emilio J. Castilla, "Gender, Race, and Meritocracy in Organizational Careers," *American Journal of Sociology* 113, no. 6 (2008): 1479–1526; Jessie Daniels, "'My Brain Database Doesn't See Skin Color' Color-Blind Racism in the Technology Industry and in Theorizing the Web," *American Behavioral Scientist*, no. 11 (2015): 35.

44 Jo Littler, *Against Meritocracy: Culture, Power and Myths of Mobility* (London; New York, NY: Routledge/Taylor & Francis Group, 2017); Alistair S. Duff, "Rating the Revolution: Silicon Valley in Normative Perspective" *Information, Communication & Society* 19, 11 (2016): 1605–21.; Alice Marwick, "Silicon Valley isn't a meritocracy. And it's dangerous to hero-worship entrepreneurs," *Wired*. November 25, 2013, https://www.wired.com/2013/11/silicon-valley-isnt-a-meritocracy-and-the-cult-of-the-entrepreneur-holds-people-back/.

In Silicon Valley, the infrastructure of meritocracy is as common as open floorplan offices or luxe cafeterias. For decades, the economic power, and cultural cachet, that meritocracy holds has been expressed by the idea that those who work there are the "best and the brightest", simultaneously dismissing the underrepresentation of Black and Latinx people in positions of power with claims that only raw talent is rewarded without respect to race or gender. Silicon Valley tech companies try to recruit "the best people" from "top schools," which popular wisdom indicates will help them triumph in a survival-of-the-fittest contest.[45] The result of this meritocratic thinking can be a misplaced confidence in existing employees to succeed at any challenge.

We often heard from ethics owners that "there are really good people working in all of these companies who try to do the right thing." In many companies, this is a valid starting point for addressing ethical challenges; ethics owners in our interviews often described Silicon Valley technologists as "grappling with hard questions on the ground," entrusting them with evaluating the ethical implications of their projects. While technologists may indeed be well-positioned within a company to evaluate whether hypothetical harm was realistic given specific technical specifications, all too often they were also seen as authorities on the nontechnical aspects of an ethical challenge, such as what defines "harm" and whose interests should be addressed.

A common refrain we heard from ethics owners was that their technical teams will gather in a room and "think really hard" about the potential ethical implications of a product or feature. The intuition that good technical teams will also be good at enumerating potential ethical implications comports with the importance of technical team members' intuition in other domains, and is underwritten by meritocratic assumptions.[46] It is one thing to ask engineers, with years of education and experience in engineering, to think of all the potential engineering problems that might arise from a design decision. It is another thing altogether to ask engineers to think of how

45 Ho, *Liquidated*.

46 Samir Passi and Steven J. Jackson, "Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects," *Proceedings of the ACM on Human-Computer Interaction* 2 (CSCW), 2018.

a design decision might affect the lives of those whose experiences and social position are drastically different from that of most Silicon Valley technologists.[47] This point is particularly salient because these teams are not reflective of the diversity of people implicated in tech development. The teams themselves lack racial, geographic, class, and gender diversity, with Black and Latinx technologists especially poorly represented. Moreover, the professional respect and value accorded to the expertise of engineering teams often does not extend to the lived experience of technicians from marginalized communities.[48]

Technological Solutionism

Ethics owners operate in a world in which technology is assumed to be the solution to most problems. Indeed, as we write elsewhere, "the idea that technology *can* solve problems has been reinforced by the rewards the industry has reaped for producing technology that they believe *does* solve problems."[49] The list of social problems—from communication (email) to entertainment (streaming services) to transportation (ridesharing)—the tech industry believes it has "solved" through new and better technology is long. In his book *Winners Take All: The Elite Charade of Changing the World*, Anand Giridharadas describes the cultural world in which these ideas, and the accolades for these ideas, circulate in ways that obscure the negative externalities of technological innovation and depoliticize some of the most serious collective issues facing society today.[50]

While many claims of social innovation through technology are dubious—hotels fulfilled the social need for temporary housing long before Airbnb came along—others are quite valid. Page-rank search algorithms really do put the world's information at our fingertips,

47 Lucy Suchman, 2002. "Located Accountabilities in Technology Production," *Scandinavian Journal of Information Systems* 14, 2 (2002): 16.

48 The use of technology for racist, anti-Black applications, including but not limited to policing and racial justice, has an even longer history as told in: Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Medford, MA: Polity Press, 2019).

49 Metcalf et al. 2019 (emphasis altered).

50 Anand Giridharadas, *Winners Take All: The Elite Charade of Changing the World*. (New York: Alfred A. Knopf, 2018).

social media connects people across great distances and at times enables new forms of political participation, etc.[51] However, most technologies are inherently dual-use, producing negative effects alongside positive ones. Even when those in the tech industry acknowledge the negative social impact of technologies they've built, they often propose technical solutions to those problems, as if more or different tech can fix the problems of tech.[52]

For ethics owners, they work within an industry in which the assumption that "tech can solve social problems" winds up framing ethics as something that can be "solved." While many ethics owners we spoke with acknowledge the futility of seeking a "silver bullet" for addressing ethics concerns, in the same breath they will also mention using "the super powers given to you by tech," a view that pervades one-on-one interviews and much of the industry's public discourse. Ethics owners search for toolkits and checklists that can be implemented as part of the design, development, and deployment process, and operate under the belief that attention to technical detail can forestall any negative social impact of a system.

However, much of the most intractable social impact of technology arises from products that work *exactly* as intended. And yet, we often heard that "you're not going to have an issue with the AI if your data is all good and you've curated that appropriately—checked for fairness, mitigated against negative bias, and carried metadata along so that you can revisit decisions … And the problem exists when you don't do those things." This attitude leads to a well-intentioned search for best practices (i.e., procedures, checklists, concrete metrics) that can guarantee an ethical outcome. But best practices are well-suited

51 Amy Langville and Carl D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings* (Princeton, NJ: Princeton University Press, 2006); Gohar Feroz Khan, Bobby Swar, and Sang Kon Lee, "Social Media Risks and Benefits: A Public Sector Perspective," *Social Science Computer Review* 32, 5 (2014): 606–27.

52 For counterpoint, see: Ben Green, and Salomé Viljoen, "Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought," In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, Barcelona, ES.

to technologists' work—to systems that can be accurately modeled in a development environment—not the social worlds in which their products are expected to function responsibly and ethically.[53]

Market Fundamentalism

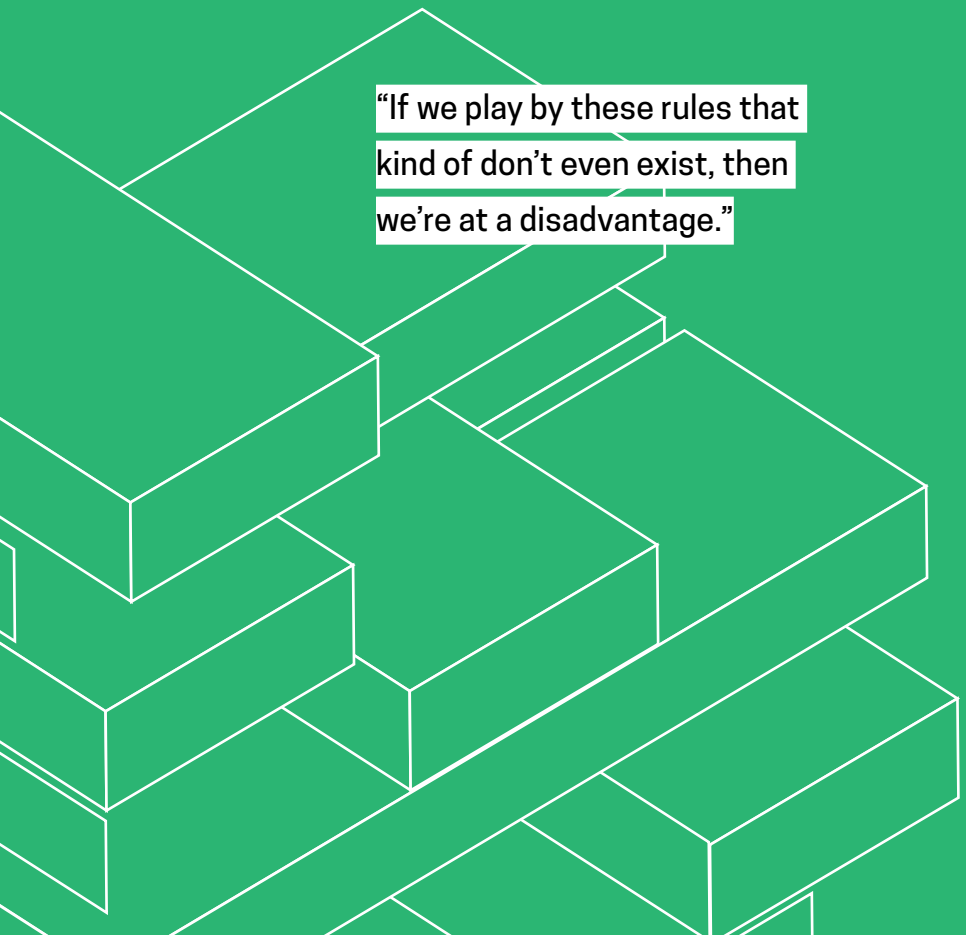
The third assumption that inextricably shapes the work of ethics owners is the idea that the market is the ultimate arbiter of which ethical approaches can work, and which cannot. We heard from ethics owners resigned to the feeling that "[market] success trumps ethics," a statement that often accompanied the fear that unethical practices might persist inside companies, as long as they remain profitable. The idea that the market can decide which ideas are good and which are bad is based on an extension of neo-classical economic ideas that have been applied far beyond the domain of economic markets.[54] Unless ethical risk can be priced into the product development process, it remains very difficult for ethics owners to argue for organizational interventions that add friction to that process. However, "if people actually stop using certain services and/or showed more preference for services that didn't have the same [flaws], you can be sure companies will be responsive," we were told.

Assumptions about the need to satisfy market demands also help explain why companies tend not to invest in ethics practices until they are fairly large or well-capitalized. Many ethics owners we spoke with expressed a desire for industry-standard, durable processes that they could adapt to their own company's needs. This implies that only the largest, most well-capitalized companies currently have the resources to address ethics. If these tools could be standardized, we heard, then even smaller startups could "do ethics," so long as the initial investments in developing those tools were borne by others.

53 Madeleine Clare Elish and Elizabeth A. Watkins, "When Humans Attack: Re-Thinking Safety, Security, and AI." Data & Society: Points (blog), May 14, 2019, https://points.datasociety.net/when-humans-attack-re-thinking-safety-security-and-ai-b7a15506a115.

54 David Harvey, *A Brief History of Neoliberalsim*. (New York: Oxford University Press, 2005)

NAVIGATING TENSIONS



Ethics owners adapt existing corporate engineering practices toward their own goals—addressing the ethical challenges that those same practices have played a role in producing. In doing so, they encounter several tensions. These tensions are fundamentally irresolvable, because they sit astride competing descriptions of the work that individuals inside of companies perform—engineers approach work differently than general counsels do, enterprise sales executives approach work differently than retail salespeople. Because ethics owners work as translators across different components of their organizations, working through these tensions without resolving them is an essential component of the role. As owning ethics becomes more concrete, moving away from the articulation of abstract principles and values towards concrete business practices, these tensions come into sharper focus.

To fully resolve these tensions, ethics owners would need to develop a generalizable method that could be applied across cases, however the specificity of the issues they deal with call for each issue to be dealt with *in context*. Therefore, the work of owning ethics in Silicon Valley is the work of navigating these tensions *without* resolving them. We identify six main tensions that highlight the specificity with which ethical issues must be addressed, how they must work against the grain of generalizable solutions.

Personal Ethics, Corporate Duties

As employees, ethics owners always have both personal ethics and corporate duties, which often conflict.

Ethics owners may work inside of companies, they may work for companies, but they also remain individuals, with moral stances and opinions that are not identical to the company itself. This is true even if they are socially situated in ways that validate their own ethical sensibilities while foreclosing other possible ways of understanding their own ethical engagement with the world.[55] They must, therefore,

55 Sareeta Amrute, 2019. "Of Techno-Ethics and Techno-Affects," *Feminist Review* 123, 1 (2019): 56–73, https://doi.org/10.1177/0141778919879744.

sit at a point of tension between the priorities of the corporation, which include a fiduciary duty to shareholders and adherence to the corporate mission, and any number of personal stances on the activities of the company. In instituting ethical practices and interventions, they must constantly translate their work back into the terms of corporate responsibility, sometimes at odds with their own understanding of personal responsibility, focusing on "metrics that go back to the stockholder," as we were told in our interviews. The interventions ethics owners make, particularly when that intervention slows or stops a product from going to market, must be justified in terms of the bottom line. Even when a decision is morally justifiable, it must also be rendered in terms of a liability risk avoided, another opportunity for profit generated, or a reputational benefit for the firm.

Tech employees often have specific, specialist knowledge of the inner workings of their companies that are inaccessible to outsiders, and have begun to assert their own moral sensibilities about their work.[56] Ethics owners—but also employees in other roles—have told us that there are certain kinds of projects they would refuse to work on, and some have left their jobs when they found out about projects they were morally opposed to.[57] They also see themselves as in the best position to hold their employers accountable to ethical principles, even if they think many of their fellow employees are "not yet motivated by ethics." This pressure is countered by internal pressure from other managers to maintain existing practices, which represent sunk costs, which avoid jeopardizing revenue streams, and which maintain legitimacy in the eyes of the public. To be fair, not all employees can push back against management for ethically dubious instructions, due to their economic situation, visa status, or any number of other issues bound up with structural forms of privilege, particularly along racial lines.

56 Dani Deahl, "Google Employees Demand the Company Pull Out of Pentagon AI Project," *The Verge*, April 4, 2018, https://www.theverge.com/2018/4/4/17199818/google-pentagon-project-maven-pull-out-letter-ceo-sundar-pichai.

57 Recent graduates entering the workforce are making similar decisions about potential employers. See: https://www.nytimes.com/2020/01/11/style/college-tech-recruiting.html.

Upside Benefits, Downside Risks

Ethics owners find that it is far easier to advocate for changes that limit potentially harmful effects of a product than it is to advocate for changes that may prove beneficial for users.

Ethics owners are well-positioned to make a straightforward value proposition about their work: It acts as a hedge against negative consequences that might arise from a public controversy surrounding a company's products or business practices. They are able to argue in clear language that if a company is caught up in a public controversy, the cost to the share price, fines, or reputational harm to the company's brand can be more than offset by the expense of ethical practices that might avoid such an end. But ethics owners must also grapple with how to articulate the upside benefits of generating better products that come out of more ethically considerate development practices.

Both the upside benefits and the downside risks are hard to quantify, but companies are already familiar with handling downside legal and public relations risks. The outcomes ethics owners aim at are distinct from the outcomes of a capable PR team, or from limiting the extent of legal liability that a general counsel's teams are concerned with. These outcomes are distinct because ethics owners are tasked with making substantive interventions into the products themselves, as they affect end-users, and not just the public messaging around those products or the contractual language that binds users and purveyors of those products.

Ethics owners described difficulties they have faced in advocating for improvements to products that don't directly affect the bottom line, compared to the relative ease they had in "killing" potentially harmful or risky aspects of a product. Even as companies make headway in seeking to improve products before they ship, some of the most powerful interventions that could help assess the ethical implications of particular technologies threaten the commercial viability of the products they are used to evaluate. The same methods that could systematically prove a product's safety could be used to steal the intellectual property behind the system or otherwise take advantage of

its design features.[58] Ethics owners worry that "the same method that helps you detect where edge cases might pop up is the same method to gaming the system."

Additionally, the basic accounting necessary for any business effort is challenging for ethics owners because of the relatively new nature of the role. Even if a business is willing to earn less profit, or accept the price of upfront investments in order to mitigate future downside risk, it is still necessary to put a concrete number on those forecasts. For example, there are industry-standard expectations about what percentage of operating expenses should be dedicated to legal teams and outside counsel. Yet we heard from ethics owners that there was a lack of clarity around relatively simple measures of expenses, such as what head count is needed to run an ethics review of a new product or how much time is reasonable to allot such a review on a product launch calendar. Even in the least-cynical sense, "ethics" must be priced on the balance sheet, and there is no clear consensus on what the cost is.

Direct-to-Consumer, Business-to-Business

For ethics owners, how they approach their work will be deeply affected by their employer's business model, and standard practices might not work across different business models.

One of the most interesting dynamics we found is that enterprise or business-to-business (B2B) companies, which sell their products and services to other companies, are quite far ahead in investing in formal ethics capacity. This is in marked contrast to the scrutiny that public scandals place on consumer-facing business-to-consumer (B2C) platforms, like Google and Facebook, and how their ethics practices directly affect customers. One reason for this is that it is more straightforward to articulate the upside benefits of robust ethics practices for B2B companies.

58 Srinivas Nidhra, "Black Box and White Box Testing Techniques—A Literature Review." *International Journal of Embedded Systems and Applications* 2, 2 (2012): 29–50.

For example, using ethical design to reduce their clients' risk exposure provides clear economic value to both parties, which has so far incentivized B2B software companies to invest in tools for mitigating algorithmic bias. At the same time, however, ethics owners ruminate on the fact that enterprise companies may have little insight into what their client companies are actually *doing* with their products and services, leaving the B2B's to wonder if their product is being used for unethical purposes. Additionally, they point out that the public has significantly less access to the workings of enterprise software systems, where even the biggest companies may not be household names.

Ethics owners at B2B companies acknowledge that "it is complicated to sit in the middle of a range of different business models that may have some tensions with each other... and we gain a lot of value as a result of having to balance those things, because it does put us in a societal position of actually having to make decisions that fit with a broad range of different kinds of complex stakeholders." In other words, ethics owners find it difficult to craft a single set of policies that can be deployed across different kinds of business models. Being an ethics owner in an enterprise company is different from being an ethics owner in a direct-to-customer company. B2B models have a more explicitly contractual relationship with enterprise clients, while B2C models rely on terms of service and technical controls to manage ethical risks. And, at a company with a mixed business model, simultaneously B2B and B2C or involving hardware and software, it can be very difficult for ethics owners to produce a single set of approaches.

For ethicists observing from the outside, there is some risk that focusing on individual and societal harms could be co-opted by the needs of enterprise salespeople to attract corporate clients. When B2B companies use ethics as a selling point for their services, it raises the possibility that "ethics" will become a service that gets bundled in with other enterprise services, and is only offered at enterprise prices. Conversely, ethical considerations may be downplayed or receive short shrift for end-users, citizens, and regulators. Insights into algorithmic harms might be behind a paywall in business-to-government and business-to-business software business models, where the services provided may be even more opaque to the public. In retrospect, it

appears obvious that different business models will approach tech ethics differently, however the public conversations about tech ethics have largely missed these complexities.

Measurable and Nonmeasurable Impacts

Developing metrics for measuring the ethical impacts of tech products, and tracking the success of intervention, remains one of the greatest challenges for ethics owners.

Ethics owners explained to us that they can use quantifiable data from end-users to assess the ethical impact of certain products. Almost all Silicon Valley products produce huge amounts of data on user activity—for services delivered to smartphones, this can include battery charge level, location, and even the temperature of the device. Some ethics owners look to translate such data into performance ratings that indicate the users for whom the product is working the best, or at least working as intended, and then check those performance metrics against their demographic status. The ethics owner might discover, then, that a particular product is performing worse for a protected class—say a particular gender or race—in a way supported by empirical data. Or, inversely, measuring negative impacts might require acquiring data that the company might otherwise not wish, or not be legally permitted, to gather directly through their service. In those cases, while the impact is theoretically measurable, an assessment requires engaging in practices that potentially have other kinds of negative impacts, such as inferring or purchasing sensitive data.[59]

But other kinds of impacts are much harder to measure, because they occur beyond the scope of data collection practices. In interviews, ethics owners admitted they are limited in their ability to understand the population of people who are either *unable* to use their products, or what would have happened had people not used their products. That said, it is exceedingly difficult, if not impossible, to understand phenomena that are not representable as data—prison recidivism

59 Jennifer Eberhardt, "Can Airbnb train hosts not to be racists?" *Daily Beast*, 2019, June 12, https://www.thedailybeast.com/can-airbnb-train-hosts-not-to-be-racists.

models cannot measure the number of prisoners who remain incarcerated but would not have gone on to commit a crime if released, and college achievement models cannot measure how those who are not admitted to higher education would have performed in college.[60]

Similarly, it is exceedingly difficult for a company to measure secondary effects from their products. For example, when Open.ai developed the text-spoofing algorithm GPT-2 as a research project, they refrained from immediately releasing it publicly out of concern that it might undermine societal trust in the mediascape, contributing to the so-called "fake news" problem.[61] Ethics owners discussed the difficulty of tracking such effects, which are widely distributed and might occur in ways that don't directly come into contact with the product itself. In this example, people who never read an article produced by GPT-2 might cease trusting other media sources due to the existence of such a technology.[62] The difference between measurable and nonmeasurable impact can lead to extra attention for the class of ethical impacts that are already heavily documented and amenable to quantitative measures, like algorithmic bias and fairness, and a relative lack of scrutiny for harder-to-measure or impossible-to-quantify impact.

Users, Nonusers

Ethics owners are given the task of understanding broader social impact of their companies' products, but have little insight into how nonusers are affected by those products.

Silicon Valley companies know far more about people who use their products than they do about people who don't. Similar to the tension between measurable and nonmeasurable impacts, the impacts on

60 Abigail Z. Jacobs and Hanna Wallach, "Measurement and Fairness," *ArXiv:1912.05511 [Cs]*, December, 2019, http://arxiv.org/abs/1912.05511.; Laurel Eckhouse, Kristian Lum, Cynthia Conti-Cook, and Julie Ciccolini, "Layers of bias: A unified approach for understanding problems with risk assessment," *Criminal Justice and Behavior* 46, no. 2 (2019): 185-209.

61 Aaron Mak, "When is technology too dangerous to release to the public?" *Slate* February 22, 2019, https://slate.com/technology/2019/02/openai-gpt2-text-generating-algorithm-ai-dangerous.html.

62 Regina Marchi, "With Facebook, blogs, and fake news, teens reject journalistic 'objectivity'," *Journal of Communication Inquiry* 36, no. 3 (2012): 246-262.

people who companies hold the most data about—their users—are much easier to grasp for ethics owners, compared to how ethics owners can assess impacts on nonusers. Companies pay attention to their users, of course, because they want to retain them as customers, and because they are already in a relationship of responsibility with them. But there are also impacts that may affect nonusers, in ways that ethics owners find difficult to address. For instance, while a ride-share company might know how the riders and drivers use their platform, it's harder to know how people *not* using that platform are affected by things like routing decisions, neighborhood rollout plans, and how public transportation planning might be affected by the decisions made by the ride-share company. Simply gathering data about nonusers may actually increase the ethical harm, particularly if that data finds its way into other business operations.[63]

Nonuser impacts, like nonmeasurable impacts, are difficult to address because of the lack of data, and because they may be outside the dominant business model for a company, but they are no less likely to be significant. Nonuser impacts are also difficult to even conceptualize, let alone address, particularly from within a company used to evaluating its products through the feedback loops its informational architecture enables.

Specificity, Generalizability

Like the rest of Silicon Valley, ethics owners face significant challenges when attempting to scale their methods across contexts.

The work of ethics owners, given the sets of assumptions that pervade the work they do inside of Silicon Valley tech companies, produces tensions that cannot be resolved. By this, we mean that there is no single, or optimal, rule that can be uniformly applied in every situation—these

63 Russell Brandom, "Shadow profiles are the biggest flaw in Facebook's privacy defense." *The Verge*, April 11, 2018, https://www.theverge.com/2018/4/11/17225482/facebook-shadow-profiles-zuckerberg-congress-data-privacy.

tensions are not resolvable through a maxim, like the Hippocratic Oath's "First, do no harm." Rather, the tensions ethics owners face in their work must be addressed in context, and specific approaches cannot be simply generalized or cleanly adapted into universal rules. In short, the work of ethics owners is never done.

Their work, given the tension described above, will always involve tradeoffs and difficult decisions. This is because of the primary tension that underlies their role, that of particularity and generality. Effectively, ethics owners work toward scalable solutions, generalizable rules and methods for ethical challenges that are inextricably contextual. However, the problem with this is that there are too many specifics—each case is unique, each company is unique—and while heuristics, or general rules, can scale, they must be fit to specific cases in practice. While experience can serve as a useful precedent, what lies between—the immediate issues at hand for individuals inside organizations addressing particular harm or risks—is hard, and requires specific interventions that cannot be determined by a heuristic process.

OPPORTUNITIES FOR ETHICS OWNERS

The work of ethics owners, as it is currently taking shape, is fully embedded within the fundamental assumptions of Silicon Valley. Ultimately, a comprehensive approach to addressing the potential for harms caused by the tech industry will have to challenge those assumptions. Ethics owners will be working from a partial set of approaches if the tech industry itself remains in the position of the entity that is best-suited to preventing the harm it creates. However, ethics owners can expand the conceptual tools at their disposal by bringing into question the capacity for technological fixes alone to minimize harms, and by placing the social consequences of technology on a different register as any financial accounting for the harm it might produce. Solutions to the problems ethics owners hope to address might cost money, might de-center technologists, and might be "bad" for business. But ethics owners, given the recent investments in strengthening their capacity to affect practices within Silicon Valley companies, are well-positioned to begin the push against this tide.

At the root of the challenges ethics owners face is the fact that while the tech industry is adept at producing scalable solutions, ethical harms remain tied to highly specific contexts. The following recommendations are intended to address this gap between scalability and specificity by acknowledging this gap as a defining characteristic of ethics owners' work. These recommendations are meant to offer ethics owners ways of addressing one of the biggest blind spots we observed in our research among tech companies: frequently, technologists sit together in a room and try to imagine how the lives of others, whose lives all too often look nothing like that of those in the room, might be affected by the products and services Silicon Valley builds. In light of this, our recommendations offer ways of bringing the experiences of those who aren't part of tech companies into the rooms where decision making happens.

It is just as important to change who is already *in* such rooms so that the people sitting around tables thinking hard about the ethical implications of products and services look, think, and experience the world in ways that are more like those who interact with those products and services. This goes far beyond the diversity and inclusion efforts that are already part of human resources initiatives in many Silicon Valley companies. As is backed up by years of research and

activism, diversity and inclusion initiatives have not been sufficient to disrupt the dominance of white and male hegemony inside the tech industry.[64] A meta-recommendation, then, is to address this problem squarely, and across the industry, but particularly within teams of 'ethics owners'. Arguably, the distributed and facilitative ethics functions we describe here are the site where diversity and inclusion initiatves could have the largest impact on shaping just and equitable product development.

Case Studies

Ethics owners need a robust body of **case studies** that generalize knowledge between companies. Excellent, general-purpose case studies already exist and are useful, but much work remains.[65] Ethics owners can actively produce case studies from their own domains that not only capture what did and did not work in the past, but help create real lines of communication between ethics owners in different companies and fields. Case studies should not be generic. Rather, they should highlight the particular, unique, and idiosyncratic details of the case, and highlight both the successful operationalizations of ethics owners' work *and* unsuccessful attempts. If case studies only report on ethics owners' successes, then the industry runs the risk of disseminating blind spots in their approaches, normalizing practices that produce failures, and failing to identify what could have been done differently. Sharing case studies, with enough practical details to be useful, and with enough documentation of potentially embarrassing failures or missteps, can be a challenging endeavor in its own right, in an industry characterized by trade secrets, cutthroat competition, and similar product offerings.

64 Safiya Umoja Noble and Sarah T. Roberts, eds., "Technological Elites, the Meritocracy, and Postracial Myths in Silicon Valley," *UCLA Previously Unpublished Works:* Preprint, 2019, 113–29, https://doi.org/10.1215/9781478003250-008.

65 Technology ethics cases. Markkula Center for Applied Ethics. Retrieved 2020, February 26. https://www.scu.edu/ethics/focus-areas/technology-ethics/resources/technology-ethics-cases/.

Informal Meetings

Ethics owners, working across the industry, have already created spaces where informal sharing of practices under the cover of secrecy has been possible. These spaces can serve as a model for the trust needed to collect the kinds of case studies that can improve ethical practices across the industry.[66] A recurring series of **informal meetings**, operating under Chatham House Rules that allow for employees to discuss issues frankly without having their remarks pre-cleared by a press department, are one model for sharing the broad base of knowledge ethics owners draw upon in their work at the requisite level of detail. When engineers, product managers, and ethics owners from different companies can meet and discuss their work, they can avoid some of the pitfalls that arise from their own, necessarily partial, perspective on problems.

At least one series of convenings like this already exists as "a process that structurally elevates those concerns, and then, two, tapping into larger communities so you can update your process with those concerns." One ethics owner who participates in the series says, "We're calling it the 'Data Ethics Salon,' so it's people interested in data ethics, in theory across the country, but the emerging working group is mostly anchored in San Francisco." These data ethics salons are hosted by various ethics owners, for other ethics owners and other tech workers with an interest in ethical issues, on a revolving basis at their employers' offices across the Bay Area, and are serving as a model that has been emulated in other cities with a strong concentration of tech companies. Other models include multi-stakeholder or cross-industry working groups that could serve as legally discrete entities capable of working across the barriers posed by nondisclosure agreements, trade secrets protection, or other proprietary protection of ethical decision-making processes.[67] In these settings, ethics owners could also strive to make their work, and that of related work inside their companies, accessible to scholarly, empirical research. Academic and independent scholars

66 As part of ongoing research and engagement with tech ethics practitioners, the authors of this report have regularly attended many such events, and have on occasion presented work as featured speakers at such events.

67 "Research, publications & initiatives," Partnership on AI, accessed February 26, 2020, https://www.partnershiponai.org/.

can serve an important coordinating function between forms of expertise at the core of ethics owners' work, and can help translate and frame issues that come from outside of companies into more tractable concerns for ethics owners to work with. Perhaps most importantly, collaborative settings such as salons and working groups can develop channels to civil society and advocacy groups that are deeply steeped in the impact of technology on society.

Partnership and Outreach

Citizen advocates and activists are closest to societal impact and ought to be closest to any solution. This is particularly true for understanding and addressing the racial and gendered impact of technologies. As Daniels, Nkonde, and Mir point out in "Advancing Racial Literacy in Tech" (2019), "Any turn toward ethics is an important one, it is incomplete without a racial lens." While already-ongoing conversations about ethics and race in technology have focused on pipeline issues, diversity and inclusion initiatives, and implicit bias training, ethics owners have the opportunity to push beyond these now-standard moves.[68] Developing ways of listening to and responsibly engaging with the concerns of those affected or potentially affected—particularly in historically disadvantaged groups—is crucial to any ethics practice that centers care and dignity over more narrow, quantified framings of risk and liability.[69]

Ethics owners, and their employers, can support the independent work of advocates who can surface actionable information about the social impacts of technology, which can then serve as the basis for interventions that ethics owners make in their own companies' corporate practices. Engagement with outside advocates ought to be recognized, compensated, and respected. Participatory development projects and community engagement frameworks have tended toward

68 Jessie Daniels, Mutale Nkonde, and Darakhshan Mir, "Advancing Racial Literacy in Tech: Why Ethics, Diversity in Hiring & Implicit Bias Trainings Aren't Enough," Data & Society Research Institute, 2019.

69 Sasha Costanza-Chock, "Sasha, Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice," *Proceedings of the Design Research Society*, June 3, 2018, https://ssrn.com/abstract=3189696.

de-politicizing the ethical stakes of many projects,[70] tokenizing the contributions of groups that hold relatively weaker positions in society,[71] and treating those in vulnerable positions as experimental subjects who bear the costs of product failures while they are still in test modes. Ethics owners are well-positioned to act as honest brokers between a company's concern not to divulge trade secrets or sensitive intellectual property, and the types of access outside advocates and academic researchers need. These outside perspectives are crucial for both the highly contextual, specific knowledge needed to address harms and the generalizable knowledge necessary for responsible, accountable tech development.

Accountability and Engagement

As we have described on multiple fronts, the cultural and technical infrastructures of Silicon Valley drive ethics owners toward practices centered on utilizing metrics. Technology companies favor metrics from top to bottom in their operations—metrics are used for rewarding staff, optimizing products, measuring product success, demonstrating diversity, etc. As a result, ethics owners are incentivized to use measurable criteria for demonstrating the value of investments in ethics to their organizations. But they need not feel limited to using measurable criteria as their sole points of leverage for bringing about change.

This is not to advocate for an anti-measurement stance—that would be a non-starter. Rather, it is to point out that measurement has a circularity that affirms its own value: technical systems can be optimized toward a measurable outcome in ways that have tangible, measurable effects on that system and therefore provide readily-visible validation for that type of intervention. This can create a deceptive impression that a particular measurable outcome is the right outcome on which to focus the organization's limited resources

70 Glyn Williams, "Evaluating participatory development: tyranny, power and (re) politicisation." *Third World Quarterly* 25, 3 (2004): 25:3, 557-578.

71 See: https://www.thestar.com/amp/news/gta/2019/10/25/indigenous-elder-slams-hollow-and-tokenistic-consultation-by-sidewalk-labs.html.

rather than some other qualitative outcome that may have more value to those affected by the technology. Furthermore, secondary metrics, like OKRs or key performance indicators (KPIs), can be attached to the primary outcome (i.e., performance metrics can be dependent on achieving algorithmic fairness metrics), reliably cementing the organization's attention.

Yet not everything that ethics owners are tasked with addressing is measurable. Indeed, the core ethical critiques of Silicon Valley are about the imbalance of power in society and the need for effective modes of accountability. These are fundamentally concerns around human values that cannot be resolved through measurement practices. Ethics owners should work to find and support ways of addressing important issues that are not amenable to measurement. Ethics owners, as they solidify their roles within an organizational structure, can be well-positioned to help facilitate the integration of community feedback into the development process. This can involve working with outside groups to detect, avoid, and mitigate unanticipated consequences of tech products and services. And it can also involve developing avenues for recourse and contestability so that tech development is not a one-way process that is deployed on a helpless public, but rather is a cooperative endeavor that integrates with society in ways that contribute to a common good.

CONCLUSION

Ethics owners, as we have discussed at length, occupy a difficult position. On one hand, they participate in the rich set of financial and social rewards Silicon Valley has reaped through producing products and services that reshape our lives. On the other hand, they must participate in acknowledging the serious harms and nearly unknowable sets of risks that these products and services sometimes create. The number of instances when the ethical implications of Silicon Valley technologies have created public outcry is unlikely to diminish in the coming months or years. At the same time, the pace of innovation and the relentless rollout of new products and services is similarly unlikely to slacken, given the existential pressure on companies large and small to maintain and grow market share.

The difficulty of their position is therefore not likely to be resolved, but rather will continue to require careful navigation through the tensions that characterize their work. Charting a course through such tensions can be aided by deliberate collaboration across and beyond the industry. If ethics owners can share their successes and failures, they will not have to rely on ethical practices from more distant domains (such as biomedical research) and will have better precedents to work from as they face new challenges. If ethics owners can reach out beyond their own companies to communities that are most directly affected by the products the tech industry builds, they will have better information about the scope and scale of the challenges they face, certainly, but also will have the opportunity to center the lives, and livelihoods, of those most deeply affected in working through those challenges.

ACKNOWLEDGMENTS

Portions of this research were conducted with support from the National Science Foundation (award # 1704425) and the Wenner-Gren Foundation. We are grateful to the interview participants who shared their time and experiences with us and the hosts of the many meetings and workshops we observed. This work benefited especially from collaborations and ongoing conversations with danah boyd, Sareeta Amrute, and Madeleine Claire Elish. We thank Ryan Hagen, Alondra Nelson, Ben Roome, Janet Vertesi, Karen Strassler, the Data & Society Raw Materials Seminar, the Data & Society AIGI team, Melissa Cefkin, Alex Hanna, and William Isaac for their thoughtful comments on previous drafts. We are grateful to Patrick Davison for his editorial prowess. We also wish to thank Sam Hinds, Natalie Kerby, Yichi Liu, Chris Redwood and the staff members at Data & Society whose work and expertise made this report possible.

REFERENCES

Adams, Janet S., Armen Tashchian, and Ted H. Shore. "Codes of Ethics as Signals for Ethical Behavior." *Journal of Business Ethics* 29, no. 3 (2001): 199–211.

Amrute, Sareeta. "Of Techno-Ethics and Techno-Affects." *Feminist Review* 123, no. 1 (November 2019): 56–73. https://doi.org/10.1177/0141778919879744.

AoIR. "Ethics," 2019. https://aoir.org/ethics/.

Arnold, Denis G., and Norman E. Bowie. "Respect for Workers in Global Supply Chains: Advancing the Debate Over Sweatshops." *Business Ethics Quarterly* 17, no. 1 (January 2007): 135–45. https://doi.org/10.5840/beq200717121.

Association for Computing Machinery. "ACM Ethics." ACM Ethics, July 17, 2018. https://ethics.acm.org/.

———. "ACM Updates Code of Ethics." Association for Computing Machinery, 2018, July 17. https://www.acm.org/articles/bulletins/2018/july/new-code-of-ethics-released.

Bahrami, Homa, and Stuart Evans. "Super-Flexibility for Real-Time Adaptation: Perspectives from Silicon Valley." *California Management Review* 53, no. 3 (May 2011): 21–39. https://doi.org/10.1525/cmr.2011.53.3.21.

Barabas, Chelsea. "Beyond Bias: Re-Imagining the Terms of 'Ethical AI' in Criminal Law." *SSRN Electronic Journal*, 2019. https://doi.org/10.2139/ssrn.3377921.

Benjamin, Ruha, ed. *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life*. Durham, NC: Duke University Press, 2019.

———. *Race After Technology: Abolitionist Tools for the New Jim Code*. Medford, MA: Polity Press, 2019.

Birhane, Abeba. "Descartes Was Wrong: 'A Person Is a Person through Other Persons,'" *Aeon*, April 7, 2017. https://aeon.co/ideas/descartes-was-wrong-a-person-is-a-person-through-other persons.

Birhane, Abeba, and Fred Cummins. "Algorithmic Injustices: Towards a Relational Ethics." *ArXiv:1912.07376 [Cs]*, December 16, 2019. http://arxiv.org/abs/1912.07376.

Birhane, Abeba, and Jelle van Dijk. "Robot Rights? Let's Talk about Human Welfare Instead." *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, February 7, 2020, 207–13. https://doi.org/10.1145/3375627.3375855.

Bloss, Cinnamon, Camille Nebeker, Matthew Bietz, Deborah Bae, Barbara Bigby, Mary Devereaux, James Fowler, et al. "Reimagining Human Research Protections for 21st Century Science." *Journal of Medical Internet Research* 18, no. 12 (December 22, 2016): e329. https://doi.org/10.2196/jmir.6634.

Brandom, Russell. "Shadow profiles are the biggest flaw in Facebook's privacy defense." *The Verge*, April 11, 2018. https://www.theverge.com/2018/4/11/17225482/facebook-shadow-profiles-zuckerberg-congress-data-privacy.

Braudel, Fernand. *Afterthoughts on Material Civilization and Capitalism*. Johns Hopkins University Press, 1977.

Brown, Sam. "Consequence Scanning: An Agile Event for Responsible Innovators." London: doteveryone, 2019.

Browne, Simone. *Dark Matters: On the Surveillance of Blackness*. Durham: Duke University Press, 2015.

Canary, Heather E., and Marianne M. Jennings. "Principles and Influence in Codes of Ethics: A Centering Resonance Analysis Comparing Pre- and Post-Sarbanes-Oxley Codes of Ethics." *Journal of Business Ethics* 80, no. 2 (June 2008): 263–78. https://doi.org/10.1007/s10551-007-9417-1.

Castilla, Emilio J. "Gender, Race, and Meritocracy in Organizational Careers." *American Journal of Sociology* 113, no. 6 (2008): 1479–1526.

Castilla, Emilio J., and Stephen Benard. "The Paradox of Meritocracy in Organizations." *Administrative Science Quarterly* 55, no. 4 (December 2010): 543–676. https://doi.org/10.2189/asqu.2010.55.4.543.

Committee on Envisioning the Data Science Discipline: The Undergraduate Perspective, Computer Science and Telecommunications Board, Board on Mathematical Sciences and Analytics, Committee on Applied and Theoretical Statistics, Division on Engineering and Physical Sciences, Board on Science Education, Division of Behavioral and Social Sciences and Education, and National Academies of Sciences, Engineering, and Medicine. *Data Science for Undergraduates: Opportunities and Options*. Washington, D.C.: National Academies Press, 2018. https://doi.org/10.17226/25104.

Cope, Bill, and Mary Kalantzis. "Big Data Comes to School: Implications for Learning, Assessment, and Research." *AERA Open* 2, no. 2 (April 2016): 233285841664190. https://doi.org/10.1177/2332858416641907.

Costanza-Chock, Sasha. "Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, June 3, 2018. https://papers.ssrn.com/abstract=3189696.

Data for Black Lives. "Data for Black Lives COVID-19 Movement Pulse Check and Roundtable Report." April 2020, https://d4bl.org/reports.html.

Daniels, Jessie. "'My Brain Database Doesn't See Skin Color' Color-Blind Racism in the Technology Industry and in Theorizing the Web." *American Behavioral Scientist*, no. 11 (2015): 35.

Daniels, Jessie, Mutale Nkonde, and Darakhshan Mir. "Advancing Racial Literacy in Tech: Why Ethics, Diversity in Hiring & Implicit Bias Trainings Aren't Enough." Data & Society Research Institute, 2019.

Deahl, Dani. "Google Employees Demand the Company Pull Out of Pentagon AI Project." *The Verge*, April 4, 2018. https://www.theverge.com/2018/4/4/17199818/google-pentagon-project-maven-pull-out-letter-ceo-sundar-pichai.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. Strong Ideas Series. Cambridge, Massachusetts: The MIT Press, 2020.

Doerr, John E. *Measure What Matters: How Google, Bono, and the Gates Foundation Rock the World with OKRs*. New York: Portfolio/Penguin, 2018.

Duff, Alistair S. "Rating the Revolution: Silicon Valley in Normative Perspective." *Information, Communication & Society* 19 (11): 1605–21.

Dunn, Alix. "Agile Ethics: Managing Ethical Complexity in Technology." *Medium*, March 15, 2019. https://medium.com/@alixtrot/agile-ethics-managing-ethical-complexity-in-technology-63131f7bab95.

Eberhardt, Jennifer. "Can Airbnb Train Hosts Not to Be Racists?" *Daily Beast*, June 12, 2019. https://www.thedailybeast.com/can-airbnb-train-hosts-not-to-be-racists.

Eckhouse, Laurel, Kristian Lum, Cynthia Conti-Cook, and Julie Ciccolini. "Layers of Bias: A Unified Approach for Understanding Problems With Risk Assessment." *Criminal Justice and Behavior* 46, no. 2 (February 1, 2019): 185–209. https://doi.org/10.1177/0093854818811379.

Elish, Madeleine Clare, and Elizabeth A. Watkins. "When Humans Attack: Re-Thinking Safety, Security, and AI." *Data & Society: Points* (blog), May 14, 2019. https://points.datasociety.net/when-humans-attack-re-thinking-safety-security-and-ai-b7a15506a115.

"Ethical OS Toolkit." Omidyar Network, 2018. https://ethicalos.org/wp-content/uploads/2018/08/Ethical-OS-Toolkit-2.pdf.

Feldman, Martha S., and Brian T. Pentland. "Reconceptualizing Organizational Routines as a Source of Flexibility and Change." *Administrative Science Quarterly* 48, no. 1 (March 2003): 94. https://doi.org/10.2307/3556620.

Freeman, R. "Managing for stakeholders: Trade-offs or value creation." *Journal of Business Ethics* 96, no. 1 (2010): 7–9.

Gawande, Atul. *The Checklist Manifesto: How to Get Things Right*. New York, NY: Picador, 2011.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. "Datasheets for Datasets." *ArXiv:1803.09010 [Cs]*, March 19, 2020. http://arxiv.org/abs/1803.09010.

Giridharadas, Anand. *Winners Take All: The Elite Charade of Changing the World*. New York: Alfred A. Knopf, 2018.

"GitHub - Drivendataorg/Deon: A Command Line Tool to Easily Add an Ethics Checklist to Your Data Science Projects." Accessed July 30, 2020. https://github.com/drivendataorg/deon.

Green, Ben, and Salomé Viljoen. "Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought." In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency.*, 13. Barcelona, ES: ACM, 2020.

Greenberg, Andy. "Security Isn't Enough. Silicon Valley Needs 'Abusability' Testing." *Wired*, January 28, 2019. https://www.wired.com/story/abusability-testing-ashkan-soltani/.

Greene, Daniel, Anna Lauren Hoffmann, and Luke Stark. "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*, 2019, 2122–2130. https://scholarspace.manoa.hawaii.edu/bitstream/10125/59651/0211.pdf.

Harvey, David. *A Brief History of Neoliberalism*. New York: Oxford University Press, 2005.

Hines, Alicia Hedlam. *Technicolor: Race, Technology, and Everyday Life*. Edited by Thuy Linh N. Tu and Alondra Nelson. New York London: NYU Press, 2001.

Ho, Karen. *Liquidated: An Ethnography of Wall Street*. Durham, NC: Duke University Press, 2009.

Holstein, Kenneth, Jennifer Wortman Vaughan, Hal Daumé III, Miro Dudík, and Hanna Wallach. "Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?" *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 2019, 1–16. https://doi.org/10.1145/3290605.3300830.

Hooker, John, and Tae Wan Kim. "Ethical Implications of the Fourth Industrial Revolution for Business and Society." In *Business and Society 360*, edited by David M. Wasieleski and James Weber, 3:35–63. Emerald Publishing Limited, 2019. https://doi.org/10.1108/S2514-175920190000003002.

Hulkower, Raphael. "The History of the Hippocratic Oath: Outdated, Inauthentic, and yet Still Relevant." *Einstein Journal of Biology and Medicine* 25, no. 1 (2016): 41–44.

IEEE. "7.8 IEEE code of Ethics." Accessed February 26, 2020. https://www.ieee.org/about/corporate/governance/p7-8.html.

———. "IEEE SA - The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems." Accessed July 30, 2020. https://standards.ieee.org/industry-connections/ec/autonomous-systems.html.

Jackman, Molly, and Lauri Kanerva. "Evolving the IRB: Building Robust Review for Industry Research," *Washington and Lee Law Review Online* 72, 3, (2016). https://scholarlycommons.law.wlu.edu/cgi/viewcontent.cgi?article=1042&context=wlulr-online

Jacobs, Abigail Z., and Hanna Wallach. "Measurement and Fairness." *ArXiv:1912.05511 [Cs]*, December 11, 2019. http://arxiv.org/abs/1912.05511.

Jensen, Michael C. "Value Maximization, Stakeholder Theory, and the Corporate Objective Function." *Business Ethics Quarterly* 12, no. 2 (2002): 235–56. https://doi.org/10.2307/3857812.

Jobin, Anna, Marcello Ienca, and Effy Vayena. "Artificial Intelligence: The Global Landscape of Ethics Guidelines," *Nature Machine Intelligence* 1, 389–399(2019). https://www.nature.com/articles/s42256-019-0088-2.

Katz, Yarden. "Manufacturing an Artificial Intelligence Revolution." *SSRN Electronic Journal*, 2017. https://doi.org/10.2139/ssrn.3078224.

Kenneally, Erin, and David Dittrich. "The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research," 2012. https://doi.org/10.2139/SSRN.2445102.

Kent Beck, Mike Beedle, Arie van Bennekum, Alistair Cockburn, Ward Cunningham, Martin Fowler, James Grenning, et al. "Manifesto for Agile Software Development." Accessed July 30, 2020. https://agilemanifesto.org/.

Khan, Gohar Feroz, Bobby Swar, and Sang Kon Lee. "Social Media Risks and Benefits: A Public Sector Perspective." *Social Science Computer Review* 32, no. 5 (October 2014): 606–27. https://doi.org/10.1177/0894439314524701.

Kolhatkar, Sheelah. "The Economist Who Put Stock Buybacks Back in Washington's Crosshairs." *The New Yorker*, June 20, 2019. https://www.newyorker.com/business/currency/the-economist-who-put-stock-buybacks-in-washingtons-crosshairs.

Langville, Amy N., and Carl D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton: Princeton University Press, 2012.

Littler, Jo. *Against Meritocracy: Culture, Power and Myths of Mobility*. London ; New York, NY: Routledge/Taylor & Francis Group, 2017.

Mak, Aaron. "When Is Technology Too Dangerous to Release to the Public?" *Slate*, February 22, 2019. https://slate.com/technology/2019/02/openai-gpt2-text-generating-algorithm-ai-dangerous.html.

Marchi, Regina M. "With Facebook, Blogs, and Fake News, Teens Reject Journalistic 'Objectivity,'" *Journal of Communication Inquiry* 36, 3 (2012). https://doi.org/10.1177/0196859912458700.

Marwick, Alice. "Silicon Valley Isn't a Meritocracy. And It's Dangerous to Hero-Worship Entrepreneurs." *Wired*, November 25, 2013. https://www.wired.com/2013/11/silicon-valley-isnt-a-meritocracy-and-the-cult-of-the-entrepreneur-holds-people-back/.

———. *Status Update: Celebrity, Publicity, and Branding in the Social Media Age*. Yale University Press, 2013. https://www.jstor.org/stable/j.ctt5vkzxr.

McGreal. "A code of ethics for software." Scrum.org. October 31, 2016. https://www.scrum.org/resources/blog/code-ethics-software.

McIlwain, Charleton. *Black Software: The Internet & Racial Justice, from the AfroNet to Black Lives Matter*. New York: Oxford University Press, 2019.

Metcalf, Jacob. "Big Data Analytics and Revision of the Common Rule." *Communications of the ACM* 59, no. 7 (June 24, 2016): 31–33. https://doi.org/10.1145/2935882.

Metcalf, Jacob, and Kate Crawford. "Where Are Human Subjects in Big Data Research? The Emerging Ethics Divide." *Big Data & Society* 3, no. 1 (June 1, 2016): 2053951716650211. https://doi.org/10.1177/2053951716650211.

Metcalf, Jacob, Emanuel Moss, and danah boyd. "Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics." *Social Research* 86, no. 2 (2019): 449–76.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. "Model Cards for Model Reporting." *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, 220–29. Atlanta, GA, USA: ACM Press, 2019. https://doi.org/10.1145/3287560.3287596.

Mittelstadt, Brent. "AI Ethics – Too Principled to Fail?" *SSRN Electronic Journal*, 2019. https://doi.org/10.2139/ssrn.3391293.

Mohr, John W. "Implicit Terrains: Meaning, Measurement, and Spatial Metaphors in Organizational Theory." In *Constructing Industries and Markets*, edited by Joseph F. Porac and Marc Ventreca. Emerald Group Publishing, 2007.

Mounk, Yascha. *The Age of Responsibility: Luck, Choice and the Welfare State*. Cambridge, MA: Harvard University Press, 2017.

Nedzhvetskaya, Nataliya, and J. S. Tan. "What We Learned from over a Decade of Tech Activism." *The Guardian*, December 23, 2019, sec. Opinion. https://www.theguardian.com/commentisfree/2019/dec/22/tech-worker-activism-2019-what-we-learned.

Nidhra, Srinivas. "Black Box and White Box Testing Techniques - A Literature Review." *International Journal of Embedded Systems and Applications* 2, no. 2 (June 30, 2012): 29–50. https://doi.org/10.5121/ijesa.2012.2204.

Noble, Safiya Umoja, and Sarah T. Roberts, eds. "Technological Elites, the Meritocracy, and Postracial Myths in Silicon Valley." *UCLA Previously Unpublished Works: Preprint*, 2019, 113–29. https://doi.org/10.1215/9781478003250-008.

Ogbonnaya-Ogburu, Ihudiya Finda, Angela D.R. Smith, Alexandra To, and Kentaro Toyama. "Critical Race Theory for HCI." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. Honolulu HI USA: ACM, 2020. https://doi.org/10.1145/313831.3376392.

Padilla, Thomas. *Responsible Operations: Data Science, Machine Learning, and AI in Libraries*, Dublin OH: OCLC, 2019.

Partnership on AI. "Research, publications & initiatives." Accessed February 26, 2020. https://www.partnershiponai.org/.

Passi, Samir, and Steven J. Jackson. "Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects." *Proceedings of the ACM on Human-Computer Interaction* 2, no. CSCW (November 1, 2018): 1–28. https://doi.org/10.1145/3274405.

Pichai, Sundar. "AI at Google: Our Principles." Google Blog, June 7, 2018. https://www.blog.google/technology/ai/ai-principles/.

Rajendran, Jeyavijayan, Vinayaka Jyothi, and Ramesh Karri. "Blue Team Red Team Approach to Hardware Trust Assessment." In *2011 IEEE 29th International Conference on Computer Design (ICCD)*, 285–88. Amherst, MA, USA: IEEE, 2011. https://doi.org/10.1109/ICCD.2011.6081410.

Raji, Inioluwa Deborah, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. "Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing." *ArXiv:2001.00964 [Cs]*, January 3, 2020. http://arxiv.org/abs/2001.00964.

Raji, Inioluwa Deborah, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. "Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing." In *Conference on Fairness, Accountability, and Transparency (FAT* '20)*, 12. Barcelona, ES, 2020.

Rose, Jacob M. "Corporate Directors and Social Responsibility: Ethics versus Shareholder Value." *Journal of Business Ethics* 73, no. 3 (July 1, 2007): 319–31. https://doi.org/10.1007/s10551-006-9209-z.

Sample, Ian. "Maths and Tech Specialists Need Hippocratic Oath, Says Academic." *The Guardian*, August 16, 2019, sec. Science. http://www.theguardian.com/science/2019/aug/16/mathematicians-need-doctor-style-hippocratic-oath-says-academic-hannah-fry.

Sayer, Andrew. "Dignity at Work: Broadening the Agenda." *Organization* 14, no. 4 (July 2007): 565–81. https://doi.org/10.1177/1350508407078053.

Simonite, Tom. "Should Data Scientists Adhere to a Hippocratic Oath?" *Wired*, February 8, 2018. https://www.wired.com/story/should-data-scientists-adhere-to-a-hippocratic-oath/.

Sine, Wesley D. "Organization Science." In *In Search of Research Excellence*, by Ronald Mitchell and Richard Dino, 14204. Edward Elgar Publishing, 2011. https://doi.org/10.4337/9781849807630.00024.

Stark, Laura Jeanine Morris. *Behind Closed Doors: IRBs and the Making of Ethical Research*. Morality and Society Series. Chicago ; London: The University of Chicago Press, 2012.

Stevens, Mitchell L. "An Ethically Ambitious Higher Education Data Science." *Research & Practice in Assessment* 9 (2014).

Subramanian, Ajantha. *The Caste of Merit: Engineering Education in India*. Harvard University Press, 2019. https://doi.org/10.2307/j.ctvr33ddc.

Suchman, Lucy. "Located Accountabilities in Technology Production." *Scandanavian Journal of Information Systems* 14, no. 2 (2002): 16.

Sundaram, A. and Inkpen, A. "The corporate objective revisited." *Organization Science* 15, no. 3 (2004): 350-363.

Markkula Center for Applied Ethics. "Technology Ethics Resources - Markkula Center for Applied Ethics." Accessed July 30, 2020. https://www.scu.edu/ethics/focus-areas/technology-ethics/resources/.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. "Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems, First Edition." Cham: IEEE, 2019. https://doi.org/10.1007/978-3-030-12524-0_2.

Upadhya, R. K. "Tech Workers against Imperialism." *Tech Workers Coalition* (blog), November 22, 2018. https://medium.com/tech-workers-coalition/tech-workers-against-imperialism-2d8024e461a7.

Varshney, Kush R. "Introducing AI Fairness 360, A Step Towards Trusted AI - IBM Research." IBM Research Blog, September 19, 2018. https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/.

Veen, Christiaan van. "Artificial Intelligence: What's Human Rights Got To Do With It?" *Data & Society: Points* (blog), May 18, 2018. https://points.datasociety.net/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d5.

Wagner, Ben. "Ethics As An Escape From Regulation. From 'Ethics-Washing' To Ethics-Shopping?" In *BEING PROFILED*, edited by Emre Bayamlioglu, Irina Baraliuc, Liisa Albertha Wilhelmina Janssens, and Mireille Hildebrandt, 84–89. Amsterdam: Amsterdam University Press, 2019. https://doi.org/10.1515/9789048550180-016.

Warren, May. "Indigenous Elder Slams 'Hollow and Tokenistic' Consultation by Sidewalk Labs." *Toronto Star*. October 25, 2019, sec. GTA. https://www.thestar.com/news/gta/2019/10/25/indigenous-elder-slams-hollow-and-tokenistic-consultation-by-sidewalk-labs.html.

West, Dave. "Why Tech Companies Need a Code of Ethics for Software Development." *Entrepreneur*, April 19, 2018. https://www.entrepreneur.com/article/311410.

Williams, Glyn. "Evaluating Participatory Development: Tyranny, Power and (Re)Politicisation." *Third World Quarterly* 25, no. 3 (2004): 557–78.

Young, Michael. *The Rise of the Meritocracy*. New Brunswick, N.J., U.S.A: Transaction Publishers, 1994.

DATA& SOCIETY

Data & Society is an independent nonprofit research institute that advances new frames for understanding the implications of data-centric and automated technology. We conduct research and build the field of actors to ensure that knowledge guides debate, decision-making, and technical choices.

www.datasociety.net
@datasociety
Illustration by Yichi Liu

DATA&
SOCIETY

www.datasociety.net
@datasociety