Resistance and refusal to algorithmic harms: Varieties of ‘knowledge projects’

Maya Indira Ganesh
University of Cambridge, UK

Emanuel Moss
Data & Society Research Institute, USA
Cornell Tech Digital Life Initiative, USA

Abstract
Industrial, academic, activist, and policy research and advocacy movements formed around resisting ‘machine bias’, promoting ‘ethical AI’, and ‘fair ML’ have discursive implications for what constitutes harm, and what resistance to algorithmic influence itself means, and is deeply connected to which actors makes epistemic claims about harm and resistance. We present a loose categorization of kinds of resistance to algorithmic systems: a dominant mode of resistance as ‘filtering up’ and being translated into design fixes by Big Tech; and advocacy and scholarship which bring a critical frame of lived experiences and scholarship around algorithmic systems as socio-technical entities. Three recent cases delve into how Big Tech responds to harms documented by marginalized groups; these highlight how harms are valued differently. Finally, we identify modes of refusal that recognize the limits of Big Tech’s resistance; built on practices of feminist organizing, decoloniality, and New-Luddism, they encourage a rethinking of the place and value of technologies in mediating human social and personal life; and not just how they can deterministically ‘improve’ social relations.

Keywords
Algorithmic, resistance, refusal, harms, epistemology, big-tech, design, bias

Introduction
AIAAIC is an independent, nonpartisan, open Google spreadsheet that painstakingly documents, verifies, and classifies incidents related to AI, algorithmic, and automation harms from around the world. At the time of writing, there were 808 rows of familiar cases of algorithmic harms
from 2014 to the present day, like semi-autonomous car crashes and racial bias resulting from faulty computer vision. The spreadsheet also has less well-known cases, such as how algorithmic errors in the Kronos automated time-tracking and scheduling software significantly disrupted the lives of low-income workers, particularly single parents, working for big retail chains like Starbucks (Kantor, 2014). Kronos’s algorithmic system was scheduling a single mother, a Starbucks barista, to report to work at 8am, which meant that she and her four-year-old son had to wake up at 5am for a three hour commute to childcare and then to work. Other cases of Kronos’ application read similarly, indicating that this software uses an algorithmic model that cannot privilege the lives and realities of workers over its own functionality. Roughly over 10,000 workers have reported being negatively affected by this software. Almost immediately after reporting by the New York Times, Starbucks retracted its use of Kronos.

The growing list of cases in the AIAAIC repository establishes the significant scale of algorithmic harms, and provokes us to ask what it means to respond to harms, and what these responses imply in discursive terms for social, academic, and policy management of algorithmic manipulation. In this paper we distinguish between two broad varieties of responses to algorithmic harms: how Big Tech—the chief actor of algorithmic harms—has quickly adopted a stance of solidarity to resist them; and refusals that emerge from beyond Big Tech and raise philosophical provocations about what it means to construct life outside the ambit of Big Tech-architected harms and mitigations.

What is distinct about the Kronos case from that of other incidents of algorithmic harms that we could open this article with? For one thing, the Kronos case was identified and addressed because of journalistic inquiry, which is one public approach from civil society to resist algorithmic manipulation. The other aspect was that it relates to someone whose life under conditions of poverty was aggravated by this algorithmic system. The range of harms listed in AIAAIC include problems such as accuracy, opacity of statistical models, appropriateness of the application, lack of consent from a community. The Kronos case evokes what the sociologist and information scholar Oscar Gandy means by rational—or, statistical—discrimination (2009). In our effort to minimize ‘irrational’ decision-making, thought to lack credibility because it is the result of “broad generalizations rather than [by] careful weighing of the relevant facts” (Gandy, 2009: 32), there has been a turn to automated statistical analysis, a “process of extensive search, reflection, and analysis” believed to result in more efficient, or rational, decision-making (ibid). But, Gandy shows, such reliance on statistics that minimizes ‘irrational’ bias also necessarily requires the normalization of the “massive impacts of system-level biases and blind spots with regard to structural impediments that magnify the impact that disparities in starting position will have on subsequent opportunities.” (Op Cit: 33).

In other words, unlike better-known cases of hostile, algorithmically-inspired sexism and racism resulting in disadvantage to a distinct, protected group of individuals, here, seemingly neutral applications of algorithmic systems in non-neutral environments have perpetuated existing systemic disadvantages. Gandy problematizes bias as flattened into just how society is, as systemic biases, rather than the understanding of bias as clearly identifiable acts of discrimination against a group of individuals that can be, eventually, designed out. In the context of the disruptive Kronos software, to address the bias faced by the single-mother Starbucks barista would be to confront the lack of adequate childcare, housing far from the center of the city, demands made of workers, their social security nets (or lack thereof), and possibly many other factors associated with the low socio-economic status that compounds and enables these factors. Now, social disadvantage is not the harm that algorithmic systems enact, but is the substrate that algorithmic systems work on and exacerbate. As we discuss in this paper, algorithmic harms like bias or unfairness have become the subject of resistance that Big Tech is now also part of. Yet, their attention to what this means does not relate to this exacerbation of the social biases Gandy refers to. A different
set of political and ethical positions however take up refusals of the social, economic, and political conditions of life itself in highly data-fied societies.

This distinction between refusals and resistance can be understood in terms of how Greene, Hoffman and Stark leverage the notion of ‘moral background’ to go beyond the normative or deterministic outcomes of correcting unethical practices, or “business disasters”, towards “a specific arrangement of second-order social assumptions about what ethics mean and how they work, above first-order claims about ethical norms or behaviors.” (2019: 2124) This distinction is valuable because we identify the contrasting loci of action by various actors, chiefly within Big Tech, responding to algorithmic manipulation, and the outcomes of these actions.

We draw attention to how Big Tech’s interventions in addressing algorithmic harms validate particular kinds of harms, re-inscribe their own influence, and thus re-configure what it means to resist algorithms. Big Tech’s actions to translate resistance to algorithmic manipulation through software re-design has resulted in a rather narrow frame of activities that coalesce around managing fairness and bias. This motivates the question of if the internalization of critiques of algorithmic systems have actually altered designed-in harms, or if it has been co-opted to fit existing engineering-friendly frames of fairness, ethics, and bias? We argue the latter, that engineering-friendly frames are an adoption of the language of resistance to algorithmic manipulation, and narrowly so.

One outcome of particular kinds of algorithmic harm being named and identified as ‘disasters’ is that who is being harmed, how that harm is named, by whom, and what gets adopted and made central to efforts at correction end up being considered as harmful. Social conditions like gender inequality or poverty tend not to be, however, it is encouraging that harms like racism are being corrected. This offers hope. So the AIAAIC spreadsheet attempts to expand the understanding of harms by documenting instances from around the world and in different applications. Refusals, by contrast, ally themselves with endemic, inter-connected societal, infrastructural, and political problems, in which small but firm tactics or actions speak to, and of, histories and cultures of social relations.

We understand resistance and refusal as ‘knowledge projects’; by studying resistance and refusal, we can identify what counts as valid kinds of knowledge—of harm, scale, change, impact, ethics—who establishes them as valid and how they do so. These work at multiple levels in terms of framing and shaping imaginaries of how to correct existing social systems and create different or better ones; but also in terms of what is practical, valuable, and effective. Thus we assemble different practices of resistance and refusal by various actors as loose groupings, varieties, or taxons—rudimentary groupings that we believe might constitute future, fuller taxonomies that other scholars might also work on. In doing so, we are not seeking a direct comparison between practices and actors of resistance and those of refusal; rather, we want to highlight the implications of contrasts between them for technology policy, design, and use. For instance, that refusals might identify terrains for existing outside of algorithmic systems but do not necessarily offer distinct protocols or steps except practices of living itself. While difficult to adopt as practical, scalable, accountable, or enforceable, refusals provoke reflection on established values of knowledge, relations, and time; which, we believe, serve as philosophical prompts to assess the terms of social progress, justice, and well-being.

This essay proceeds as follows. First, we describe ‘power/knowledge’ as a theoretical influence that allows us to identify the interplay of material, situated, historic, and political-economic practices of powerful technologists that accrue to what we eventually understand resistance to be and mean. It is only fitting then that we also identify our own situations, and we offer this as our methodology. Then we turn to a key taxon in our assembly of varieties of responses to algorithmic manipulation, how change or improvement emerge from ‘Big Tech’, ‘platform companies’, and/or ‘Silicon Valley’. In ‘filtering up’ to Big Tech, resistance is transformed into computational or
interface design specifications that can scale along with the algorithmic modalities and logics in order to be successful. While this does not challenge the logics of Big Tech, it does portend, possibly, a dissolving of critical concerns into standard product features.

Then, we explore the limits of resistance by arguing that the telos of resistance always remains within the ambit of Silicon Valley logics and the power of Big Tech companies to make changes to their own practices. We emphasize this through a brief detour through three separate kinds of harms—online harassment, challenging the right to encryption, and racist language—that Big Tech has intervened to address, and with mixed outcomes. Onto-epistemic dynamics are at work here, with online harassment being framed as a niche social problem to be fixed by technical design, whereas encryption and anonymity are values that can be defended through policy and technical design. Racist language in computing organizations gets addressed very quickly in the year of worldwide protests against the killing of George Floyd. However, eventually, affected communities and their allies do not look explicitly to Big Tech to architect changes, but seek to influence the discursive constructions of ‘resistance’ through interdisciplinary education, research and practice. These are ‘good faith’ attempts to bring sociological, socio-technical and lived experiences to the (re)framing of resistance. We understand our own work as scholars and practitioners in this light.

So, while ‘resistance’ might attempt to improve or reform current socio-technical systems, ‘refusal’ alerts us to how we must consider the quality and experiences of life under conditions beholden to algorithmic logics. We find entwined strands of new Luddism, feminism, and decoloniality that negate the possibilities for repair. These iterations argue for conditions of everyday life that lie outside of machine logics. They decline the power of Big Tech over social, biological, and interior life. However, refusal is also being quickly co-opted without concerted efforts to act and effect change. We conclude by urging for varied, critical, and creative practices to address everyday life outside and alongside digital technologies.

**Power/knowledge**

Resistance to algorithmic systems relies on measurement and documentation of harms, often requiring computational science skill sets. Resistance understood within the imaginary of the law and policy extends to questions of fairness within the letter of local laws, or in setting out the lines around what a minimally-not-invasive, minimally-acceptable technology would look like. We understand ‘resistance’ to algorithmic systems as implicated in power/knowledge (Foucault, 1980), so our intervention is a critical reflection on expert and epistemic communities, which ones are likely to be successful and why, and how these move practices of justice and equity forward, or not. Here, by ‘power/knowledge’, we refer to the intersections of practices of scientific and technical knowledge-making that constitute the truth of algorithmic harms, and the social, cultural, and political institutions, discourses, and social relations that enable this knowledge to take shape in the world as truth. This entails significant kinds of social, political, industrial, and epistemic power. In other words, we are interested in the dispositif, the complex fabric of historical and institutional practices that constitute worlds, as well as the devices or instruments that identify and quantify harms, thereby legitimizing them. How we measure and know something also implicates the contexts in which that measure is created in the first place.

In taking the entwining of knowledge-making with power, we identify both resistance and refusal as political in epistemic terms: it suggests that some kinds of knowledge are limited, that other ways of knowing are possible. But, as we show, knowledge is valued differently, so not all kinds of knowledge that resists is considered valuable. We are inspired by two strands of related scholarship. One is that which is tied to Michel Foucault, who was a meticulous scholar
of over three hundred years of European society in almost all its institutional, philological, meta-epistemological, and interpersonal forms. The other is that which is tied to the eminent feminist technoscientist, Karen Barad, who organizes her complex techno scientific explorations of the state of reality and existence through a critical re-reading of one of the most fundamental experiments in Western science of the past century: Niels Bohr’s two-slit gedanken experiment that took on the ontology of light itself (Barad, 2007). Yet, both significantly contribute to our study of apparatuses concerning the production of knowledge. Barad’s work addresses nonhuman technologies, like the ultrasound scanner, and laboratory practices of quantum physics; and is aligned with a socio-technical approach that attends to how technologies actually work, and how in that process they become potent epistemological forces. Her work identifies how practices of measurement, employed as epistemological instruments, necessarily create the worlds they purport to objectively evaluate, and which come into being through representational practices of language and quantification. The act of world-making cannot, for Barad, be considered apart from the ethical and political implications of the socio-technical construction of reality (Barad, 2011). This onto-ethico-politico-epistemology, taken with the stakes of Foucauldian knowledge/power allows us to recognize knowledge-making and its instruments as always situated.

Situating this work

This paper emerges from multiple streams of the authors’ individual and shared research practice. We have the common experience of working with and contributing scholarly work to the Fairness, Accountability and Transparency conference (FAccT) between 2017-2020 in the capacity of doctoral students in the Humanities. One of us has served on a program track at FAccT to integrate Social Sciences and Humanities scholarship into what is predominantly a Computer Science and Law conference; and finding gaps and challenges in establishing interdisciplinary practices of knowledge-making. Experiences of working within the FAccT frame have given us an opportunity to assess the limits of this discourse as it relates to other movements for technology justice and equity we have been associated with. For instance, one of us comes to this critique through a history of feminist organizing and knowledge-making through resistance to gendered online harassment on big social media platforms. This offers a parallel view of how practices of resistance operate and do not scale when emerging from a traditionally de-centered and marginal space. As such, this paper is built on reflections from primary research and ethnographic fieldwork.

Varieties of resistance

Resistance flowing upstream: Big tech mitigates algorithmic harms

Silicon Valley has made an indisputable effort to be seen as responding to resistance. For almost half a decade now, industry investments in ‘algorithmic fairness’ and ‘AI ethics’ have supported academic and industrial research, dedicated organizational divisions within the largest companies to incorporate that research into products, and backed venture capital for startups turning ethics and fairness into consulting services for enterprise clients. The net effect of these investments is to have transformed ethics and fairness from contested values about how best to live life and organize society into components of the product development process. In the absence of any other regulatory limits, this ultimately places Silicon Valley executives and other industry technologists as the final arbiters of ethical contestation.
This transformation of resistance into software development methodology is most obvious in the case of algorithmic fairness, but has proven impressively adaptable to a wide range of resistance movements. Algorithmic fairness can be traced back to a niche academic computer science research topic nested within privacy and security research (see Dwork et al., 2012). But a germinal investigative journalism project, Machine Bias (Angwin et al., 2016), revealed to a wider world how algorithmic systems can replicate and magnify racial biases in criminal justice. This reporting was followed by subsequent work that demonstrated a host of harms produced by gendered and racialised biases across many domains in which algorithmic decision systems were being used. Rather than questioning whether or not the use of algorithms was appropriate within these domains given the potential harms, the tech industry developed a massive body of work oriented toward being able to demonstrate compliance with civil rights legislation prohibiting discrimination of certain protected features (race, gender, sexual preference, etc.) in domains regulated by those laws (housing, employment, criminal justice, finance, etc.).

A similar process can be observed around AI ethics. Trenchant outside critiques of the use of artificial intelligence technologies in national security, corporate surveillance, and media manipulation have been transformed into organizational practices meant to make design tradeoffs that blunt resistance without compromise shareholder values (Moss and Metcalf, 2020). But most importantly, they shift ethics from a zone of contestation that implicates a wide range of solidarities—publics, government institutions, advocacy organizations—into a technology development process over which Silicon Valley companies hold dominion. For example, the Omidyar Foundation-funded ‘Ethical OS Toolkit’ suggests to senior executives in venture capital and Big Tech that they can be ethical through adopting a set of design procedures, an approach that is both deterministic and systemic. It is systemic in that as top-down education it recognizes that there is a need to change business practices by influencing leadership; however such attention is also deterministic and cannot address the wider political, economic, business, organizational, human and nonhuman infrastructures that sustain leadership. ‘Ethics’ becomes a practical set of tools and processes, a to-do checklist rather than a moment for critical reflection on power itself (Bietti, 2020; Rességuier and Rodrigues, 2020).

Such approaches are justified by claims that algorithmic technologies are too complex for regulators or the public to understand or that the pace of technological development cannot afford to be stymied by time-consuming regulatory compliance. Regulators have largely accepted these arguments, opting for regulatory requirements that only burden tech companies with self-assessments that allow them to write the standards to which they are held.

In correcting algorithmic bias through the emergent field of ‘Fair ML’ (machine learning), we argue that resistance-architected-into-design must scale to meet the computational and business logics of automated decision-making. Algorithmic bias is contextual in its effects, even if it is produced through technologies that operate at scale, i.e. independent of context. Whereas resistance to the (contextual) harms of algorithmic bias is predominantly local, this resistance eventually has to be transformed into design specifications that, while addressing harms, do not interfere with computational systems’ ability to scale (See for instance, Gebru et al., 2018, and Mitchell et al., 2019). We propose that design be understood here as its own rather distinct onto-epistemology; a powerful practice of knowledge-making that creates and shapes worlds. What matters is where and how design is shaped thus; Silicon Valley’s design-as-onto-epistemology constitutes a strange site for resistance-based activities to shape corporate power and for corporate power to recast resistance in the mold of its own logics.

This is perhaps what Twitter enacts with their ‘algorithmic bias bounty’ (Chowdhury and Williams, 2021), which proposes drawing on contextually-bound perspectives of those outside the company to reveal algorithmic biases on their platform that might not otherwise be readily
visible to the company itself, which necessarily operates fully ‘at scale’. ‘Bias bounties’ are inspired by a practice familiar to the butch culture of the infosec community, that of ‘bug bounties’ (Bar On, 2018). Bug bounties are hunts for malware, usually between elite (in terms of skills) hackers aligned against each other along 20th century geopolitical-military arrangements, or corporate antagonisms. ‘Bounty-hunting’ is a practice rich in metaphors of rough justice, frontierism, and violence. In computational terms, a bug destroys the smooth functioning of a system. In social terms a bug is infinitesimally small—almost immaterial—but inordinately powerful. Timothy Mitchell discusses mosquitos’ agency not only as a vector of disease, but also as the fulcrum on which claims experts make about their socio-technical efforts to counteract the mosquito’s agency (Mitchell, 2002). The algorithmic bias bounty, in extending practices of bug bounties to algorithmic bias, therefore suggests that bias is an anomaly, rather than a feature, of automated decision-making; and something to be identified and stamped out. But it also situates the ‘hunting’ of bias within a particular set of expert practices—those of the bounty hunter—already aligned with software and algorithm development.

A different set of dynamics emerges through the scandal at Google’s Ethical AI Lab that resulted in the termination of the Lab’s co-founders; this began over a paper that outlined how Google and its engineers constitute a site of ethical decision-making (Bender et al., 2021). This work centers environmental damage, automated bias, and “hegemonic world views” as the subject of ethical AI, and shows how current practices of building large-scale computational models contribute to such unethical business practices. The authors proposed practical recommendations that assume ethical technology as a relational, socio-technical process, emphasizing how the work is done from within the organization, rather than external laws or regulation, or tweaking optimization functions and weights in machine learning systems. They push toward advocating practical and professional approaches: assessing downstream effects of technologies; asking if these technologies are beneficial to a variety of communities around the world; consistently reflecting on the values through ‘value sensitive design’ processes; performing post-mortems and ‘pre-mortems’ of business products as they get released into the world; documenting how datasets are selected and how they inform model-building; among others. Their paper was not about the encoding of human values into computation in a traditional machine ethics style approach, i.e. how to comport the company computationally in accordance with terms set out by law or principled ethical frameworks (see Metcalf et al., 2019). Rather, it suggested attempting to change the practices of Google internally, through alterations to the practices of engineering.

But while their approach is not about ‘fixing’ specific algorithms to ‘be’ ‘less racist’, they are in fact suggesting we re-architect the system from the top down. The authors of the paper in question are computer scientists after all, so their locus of intervention can only come from the top-and Google’s Ethical AI Lab is as powerful as it gets—to “rearrange the algorithm for the good of society” (Amoore, 2020: 7) Perhaps the internal resistance faced by two of the authors, Gebru and Schmitchell, resulted from their critique of practices that constitute how computational infrastructures are built and maintained. Eventually, the system they wanted to re-architect is entirely socio-technical and relates to the industrial, organizational, and socio-cultural of Big Tech; this was understood as being at odds with Google’s core business proposition.

By contrast, activist technologists insist on bias mitigation through social and cultural change, resisting the “seduction” of architecting bias mitigation solutions from the top-down (Powles and Nissenbaum, 2018). These latter communities argue that technical fixes are limited in range and scope, and will need the constant engagement and work of multiple communities and the development of a range of literacies. But automated or top-down solutions are seductive because it is hard to manage harassment, bias, and discrimination at scale. What is difficult about this seduction is the entanglement of automation and computation with the ease of modern life itself.
Yet there are precedents of ground-up community-based research, advocacy, and policy engagement that have been taken to platform companies but have stalled, never scaled, or have been contentious; online harassment is one of them. In the next section, we make a tangential move to reflect on the fault lines in how and why this resistance-to-seduction both has and has not filtered up to scale through Big Tech.

Resistance from the ground-up

Three cases—racist language, encryption, and online harassment—each a distinct domain in the study of interactions between policy, values, and technical design, are included here because they deepen the context, situate, and historicize Big Tech’s responses to harms discussed here. The inclusion of these cases might be considered unusual here, unusual for current discussions of algorithmic antagonisms, because even how we understand what ‘algorithmic’ means is shaped by a combination of what research about algorithms already exist, who has had access to algorithmic systems in order to critically examine them, and what we think ‘algorithms’ are in the first place. But, ‘algorithm’ itself does not necessarily hold up as a clear analytic category or a singular object of productive inquiry (Barocas et al., 2013) and ‘disappears’ into ‘material history’, ‘socio-technical systems’, or ‘culture’ (Seaver, 2017), hence becomes unstable and difficult to study, let alone govern (Ziewitz, 2017: 8) More critically perhaps, accountability itself is produced, is relational, how its actors are identified and algorithms become “accountabilia” (Ziewitz, 2017; Woolgar and Neyland cited in Barocas et al., 2013). So, the following examples serve as a contrast, allowing us to identify how Big Tech’s current adoption of ‘resistance’ from the top-down is not new, does not include affected communities, and eventually aligns with its own ambitions.

The most recent example of re-designing software to mitigate a historic harm has been in efforts to remove racist terms such as ‘master’ and ‘slave’ embedded in computer engineering language; and movements that offer alternatives (Conger, 2021). This comes in the wake of global protests in support of the Black Lives Matter movement in 2020 following the death of George Floyd at the hands of local police. That it took so very long for this to be enacted despite research that has shown how racist values are deeply coded into software and technical design (Eglash, 2007; Roth, 2009) is indicative of how the powerful can be forced to act only when things are perceived to be particularly egregious to their own reputations.

A second case is Apple’s defense of the right to encryption and freedom from surveillance. Activists and human rights defenders have strongly supported the use of encryption tools to enable the conditions that protect freedom of speech, and vulnerable communities of activists and targets of online violence. This is a challenge to governments who invariably want to limit and bypass encryption ostensibly in the interest of national security, but that also has the tradeoff of mass surveillance (APC, 2015). A fairly niche and complex technical domain, encryption achieved public visibility when the US government wanted Apple to build a custom operating system that would disable the security features on a terrorist’s iPhone to give them (the US government) access to the information on it. In effect, this was a backdoor that would set a precedent to undermining the security of Apple’s products, as well as violating constitutional guarantees to freedom of speech and expression (Electronic Privacy Information Center, 2014).

Apple eventually did not capitulate to the government’s demands, and the government dropped the case. At the time of writing this article, Apple has taken a stand against the Israeli surveillance technology company, NSO Group, suing them for installing malware on a small number of Apple devices belonging to human rights defenders and journalists, and banning them from developing
any software that would work on Apple devices (Apple, 2021). Apple has also made generous
donations to the Citizen Lab, a research group at the University of Toronto, and Amnesty
International, that have worked over years to expose these attacks and secure activists. Security
and privacy are central to Apple’s brand, so it follows that the company has prioritized security
as a design feature and as policy, even in direct resistance to the law. This sets a precedence for
how anonymity and privacy online are valued. Also germane to our discussion is how what is con-
sidered a niche standard is in fact quietly encoded into most of our digital messaging apps, from the
widely-used WhatsApp, to the more niche app, Signal. The difference is in how companies (like
WhatsApp/Meta) do capitulate to law enforcement⁴ whereas others do not.

A third case relates to gendered online harassment. Feminist technologists, activists, journalists,
and researchers have shown that online harassment is enabled and sustained by the interaction of
social media platform interface design and its affordances, and the situated, embodied, social
and cultural contexts of social media and technology use (Kovacs and Ranganathan, 2017; Sim
and Zevenbergen, 2017⁷, PEN America, 2021). But, digital security tends to be framed in terms
of securing digital devices, rather than in terms of social, intimate, and interpersonal worlds in in-
teraction with technological systems (Interview with Mallory Knodel, 2019). So, online harassment
becomes transformed into a series of discrete computational events that can be identified and con-
trolled; however, violence is insidious in its manifestations and mutations and thus difficult to keep
up with. So, even when technical fixes are designed to mitigate harms, they fall short because the
socio-technical aspects of how violence happens are not fully addressed by re-design alone.

For example, a design team at Twitter enabled an adjustment that silenced notifications sent to
trolling victims when they were added to malicious Twitter lists. But, removing list notifications
allows harassers to compile and share lists of targets undetected. Victims are unaware they are
shared targets and can’t fight back. The violation is so obvious that the feature is reversed within
hours (Perez, 2017).⁵ Similarly, Google’s Perspective AI was an attempt to use algorithmic tech-
nologies to identify hateful and misogynist speech online, but was largely ineffective because it
failed to engage with the socio-technical reality of online harassment, as well as how language
itself can code meanings that are intended only for specific audiences, and thus evade algorithmic
recognition (Weimann and Ben Am, 2020). Facebook’s experiments with ‘hashing’ personal
images to prevent their unauthorized sharing (aka “revenge porn”) was also criticized by security
experts because it involved uploading nude photographs to the website (Schneier, 2017).

It takes resources to investigate and compile socio-technical harms; to amplify them to mount
resistance and build coalitions around this resistance; and then to take that resistance to the com-
panies that architect those systems. Perhaps this is why issues of encryption and freedom of
speech and expression that are playing out between transnational corporations and nation states,
and taken up by Amnesty Tech and Citizen Lab, can be seriously addressed by Apple. The
attacks on women journalists and activists—also a freedom of speech issue—appear more like bother-
some, complicated, social concerns. Silicon Valley’s architects of digital platforms do not consider
gendered online harassment either easy or profitable to address, possibly because to actually address
harassment is to address its social and structural contexts, to recognize user knowledge and feminist
knowledge, and to acknowledge the value of an expert community that mediates design and policy
between users and tech companies. No efforts to minimize harassment have effectively scaled
through Big Tech interventions.⁶

Eventually, however, most marginalized users are not waiting for Big Tech to deliver the solu-
tions—perhaps because they cannot. They may not have the reach, resources, or access to persuade
Big Tech to solve the problems it has created for them; and communities may not have the time to
wait while they are actively experiencing harms. They are organizing against the use of facial rec-
ognition by their landlords (Moran, 2020), they are interrogating the systems’ claims themselves
(Buolamwini and Gebru, 2018, Costanza-Chock, 2018), they are engaging in direct challenges to the logics the systems embody through “social audits” (Vecchione et al., 2021), and finding ways to secure themselves online by becoming more digitally savvy.\(^7\) These do not necessarily require access to the full software engineering stack behind the front ends of these systems. The mitigation and prevention of abuse relies on practices of diffuse mutual support, adoption of alternate digital infrastructures,\(^8\) or more organized resources like helplines, or DIY security tips for vulnerable communities provided by international civil society organizations.\(^9\) Many of these organizations and projects serve journalists and activists in particular, people who are actually engaged in the work of resistance to power.

**Varieties of refusals**

In recounting how the problem of Big Tech-enabled online harassment has been (un)addressed by Big Tech, we argue that there is a dilution of the terms on which digital participation is understood. Because these terms are bound more closely to industry and its allies, it becomes the locus of our agitation, notions of speech, utopia, freedom, our constructions of self and identity, repositories of history (or manipulations thereof). In some quarters, there has been an important shift from the language of resistance to that of refusal. (We do not suggest that refusal as a politics is new). And the shift towards ‘refusal’ is not an unequivocal good or positive trend necessarily: refusal to vaccines, the shape of the earth, of situated knowledge, and to modern knowledge-making practices, are rife and particularly across the digital. Holding on to this contradiction, we turn to a discussion of refusals that are unsatisfied with the temporalities, terms, and terrain on which Big Tech operates.

We point to a taxon of refusal in which Luddism, feminism,\(^10\) and decoloniality are closely entwined. This exists alongside what has become a cottage industry of pure critique that dismisses engagement to reform, contextualize, or otherwise situate algorithmic technologies by bounding and dismantling their differential impacts, or by attempting to bring them to heel under political and social forms of accountability. Keeping aside these recent ‘bad faith’ approaches to dismiss all kinds of resistance, we understand entwined and intersectional approaches to New Luddism, feminism and decoloniality as refusals that utter distinct kinds of ‘no’ even in their resonant convergences.

It must be possible to refuse being implicated in big technological data systems, as the authors of the Feminist Data Manifest-No suggest, without also refusing a place in society.\(^11\) The spirit of a new Luddism is entwined with this; rather than opposing technology for opposition’s sake, New Luddites suggest that to smash the apparatuses that make life less livable is to prioritize human dignity and flourishing over that of capital (Mueller, 2021). In reflecting on the limits of resistance to the harms of algorithmic systems when it is articulated through the practices of Silicon Valley, New Luddites ultimately seek a position that lies not only outside of Silicon Valley, but also prior to it. By this we mean that there are some aspects of life—social life, biological life, interior life—that ought not be subject to the logics of digital, surveillant capitalism. There is no ‘safer’ way to be online in a world that algorithmically re-embeds racism and misogyny, or that facilitates harassment. Products that do so are “unsafe at any speed” (Nader, 1972). They present “designed-in dangers” that can be (and should be) minimized and mitigated, but never eliminated. And because these dangers cannot be eliminated, the products that produce them must not be made so central to our lives. So, practices of keeping personal and private aspects of life offline, such as not sharing images of young children and babies online, using encrypted services like Signal, or through the adoption of anonymous social media accounts to speak to a smaller circle of confidants (e.g. “Small Twitter”) are small acts of refusal in this vein.
While it is true that staying off Big Tech platforms and services is not an option for most people, efforts to continually develop and maintain access to information and communications technologies outside of monopolistic control are essential. Free, open source, decentralized, autonomous, decolonial, techno-shamanistic, or ‘gambiarra’ ("make-shift") approaches and communities are doing just this, and chiefly by rejecting notions of scale, and connection, integrated into the hegemonic capitalism of Big Tech (Roussel and Stolfi, 2020). There are small and resolute communities of autonomous feminist tech infrastructure developers and technologists who are trying to exist outside these systems by literally architecting their own hardware, data storage, and servers, as well as the values that come with small infrastructures, such as slowness, care, and repair (Savic and Wuschitz, 2018; Smith and Kefir, 2017; Numun Fund). In parallel to these refuseniks are artists, practitioners and scholars who are developing new visions for technology infused with social values from intersectional and marginal perspectives, such as: the New New about the ethics of algorithms; ‘data feminism’ (D’Ignazio and Klein, 2020); Whose Knowledge?; feminist data sets; feminist principles of the internet (which speak directly to matters of encryption and anonymity in the context of gendered harassment); feminist open source investigations (Dyer and Ivens, 2020).

These projects and practices do not investigate or expose specific algorithms, but are positions of refusal to the socio-technical power, policies, and politics that sustain algorithmic systems. These positionalities inform methodological approaches and practices to working with technology. Its practitioners bring attention to how endemic social conditions of poverty, misogyny, racism, climate injustices, and disability are exacerbated by poorly constructed algorithms, and entwined with the values of scale and speed built into these technologies. And that because of this, practices of working with data should recognize the unique perspectives of people who are marginal to technology development. But, these initiatives are not necessarily aligned with refusal-as-disconnection, or in terms of being outside of Big Tech infrastructures; some may suggest strategic visibility of marginal perspectives through and within Big Tech. (It is perhaps a point for future discussion elsewhere that Big Tech remains the hegemonic center, the measure of the distance from which is also a measure of the validity or purity of politics.) Still, refusal initiatives are an opportunity to empower marginal perspectives as valid sites of knowledge and knowledge-making about what constitutes harm, but also new baselines for co-existing with technology.

Similarly ‘decoloniality’ could be understood as a politics of existing in more than one place at a time, metaphorically speaking, to refuse notions of identity, place, and borders, while still existing therein. Decoloniality can span questions of indigenous (re)ownership of land in North America, Australia, New Zealand; to the rejection of the English language as the global lingua franca. Between these lie a spectrum of demands and solidarities that highlight a desire to shift dominant onto-epistemic frames and dynamics, questioning what constitutes the basis for knowing and knowledge. Projects and locations of the kinds discussed above might be considered ‘decolonial’ in this sense. Frank Pasquale identifies ‘waves’ of algorithmic accountability, with a first wave oriented towards improving algorithmic systems, as resistance approaches outlined here do; and a second wave that asks if these algorithmic systems are required at all (Pasquale, 2019). In that vein, Decoloniality questions the centrality of particular kinds of concerns and harms as defined by hegemonic centers and institutions.

But, oddly, decolonial critiques are being adopted within the very technology companies that have hegemonic and imperial ambitions. Recent work in the tech industry has examined a range of technology critiques grounded in decolonial theory and claimed this form of critique as promising cues for more value-sensitive development practices. For instance, “decolonial AI” argues that forms of resistance to tech dominance can be used to improve technology products, and without
ever grappling with the power imbalances between technology companies and the various publics implicated in its products. For instance, some technologists argue that artificial intelligence “can be used as a decolonizing tool” (Mohamed et al., 2020: 16) by incorporating decolonial theory as a set of design methods for technologists to use, a nonsensical proposition that elides any meaningful consideration of power held by those very technologists.

Conclusions

This essay has discussed two broad varieties of resistance and refusal practices that have emerged in response to contemporary kinds of algorithmic harms: practices of resistance emerging from Big Tech, and refusals that emerge and remain outside of it. “Inside’ and ‘outside’ Big Tech might not always hold up as water-tight categories however; as we discuss, proposals from within Big Tech, such as the work of Google’s Ethical AI lab under Dr Timnit Gebru, and Twitter’s efforts to bring greater transparency to its platform are sincere efforts to limit harms. Yet, our interest is in the discursive implications of different kinds of responses to algorithmic harms. Influenced by Foucauldian and Baradian theories of how material practices generate particular kinds of knowledge and knowledge-making, this article is concerned with how practices of measuring, identifying, framing, and naming harms establish the epistemic and philosophical terms of what freedom from harm mean. Hence, how harms are being addressed by different actors is important to our categorization.

‘Resistance’ to algorithmic harm, identified by scholars, activists, and communities experiencing that harm, has quickly become adopted by different industrial actors as a political stance, a design method, an industrial practice to make tech better. We identify ‘resistance’ as a zeitgeist and key to Big Tech businesses; it is crucial for companies to be seen as responsive to outside demands for more ethical products and practices. Algorithmic bias, faulty machine learning, and poorly made automated decisions are just not good business in the context of acute, global, sensitivities to power. Big Tech’s adoption of resistance to algorithmic harms ends up setting the terms that overdetermine the scope of Big Tech-oriented resistance movements. Fair ML, ethical AI, and bias concerns have dominated the discourse, but eventually only serve to denature, without resolving, many critiques of Big Tech’s practices. While there are resistance movements that are orthogonal to these narrow channels (Stark et al., 2021) these dominant axes actually re-inscribe the concerns of private companies that develop algorithmic systems. We thus identify on one hand, that ethics, fairness, and bias concerns frame, or re-frame, the vulnerabilities algorithm developers themselves perceive to be legal transgressions and reputational harm, and mobilize external actors to work towards solutions that address these vulnerabilities.

At the same time, targeted resistance to and critique of these systems is being levied from outside these frames, by those directly affected by algorithms or primarily concerned with their effects (see Garcia et al., 2020; Nkonde, 2020; Roussi, 2020).

While the scope of resistance to Big Tech has expanded over time, as more controversies about tech development are uncovered, new solidarities emerge, and new critical frames are deployed, tech development practices have proven their agility in incorporating these forms of resistance into the development process. Many important efforts are being undertaken to ‘design out’ the dangers Silicon Valley has built. But so long as those efforts only seek to make its products ‘safe enough’ to quell resistance, they will have failed to put the claims at the center of that resistance above the avaricious tendencies of development.

We also tabled three cross-discipline cases related to racist language, encryption, and online harassment that offer insights into the experience and management of harms by Big Tech. Contrasting
responses to online harassment and the defense of the right to encryption and anonymity—both related to freedom of speech and expression online—we argued that conditions perceived endemic to how society is, such as gendered inequality, exacerbated by digital technologies, are hard problems that Big Tech, for all its solution-ism, finds intractable (Sloane, 2019). So these problems remain exacerbated and unresolved. The mixed fortunes of mitigating gendered online harassment through partnerships and negotiations with Big Tech suggest that constant appraisals of what kinds of accountability we are pursuing, for who, and towards what ends, could lead us to attend to urgent situations of inequality and discrimination in society.

Refusal practices situated in local movements, lived and situated histories of marginalization, decline the limits of categories and groupings adopted by Big Tech in their resistance efforts, and recognize the difficult, messy whole of the social. New Luddite, feminist, and decolonial approaches are refusals of “a particular series of declinations” that translate the problems of bodies into problems for machine structures and architectures. These approaches to working with data and the digital turn away from the But, “the beautiful structure, the new machine, refuses to acknowledge or accept where it came from, and gives itself as the only possible answer, the solution.” (Bassett, Kember, O’Riordan, 2019: 28) In presenting itself as a problem-solving device, Big Tech’s beautiful machines carve out the shape of what socio-technical problems are in the first place, and how they are to be addressed. In this way, Big Tech positions itself outside the endemic social biases it aggravates, as in the case of Kronos.

But if we acknowledge that our contemporary societies are highly data-ied, and substantially algorithmically shaped and governed, then, the Refuse-rs argue, there has to also be a concerted movement in the other way: of the social not just as sites of problem-solving, but more explicitly as a place where directions to answers and solutions to harms actually exist.

Interviews

Interview with Mallory Knodel. January 25, 2019 in the context of a study about feminist approaches to digital security. Interview by first-named author.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Maya Indira Ganesh https://orcid.org/0000-0002-0966-5880

Notes

1. The repository can be accessed online here: https://docs.google.com/spreadsheets/u/0/d/1Bn55B4xz21-_Rgdr8BBb2lt0n_4rzLGxFADMIVW0PYI/htmlview#.
2. This conference has undergone several name-changes in its short history. It started as a small academic workshop called Fairness, Accountability and Transparency in Machine Learning (FATML), quickly becoming the ‘FAT* Conference’ (pronounced ‘Fat Star’); to its most recent avatar, FAccT that has been in use since January 2020. See: https://facctconference.org.
3. See https://ethicals.org/.
4. https://faq.whatsapp.com/general/security-and-privacy/end-to-end-encryption/?lang=en.
5. In fairness to Twitter, the company has worked consistently to address online harassment; this is one example from many design and UX fixes it has rolled out over the past years that are motivated by safety concerns.
6. In the past few years, most social media companies have instituted ‘trust and safety’ roles to liaise with civil society groups who have been documenting social media platform harms. We would argue that the results of these engagements are mixed because the contexts and realities of harm often require substantial policy and technology change if they are to be mitigated, and often in ways that might clash with business aims.
7. We refer to the work of organisations such as Chayn, Heart Mob, Trollbusters, Access Now’s Digital Security Helpline, the Gender-Tech network and GenderSec Curriculum from Tactical Tech, Glitch UK, Coding Rights, APC’s Women’s Rights Program, among others, who are addressing gendered online harassment.
8. The increased popularity of free, open source, and fully encrypted messaging platforms like Signal are a good example.
9. For instance, Access Now, the Digital Rights Foundation, or the Women and Media Center’s Speech project.
10. We acknowledge that there are many kinds of feminisms, however, we adopt the singular ‘feminism’ because it is beyond the scope of this paper to go into these differences. In this essay we refer chiefly to feminist movements and practices that are majority world, and influenced by perspectives on digital rights, and loosely open source, and cipherpunk values. Feminism is a strong motivating politics that centres the experiences of power/disempowerment, misogyny, discrimination, and finds solidarity with multiple other struggles for justice for marginalised communities.

References

Amoore L (2020) Cloud ethics: Algorithms and the attributes of ourselves and others. Durham: Duke University.
Angwin J, Larson J, Mattu S, et al. (2016) Machine bias. ProPublica, May 23, 2016. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed on December 15, 2021).
Apple (2021) Apple sues NSO Group to curb the abuse of state-sponsored spyware. Press Release, November 23, 2021. Available at: https://www.apple.com/newsroom/2021/11/apple-sues-nso-group-to-curb-the-abuse-of-state-sponsored-spyware/ (Last accessed November 25, 2021).
Association for Progressive Communications (2015) The right to freedom of expression and the use of encryption and anonymity in digital communications Submission to the United Nations Special Rapporteur on the Right to Freedom of Opinion and Expression by the Association for Progressive Communication (APC) Association for Progressive Communications (APC) February 2015. Available at: https://www.ohchr.org/Documents/Issues/Opinion/Communications/AssociationForProgressiveCommunication.pdf (Last accessed November 23, 2021).
Barad K (2007) Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning. Durham: Duke University.
Barad K (2011) Nature’s Queer Performativity. Qui Parle 19(2): 121–158. DOI: https://doi.org/10.5250/quiparle.19.2.0121.
Barocas S, Hood S and Ziewitz M (2013) Governing algorithms: A provocation piece. Available at: https://ssrn.com/abstract=2245322.
Bar On AE (2018) We need bug bounties for bad algorithms. Vice. May 3, 2018. Available at: https://www.vice.com/en/article/8xkyj3/we-need-bug-bounties-for-bad-algorithms.
Bassett C, Kember S and O’Riordan K (2019) *Furious: Technological Feminism and Digital Futures*. London: Pluto Press.

Bender EM, Gebru T, McMillan-Major A, et al. (2021) On the dangers of stochastic parrots: Can language models be too big? In: *proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, Online: 3–10 March 2021. DOI: 10.1145/3442188.3445922.

Bietti E (2020) From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy. In: *proceedings of the 2020 conference on fairness, accountability, and transparency*, pp.210–219. New York, NY, USA: Association for Computing Machinery (FAT* ‘20). DOI: 10.1145/3351095.3372860.

Buolamwini J and Gebru T (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of Machine Learning Research*. Vol 18, pp. 1-15. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Chowdhury R and Williams J (2021) Introducing Twitter’s first algorithmic bias bounty challenge. Twitter Blog. Available at: https://blog.twitter.com/engineering/en_us/topics/insights/2021/algorithmic-bias-bounty-challenge (Last accessed August 20, 2021).

Conger K (2021) ‘Master’, ‘slave’ and the fight over offensive terms in computing. The New York Times. April 13, 2021. Available at: https://www.nytimes.com/2021/04/13/technology/racist-computer-engineering-terms-ietf.html.

Costanza-Chock S (2018) Design justice, AI, and escape from the matrix of domination. *Journal of Design and Science*. https://doi.org/10.21428/96c8d426

D’Ignazio C and Klein F (2020) *Data Feminism*. MIT Press. https://data-feminism.mitpress.mit.edu/.

Dwork C, Hardt M, Pitassi T, et al. (2012) Fairness through awareness. In: *proceedings of the 3rd innovations in theoretical computer science conference on - ITCS ’12*, pp.214–226. Cambridge, Massachusetts: ACM Press. DOI: 10.1145/2090236.2090255.

Dyer S and Ivens G (2020) What would a feminist open source investigation look like? *Digi War* 1: 5–17.

Eglash R (2007) Broken metaphor: The master-slave analogy in technical literature. *Technology and Culture* 48: 360–369.

Electronic Privacy Information Center (2014) Apple v FBI. Available at: https://epic.org/documents/apple-v-fbi-2/.

Foucault M (1980) *Power/Knowledge: Selected Interviews and Other Writings, 1972-1977*. G Colin (Ed). New York: Harvester.

Gandy OH (2009) Engaging rational discrimination: Exploring reasons for placing regulatory constraints on decision support systems. *Ethics And Information Technology* 12: 29–42

Garcia P, Sutherland T, Cifor M, et al. (2020) No: Critical refusal as feminist data practice. In: *conference companion publication of the 2020 on computer supported cooperative work and social computing*, pp.199–202. Virtual Event USA: ACM. DOI: 10.1145/3406865.3419014.

Gebru T, Morgenstern J, Vecchione B, et al. (2018) Datasheets for datasets. In: *proceedings of the 5th workshop on fairness, accountability, and transparency in machine learning*, Presented at the PMLR, Stockholm, SE 10–15 July 2018. CoRR abs/1803.09010.

Greene DM, Hoffmann AL and Stark L (2019) Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. In: *proceedings of the 52nd Hawaii international conference on system sciences 2019*.

Kantor J (2014) Times article changes a Starbucks policy, fast Times Insider. The New York Times. Available at: https://www.nytimes.com/times-insider/2014/08/22/times-article-changes-a-policy-fast/ (accessed 19 August 2021).

Knodel Mallory (2019) *Online interview*.

Kovacs A and Ranganathan N (2017) Gendering surveillance. Available at: https://genderingsurveilance.interndemocracy.in/ (accessed 20 August 2021).

Metcalf J, Moss E and Boyd D (2019) Owning ethics: corporate logics, silicon valley, and the institutionalization of ethics. *Social Research* 86: 449–476.

Mitchell M, Wu S, Zaldívar A, et al. (2019) Model cards for model reporting. In: *proceedings of the conference on fairness, accountability, and transparency, FAT* ‘19*, pp.220–229. Atlanta, GA, USA: ACM Press. DOI: 10.1145/3287560.3287596.
Mitchell T (2002) Can the mosquito speak? In: Mitchell T (ed) Rule of Experts: Egypt, Techno-Politics, Modernity. Berkeley, CA: University of California, pp.22–45.

Mohamed S, Png M-T and Isaac W (2020) Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. ArXiv:2007.04068 [Cs, Stat], July. DOI: 10.1007/s13347-020-00405-8.

Moran T (2020) Atlantic Plaza Towers Tenants won a halt to facial recognition in their building: Now they’re.... AI Now Institute (blog). January 9, 2020. Available at: https://medium.com/@AINowInstitute/atlantic-plaza-towers-tenants-won-a-halt-to-facial-recognition-in-their-building-now-theysre-274289ad6d8eb.

Moss E and Metcalf J (2020) Ethics owners: A new model of organisational responsibility in data-driven technology companies. Report by Data & Society Research Institute, New York. Available at: https://datasociety.net/pubs/Ethics-Owners.pdf (accessed 20 August 2021).

Mueller G (2021) Breaking things at work: The Luddites were right about why you hate your job. London and New York: Verso Books.

Nader R (1972) Unsafe at any speed: the Designed-in Danger of the American Automobile. New York: Bantan Books.

Nkonde M (2020) Automated anti-blackness: Facial recognition in Brooklyn, New York. Harvard Kennedy School Journal of African American Policy: Anti-Blackness in Policy Making: Learning from the Past to Create a Better Future, 2019-2020 Volume. Available at: https://hjaap.hkspublications.org/wp-content/uploads/sites/14/2020/05/Final-PDF-for-Website.pdf. Last accessed January 27, 2022. pp. 30–36.

Numun Fund. Available at: https://numun.fund/.

Pasquale F (2019) The second wave of algorithmic accountability. The LPE project blog. Published November 25, 2019. Available at: https://lpeproject.org/blog/the-second-wave-of-algorithmic-accountability/.

PEN America (2021) No excuse for abuse: What social media companies can do now to combat online harassment and empower users. Report by PEN America 31 March 2021. Available at: https://pen.org/report/no-excuse-for-abuse/ (Last accessed August 21, 2021).

Perez S (2017) Twitter quickly kills a poorly thought out anti-abuse measure. Tech Crunch, February 14, 2017. Available at: https://techcrunch.com/2017/02/14/twitter-quickly-kills-a-poorly-thought-out-anti-abuse-measure/ (Last accessed November 29, 2021).

Powles J and Nissenbaum H (2018) The seductive diversion of ‘solving’ bias in artificial intelligence. Medium. DOI: https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53.

Rességuier A and Rodrigues R (2020) AI Ethics should not remain toothless! A call to bring back the teeth of ethics. Big Data & Society 7(2). DOI: 10.1177/2053951720942541.

Roth L (2009) Looking at Shirley, the ultimate norm: Colour balance, image technologies, and cognitive equity. Canadian Journal of Communication 34: 111–136.

Roussel N and Stolfi A (2020) Taking back the future: A short history of singular technologies in Brazil. Catalyst: Feminism, Theory, Technoscience 6(2), no pagination. Special Section: Computing in/from the South. DOI: 10.5399/uo/ada.2018.13.10.

Roussi A (2020) Resisting the rise of facial recognition. Nature 587: 350–353.

Savic S and Wuschitz S (2018) Feminist hackerspace as a place of infrastructure production. Ada: A Journal of Gender, New Media, and Technology (Issue 13). DOI: 10.5399/uo/ada.2018.13.10.

Schneier B (2017) Available at: https://www.schneier.com/blog/archives/2017/11/facebook_finger.html.

Seaver N (2017) Algorithms as culture: some tactics for the ethnography of algorithmic systems. Big Data and Society 4(2). DOI: 10.1177/20539517177738104.

Sim K and Zevenbergen B (2017) Design ethics for gender-based violence and safety technologies. Freedom to Tinker blog. Available at: https://freedom-to-tinker.com/2017/07/24/design-ethics-for-gender-based-violence-and-safety-technologies/ (Last accessed August 21, 2021).

Sloane M (2019) Inequality Is the Name of the Game: Thoughts on the Emerging Field of Technology, Ethics and Social Justice. In: Proceedings of the Weizenbaum Conference 2019 "Challenges of Digital Inequality - Digital Education, Digital Work, Digital Life." Berlin pp. 1–9. https://nam12.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdoi.org%2F10.34669%2Fwi.cp%2F2.9&
Smith E and Kefir (2017) An ongoing conversation on feminist autonomous infrastructure. Genderit, August 2, 2017. Available at: https://genderit.org/articles/ongoing-conversation-feminist-autonomous-infrastructure-erika-smith-and-k%2Feir. (Last accessed December 15, 2021)

Stark L, Greene D and Hoffmann AL (2021) Critical perspectives on governance mechanisms for AI/ML systems. In: Roberge J and Castelle M (Eds) The Cultural Life of Machine Learning, pp.257–280. Cham: Springer International. DOI: 10.1007/978-3-030-56286-1_9.

Vecchione B, Barocas S and Levy K (2021) Algorithmic auditing and social justice: Lessons from the history of audit studies. ArXiv:2109.06974 [Cs], September. DOI: 10.1145/3465416.3483294.

Weimann G and Ben Am A (2020) Digital dog whistles: The new online language of extremism. International Journal of Security Studies 2: 1–24.

Ziewitz M (2017) A not quite random walk: Experimenting with the ethnomethods of the algorithm. Big Data & Society 4(2): 2053951717738105.