No Simple Fix

How AI harms reflect power and jurisdiction in the workplace

Nataliya, Nedzhvetskaya*
University of California, Berkeley
nataliyan@berkeley.edu

JS, Tan
Massachusetts Institute of Technology
js_tan@mit.edu

ABSTRACT

The introduction of AI into working processes has resulted in workers increasingly being subject to AI-related harms. By analyzing incidents of worker-related AI harms between 2008 and 2023 in the AI Incident Database, we find that harms get addressed under considerably restricted scenarios. Results from a Qualitative Comparative Analysis (QCA) show that workers with more power resources, either in the form of expertise or labor market power, have a greater likelihood of seeing harms fixed, all else equal. By contrast, workers lacking expertise or labor market power, have lower success rates and must resort to legal or regulatory mechanisms to get fixes through. These findings suggest that the workplace is another arena in which AI has the potential to reproduce existing inequalities among workers and that stronger legal frameworks and regulations can empower more vulnerable worker populations.

CCS CONCEPTS

• **Social and professional topics**; • **Socio-technical systems**;

KEYWORDS

artificial intelligence, harms, work, expertise, algorithmic management, governance, regulation, safety

ACM Reference Format:
Nataliya, Nedzhvetskaya and JS, Tan. 2024. No Simple Fix: How AI harms reflect power and jurisdiction in the workplace. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3630106.3658915

1 INTRODUCTION

The introduction of AI can produce numerous changes in the workplace– challenging human expertise, expanding the predictive potential of large datasets, and replacing organizational goals with algorithmic incentives [9, 15, 18, 24]. Not all changes are beneficial and workers themselves are frequently subject to harm. While studies have documented the harmful effects of algorithmic mismanagement, labor exploitation in the AI supply chain, and social bias in machine learning models, less work has been done to examine the ways that workers have responded to such harms [8, 17, 29, 37, 40, 41, 57]. To our knowledge, no work has yet systematically examined what factors increase the likelihood of AI workplace harms being fixed, either by offering a technical solution to fix the origin of the harm, discontinuing the use of the technology itself, and/or offering compensation to workers for the harms incurred.

An extensive literature within the computer sciences has sought to provide preventative measures and technical fixes to AI harms through the use of audits and assessments [13, 38, 47]. This research, while valuable, is predicated on the expectation that creators of AI systems want to minimize harms and agree with users and community members on what those harms look like. Our study begins first and foremost with the assertion that AI is a sociotechnical system [33]. In addition to focusing on the micro-level processes and interventions that lead to AI harms, we should also seek to understand how these harms are received at the level of the workplace, supplementing our technical understanding with social-scientific explanations [7, 41].

This study focuses on one population that has been impacted by AI's adoption and development: workers. Using data from the AI Incidents Database (AIID), a publicly accessible global dataset of AI harms, we evaluate all reports of workers being harmed by AI and determine which configurations of power have been most successful in providing fixes for AI harms. Power resource theory, which maintains that workers can utilize different forms of collective power to achieve their demands, provides a theoretical foundation for understanding the ways that workers can assert power over the governance of AI in the workplace [32, 50, 51, 58]. We bridge an existing sociological literature on power relations in the workplace with discussions of safety and fairness in the AI literature.

We use the term "artificial intelligence" to mean technologies employing machine learning, which includes but is not limited to software employing algorithms and hardware such as robots or self-driving cars. At times throughout the paper, we may use more specific terms such as "machine learning" or "algorithms" when describing specific cases. "AI workplaces," another term we employ throughout, simply means any physical or virtual workplace that incorporates the use of AI technologies in a non-incidental way.

Through our analysis, we find that AI harms get addressed under considerably restricted scenarios. Workers with more power resources, either in the form of expertise or labor market power, have a greater likelihood of seeing harms fixed, all else equal. By contrast, workers lacking recognized expertise or labor market power have much lower success in getting harms fixed and must resort to legal or regulatory mechanisms, to get fixes through. These findings suggest that the workplace is another arena in which AI has the potential to reproduce existing inequalities among workers

*Place the footnote text for the author (if applicable) here.

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-0450-5/24/06
https://doi.org/10.1145/3630106.3658915

and that stronger legal frameworks and regulations can empower more vulnerable worker populations.

We begin this paper with an overview of the existing literature on AI harms in the workplace. We unpack the ways that harms in the field are inherently power-laden and then introduce power resource theory as a means to understand AI systems from a sociological perspective. We introduce a typology of AI harms that includes algorithmic labor, algorithmic management, algorithmic recommendation, and data misuse. Following this, we introduce our dataset, methods (qualitative comparative analysis, or QCA) and variables. We explain our findings, presenting several cases from our database for illustrative purposes. Finally, we end with a discussion of our findings, the limitations of our analysis, and avenues for further research.

2 AI HARMS IN THE WORKPLACE

Harms are by no means unique to AI workplaces but the presence of algorithms can make it more difficult to identify both the source of a harm and its potential fix [21]. Harms can occur in situations of negligence– where the outcome was clearly an accident—or in situations of normative uncertainty– where the boundaries of what is considered safe and fair are contested [3, 39]. The challenge often lies in determining which it is and finding a consensus path toward fixing the harm. As literature in both sociology and AI has demonstrated, whether or not a harm is categorized as an accident is ultimately a question of power and has a significant effect on the likelihood of a harm being addressed [4, 44]. Viewing the AI workplace through the lens of power resource theory allows us to determine who holds power, who defines harms, and the ways in which power may be leveraged towards finding recourse for harmed workers.

Power resource theory has been applied to the study of organized labor in order to determine which strategies are most successful in forcing employers to meet worker demands [32, 51, 58]. The premise of power resource theory is that a group of workers in a structurally disadvantaged relationship with their employer can utilize different forms of collective power– structural, associational, institutional, and societal– in order to force their employer to give in to their demands [50]. Political scientists have expanded on this theory to argue that countries with strong labor movements, such as Sweden, Germany, or Norway, have strong welfare states while countries with weak labor movements tend to have weak ones [35, 52]. While power resource theory scholars have mostly examined organized labor, workers taking any sort of group action— whether a lawsuit, media campaign, internal or external protest— can be thought of as utilizing their power collectively. Power resource theory rests on a Weberian perspective of power— that success rests upon convincing another party to carry out your will despite their opposition— and therefore is especially relevant in the study of harms in the workplace [55]. To construct our argument, we first identify the ways in which existing literature suggests AI workplaces may structurally differ from non-AI workplaces before examining potential protest strategies and relevant forms of power.

Existing literature on AI and work suggests three potential challenges to identifying and fixing AI harms in the workplace. First, because AI, and in particular, machine learning algorithms are "black box" technologies, workers may struggle to explain the source of harm and suggest a specific solution [43]. Burrell (see [14]) describes AI systems as possessing three forms of opacity: corporate secrecy, technical illiteracy, and opacity of scale. All three are potentially problematic to the resolution of AI workplace harms. In their study of worker protests of AI, Nedzhvetskaya and Tan (see [40]) show that certain types of workers have a privileged ability to "unpack" the black box. Highly trained, technical workers are able to claim proximate knowledge, or an in-depth understanding, of the ways that these technologies operate and can thus incorporate that in their response to harms and speak to whether or not harms are justifiable (see also [57]).

More precarious workers, by contrast, are rarely able to make such claims and are more likely to be involved in actions where they themselves are the subject of harm (cf. [45]). The opacity of algorithms coupled with a workers' technical background, or lack thereof, is a factor in their ability to make claims to certain forms of harm and, one might expect, their ability to advocate for fixing these harms. The algorithmic recourse and accountability literature argues that individuals should have the ability to reverse unfavorable decisions from classification models through changes to the input variables [53, 54]. In other words, algorithms should be kept accountable to the populations they serve and allow for agentic human intervention. Precarious workers especially lack the power to hold algorithms accountable if they lack subject area expertise or access to the social, political, and economic resources that are necessary to demand recourse.

This form of power has traditionally been conceived of as professional jurisdiction. Professional jurisdiction is defined as an exclusive set of rights held by a profession that gives its members control over a set of practices, the boundaries of its membership, and its self-discipline, among other aspects [1, 6]. Eyal (see [25]) disaggregates experts (i.e., professionals) from expertise entirely. To explain the precipitous rise of autism diagnosis in the late 20th century, he argues the deinstitutionalization of mental retardation allowed a new set of actors—parents of children with autism working in alliance with psychologists and therapists—to develop a non-professionalized type of expertise that relied not on formal professional institutions but on networks that link agents, devices, concepts, and institutional and spatial arrangements together. Like professional associations, unions, and other forms of organized labor, expertise relies heavily on associational power, the ability to act collectively with fellow members to impose standards or barriers to the profession [27]. Expertise has been shown to play a role in driving the success of occupational activists, individuals who use their profession—and its associated expertise—to make moral or political stands [16]. The ability to assert expertise over the area in which a harm occurred allows workers some recourse over their employers and may allow them greater agency and power to act.

The second reason why AI workplaces might be challenging places to identify and fix harms is their use of algorithmic management. An increasing number of workplaces do not just implement AI systems– they use AI systems to manage workers, making decisions about how they spend their time and resources [29, 36]. This expands the space for AI-related harms into the realm of workplace control and labor relations. Workers must also contend with an

algorithm controlling the labor process and/or wages thus causing them direct harm. In their study of Uber drivers, for example, Rosenblat and Stark (see [49]) describe how Uber drivers see themselves as harmed by the platform's opaque pricing algorithm, an intentional feature implemented by the company for more effective management. Lei (see [37]) answers an even more critical question about algorithmic management– can algorithmic management itself cause more conflict and harm than traditional management? Through a comparative study of food-delivery workers in China, one set employed on a traditional service platform and another set employed on a gig platform, she finds that algorithmic management itself leads to a greater perception of exploitation and harm among workers, even when working conditions are otherwise similar.

However, other studies suggest that we should be wary of viewing all algorithmic management as exploitative and inferior to traditional forms of management. In their survey of food delivery workers, Griesbach, et al. (see [29]) find that not all algorithmic management systems are created equal. Their survey finds significantly worse job satisfaction for Instacart workers compared to other food delivery platforms due to their stringent and despotic management style. Moreover, in the Global South where informal employment relations are dominant, platforms and algorithmic management have been characterized as creating a more formal employment relationship than what would otherwise be present [11, 30]. To date, most studies of algorithmic management have involved gig workers and this is no coincidence. Gig work platforms were among the earliest workplaces to introduce algorithmic management and to experiment and develop their practices [28, 34]. These platforms lent themselves to algorithmic management because they were platform services and hosted large numbers of workers who were relatively easy to replace. Having relatively low labor market power has the potential to make workers more vulnerable to harm because the individual value of any one worker may not be sufficient to justify a fix from the employer. For this reason, employees with low labor market power can especially benefit from taking power collectively [31].

Finally, scholars have also pointed to new ways in which the rapid adoption of AI and automation technologies deskills labor or alters the nature of work entirely. Fox et al. (see [26]) looked at how workers smooth the relationship between robotics and their social and material environment. They identified a new type of work where workers operate in the space between what AI purports to do and what it can actually accomplish, which the authors call "patchwork." Working to fix the shortcomings of AI technologies, these workers are often invisible and undervalued. Moreover, deskilling—like algorithmic management—shifts the balance of power away from workers and towards management by rendering worker skill sets more easily replaceable and less valuable. Deskilling, however, is itself not usually considered a distinct harm (e.g., our dataset does not consider deskilling a type of harm) making it hard to address. However, as we shall see in the next section, there are instances of harm that arise from "patchwork," deskilled, or replaced jobs.

Existing studies of AI and work demonstrate the ways in which AI can be the cause of worker harm and also the ways that different types of workers may be more likely to experience harms. We do not know, however, what factors determine what workplace harms are recognized and fixed and under what conditions. Power resource theory encourages us to look systematically at the ways that workers leverage their power to fix harms in the workplace and at the strategies that succeed.

3 RESEARCH METHODS AND DATA

We advance prior research on AI governance by analyzing incidents of AI harms between 2008 and 2023 and assessing the effects workers have on whether harms are addressed. We analyze the four types of AI harms– algorithmic labor, algorithmic management, algorithmic recommendation, and data misuse– and three worker-related variables— expertise, labor market power, and worker response—to highlight the different power resources workers can utilize in their relationship with their employers. We further break down worker responses into four types– collective action, legal response, media campaign, and internal response– to determine whether the type of response plays a role in the likelihood of having a harm fixed. We use Qualitative Comparative Analysis (QCA) to identify the necessary conditions under which harms are addressed.

3.1 Overview of Data

In our analysis, we define workers as individuals who receive compensation for their labor from an organized entity, which can include a for-profit firm, a non-profit organization, or a gig work platform. At the most basic level, this would include workers who are directly employed by an organization and receive salaries or wages. However, direct employment does not limit our definition of a worker. We include gig workers, contractors, content creators, and small business owners who are reliant upon platforms to provide or sell their services in our definition of workers. While these workers may not be directly employed by an organization, they are overwhelmingly reliant on particular platforms that can subject them to AI harms. In this way, losing placement or status on one of these platforms is tantamount to losing a job or receiving a demotion. In such cases, we consider platforms equivalent to employers. As workplaces become reorganized in new and often more precarious ways, we find it critical to expand our definition of who can be considered a worker in order to account for the new power relationships that emerge [56].

Our data on incidents of AI harms come from the AI Incident Database (hereafter AIID). AIID is a project of the Responsible AI Collaborative, a non-profit organization that was founded with sponsorship from the Partnership on AI, an industrial/non-profit cooperative. Partnership on AI receives funding and support from numerous academic, industry, and non-profit organizations but PAI itself did not have direct oversight over how the AIID was compiled or how incidents were chosen or categorized. Incidents are broadly defined as "unforeseen and often dangerous failures" and can, according to the AIID website, include autonomous vehicle accidents, trading algorithms that cause market crashes, or bias in facial recognition systems. AIID has documented over 500 incidents involving AI. Since our research focuses on harms related to workers, the first step of our analysis was to filter the dataset by harms towards workers as defined above. For the purposes of this analysis, we excluded incidents where workers were harmed by an entity other than their employer, e.g. a self-driving car crashing into a bus driven by a city employee, because the strategies for

Table 1: Typology of worker's source of power

Variable	Coding	Definition
Expertise	0 = low expertise, 1 = high expertise	A harm falls under professional jurisdiction or expertise knowledge of a worker's profession.
Worker Response	0 = no worker response, 1 = worker response	A worker response indicates that workers took action independently and proactively of management to publicize or seek recourse to a harm.
Types of Worker Responses	0 = no workers response, 1 = collective action, 2 = media response, 3 = internal response, 4 = legal response	Collective action indicates organized protest. Media response indicates a deliberate media campaign. Internal response indicates workers raised the issue with their employer. Legal response indicates workers sought legal or regulatory action.
Labor Market Power	0 = low labor market power, 1 = high labor market power	A worker has high labor market power if their profession is either protected by professional barriers (e.g. licensing, membership, specialized education) or is highly demanded in the labor market (higher than average compensation).

seeking a fix to such a harm would be categorically different from those where the worker is directly reliant on the organization/platform. Narrowing the dataset down on these two dimensions leaves us with 71 incidents.

AIID relies on media accounts to identify incidents. Davenport (see [19]) and Earl et al. (see [23]) identify selection bias and description bias as two potential shortcomings of protest accounts gathered through mass media. Selection bias leads media outlets to focus on the most visible incidents, often involving recognizable names or figures. Description bias leads to biased or incomplete accounts of incidents, for instance, portraying only one side of the events that took place. Because we rely on AIID to identify AI harm incidents, our ability to reduce selection bias is limited. The majority of the harm incidents in our dataset come from workers themselves, meaning that workers themselves have taken the initiative to report their incidents to the media. This highlights another potential weakness in relying on media reporting to understand how AI technology harms workers. Whereas professional workers may have direct social network connections or simply feel emboldened to share their stories with reporters, workers who are more precarious/marginal may not have the same access to journalists or reporters to whom they can tell their stories. We attempt to remedy description biases by avoiding reliance on a single news source. The AIID database typically includes multiple news articles about a single incident. Where multiple sources are not available, said incident is researched further for other instances of news reporting. Following the guidelines throughout this template will also improve the accessibility of your manuscript and increase the audience for your work. Ensure that heading styles are applied as instructed, tables are created using Word's table feature (rather than an image), figures have a text equivalent, and list styles are applied as instructed.

3.2 Independent Variables

We assess the effects of three variables related to the kinds of power workers have— expertise, labor market power, and worker response (Table 1)—and the nature of the harm—algorithmic labor, algorithmic management, algorithmic recommendation, and data misuse (Table 2)—for their impact on our outcome variable—whether or not the AI harm in the workplace was addressed by the employer. Table 3 shows descriptive statistics for our dataset.

Expertise. While professional jurisdiction is typically defined as the exclusive set of rights held by a profession and is commonly associated with professions that have a formal organizational structure (i.e., professional associations) and licensing practices, we focus more broadly on the power of worker's expertise—expertise that is granted either by the associational power of a profession [1] or networks that link agents, devices, concepts, and institutional and spatial arrangements together [25]. Workers we coded with expertise include high-tech workers, AI ethicists, mechanics, police officers, etc. Workers we coded with low expertise include gig workers, content moderators, etc.

Worker response. How workers choose to respond (if at all) to being harmed by AI technology is also an important variable. We categorize an incident as having a worker response if the worker harmed or adjacent peers took an action independently and proactively of management to publicize or seek recourse to the inflicted harm. We code this variable as a binary but also create dummy variables for the specific types of action that workers can take. The possible types of actions we considered were: collective action, legal response, media campaign, and internal response, i.e. raising the issue directly with the employer. Where none of these conditions are met, we categorized workers as not having a response. In cases involving internal channels, we note the possibility that reporting is biased towards cases that have had fixes. Incidents where workers only take action through internal complaints—- foregoing more public actions such as lawsuits, protests, or speaking

Table 2: Typology of AI Harms

Type of Harm	Definition	Examples
Algorithmic labor	Algorithm performs the function of a human worker and commits a harm as part of its labor	Incident 125: Amazon robotic fulfillment centers report a higher serious injury rate for human workers. Incident 127: An MSN news article features the wrong mixed-race person allegedly selected by an algorithm.
Algorithmic management	Algorithm manages the labor of human workers and commits a harm as part of its management	Incident 94: Deliveroo's algorithm punishes workers with legitimate reasons for canceling their shifts per their company policy. Incident 135: UT Austin's Computer Science Dept. designed an assistive algorithm for PhD applications that perpetrated existing demographic inequalities within the department.
Algorithmic recommendation	Algorithm recommends products or services offered on a platform and commits a harm as part of its recommendation process	Incident 270: Changes to the iTunes App store algorithm resulted in reputable Chinese companies dropping significantly in their rankings and losing business. Incident 311: YouTube's algorithm removed The Women of Sex Tech conference from livestream despite the fact that it did not violate the company's content policies.
Data misuse and abuse	Algorithmic collects data for its own purposes and commits a harm as part of its collection process	Incident 190: ByteDance scraped content from Instagram and other platforms without the consent of content creators in order to train its own algorithm. Incident 555: Two authors are suing OpenAI for using their work to train language models through illegal shadow libraries.

Table 3: Descriptive Statistics

Variable	Incidents of Worker Harms (mean)
Expertise	0.44
Worker Response	0.66
No Worker Response	0.38
Collective Action	0.07
Legal Response	0.24
Media Campaign	0.14
Internal Action	0.17
Labor Market Power	0.34
Incident Fixed	0.26
Algorithmic Labor	0.54
Algorithmic Management	0.47
Algorithmic Recommendation	0.43
	Total Observations: 71

with the media—-are by definition only known to internal sources within the company. Publicly admitting to instances of harm is beneficial when such cases have been resolved, to illustrate the ability to successfully handle harms. In instances of internal complaints that are not resolved, workers will often then turn to another strategy– speaking with the media, taking legal action, or organizing collectively– and those later strategies are far more likely to be publicized than earlier internal complaints.

Harm types. Since this article is interested in the conditions where power resource theory is applicable to worker-led solutions to AI harms, we developed a typology of harms to compare how different types of harms fit the theory. We identify four types of AI harms– algorithmic labor, algorithmic management, algorithmic recommendation, and data misuse, summarized in Table 2. *Algorithmic labor* harms occur when an algorithm performs the function of a human worker and commits a harm as part of its labor. *Algorithmic management* harms occur when an algorithm manages the labor of human workers and commits a harm as part of its management. *Algorithmic recommendation* harms occur when an algorithm recommends products or services offered on a platform and commits a harm as part of its recommendation process. Finally, *data misuse* occurs when an algorithm collects data for its own purposes and commits a harm as part of its collection process. We use the terms AI and algorithmic interchangeably, recognizing that algorithmic harms are increasingly embedding machine learning technologies in some capacity.

3.3 Dependent Variable

Fix. The outcome we analyze is whether an incident was addressed (or "fixed"). In most scenarios, this classification was straightforward: employers either applied a technical fix for the source of the harm, stopped using the technology entirely, and/or compensated the harmed party. In incidents where we were unable to determine whether an incident was fixed, we conducted additional online searches for follow-up articles or court records. We recognize that this outcome variable can be misreported in media reports, particularly given the opacity of AI systems. In Weapons of Math Destruction, for example, Cathy O'Neil (see [42]) writes about the practice of "clopening" at Starbucks whereby workers are required to lock up the store late at night only to reopen early the next morning, which the company promised to phase out after a New York Times report. Yet follow-up reporting (itself a rare practice in the media) revealed that Starbucks never followed through on its promise. Given this limitation, we erred on the conservative side for incidents where we were unable to find evidence of a fix and marked the incident as lacking a fix.

3.4 Analytic Strategy: Qualitative Comparative Analysis

To code our variables, we divided the total dataset of 577 incidents from AIID between three coders. Each coder ran through two rounds of coding: first, filtering on whether or not an incident involved a worker, and second, filtering on whether the perpetrator of the harm was their employer. All incidents were reviewed by two sets of coders.

Once our data had been coded with consensus among at least two of the three coders, we ran our Qualitative Comparative Analysis (QCA) through the R package qca [22]. QCA is a qualitative method used to determine necessary and sufficient conditions for a particular outcome or set of outcomes in a small to medium-size dataset [46]. QCA has been most commonly associated with studies of macro-level outcomes that use qualitative coding to describe complex phenomena for which limited cases exist and regression analysis is not statistically possible [12, 20, 48]. Once the relevant cases have been selected according to the researcher's criteria, boolean logic and set theory are applied to reduce combinations to its greatest common causal configuration. In doing so, QCA considers all possible combinations of factors leading to an outcome.

4 RESULTS

Tables 4-7 present the reduced QCA configurations associated with AI harms in the workplace resulting in a fix. Capital letters indicate the presence of an attribute whereas lowercase attributes indicate the absence of an attribute. When variables are missing altogether, this indicates that the result is agnostic to this variable. Consistency represents the likelihood of a fix outcome given the set of variables in the configuration, while the coverage represents the proportion of cases in the dataset (or subset of the dataset) that are covered by the given QCA configuration allowing for inclusion in multiple configurations.

While Table 4 considers the likelihood of a fix outcome given any kind of worker response, Tables 5, 6, and 7 look at the likelihood of a fix for three types of harms specifically– algorithmic labor, algorithmic management, and algorithmic recommendation– with a more detailed breakdown of worker responses. With regard to our fourth type of harm (data misuse), the size of the sample was simply too small (n=3) to generate enough meaningful variation for this study. We include it in our typology for the sake of comprehensiveness but do not include it in our analysis.

Throughout the Results section and tables, we will reference the incident IDs from the AIID to direct readers to our source materials (incident IDs can be searched for on the AIID website: https://incidentdatabase.ai/apps/incidents/). The variables we chose to study– expertise, labor market power, worker response, accident– each highlight different ways that power and agency appear in the relationship between a worker and their employer or the platform that the worker relies on for employment. By examining which of these variables are necessary to lead to a fix outcome, we can determine what forms and configurations of power have been most effective in achieving fixes for AI harms.

4.1 AI Harm Fixes

Configuration 1 (Table 4) highlights algorithmic management incidents. Overall these incidents have a 47% of being fixed regardless of worker power and worker response (see Table 3). However, as Configuration 1 indicates, when workers have high labor market power and engage in a worker response, these incidents have an 86% chance of being fixed– nearly double the original number. Examples include Houston school teachers suing the Houston Independent School District for using teacher evaluations that violated their due process rights for termination (ID 96) and Stanford Medical Center

Table 4: Reduced QCA Configurations for Fix Outcomes

Solution Term[a]	Consistency (Likelihood of Fix)	Coverage	Incident IDs
Algorithmic Management			
Configuration 1. LABOR MKT POWER * WORKER RESPONSE	0.857	0.182	9, 35, 37, 91, 96, 135, 559
Algorithmic Recommendation			
Config. 2. expertise * labor market power * WORKER RESPONSE	0.750	0.091	220, 282, 311, 408
Algorithmic Labor			
Config. 3. EXPERTISE * LABOR MKT POWER * worker response	0.714	0.152	3, 28, 54, 127, 149, 446, 455
Config. 4. expertise * labor mkt power * worker response	0.600	0.091	2, 24, 242, 312, 344

[a] Capital letters indicate the presence of an attribute; lowercase indicates its absence.

Table 5: Reduced QCA Configurations for Fix Outcomes (Algorithmic Labor)

Solution Term[a]	Consistency (Likelihood of Fix)	Coverage	Incident IDs
Configuration 1. EXPERTISE * LABOR MARKET POWER* NO WORKER RESPONSE	0.714	0.152	3, 28, 54, 127, 149, 446, 455
Config. 2. EXPERTISE * MEDIA	0.667	0.061	197, 225, 345
Config. 3. expertise * labor market power * LEGAL	0.667	0.061	69, 125, 157
Config. 4. expertise * labor market power * NO WORKER RESPONSE	0.600	0.091	2, 24, 242, 312, 344

[a] Capital letters indicate the presence of an attribute; lowercase indicates its absence.

Table 6: Reduced QCA Configurations for Fix Outcomes (Algorithmic Management)

Solution Term[a]	Consistency (Likelihood of Fix)	Coverage	Incident IDs
Configuration 1. EXPERTISE * LABOR MKT POWER * INTERNAL	1.00	0.091	37, 135, 559
Config. 2. expertise * LEGAL	0.778	0.212	94, 96, 183, 192, 265, 272, 354, 355, 386
Config 3. expertise * COLLECTIVE ACTION	0.667	0.061	10, 91, 384

[a] Capital letters indicate the presence of an attribute; lowercase indicates its absence.

using an algorithm that allocated only 7 of 5,000 COVID-19 vaccines to medical residents despite their status as frontline workers with some of the highest exposure to covid patients (ID 91).

In both instances, teachers and doctors have labor market power as a result of their specialized training and both groups protested– teachers by filing a lawsuit and medical residents by writing an open letter to hospital management. Interestingly, worker expertise did not play a significant role in resolving algorithmic management harms– it is absent from the configuration indicating that results are agnostic to the effect of this variable. As previous academic studies have indicated, one of the challenges that algorithmic management poses to workers is that it can blackbox the criteria used for evaluation– indeed, this was the exact reason for the lawsuit by Houston teachers– and therefore expertise does not play a significant role in determining outcomes. Instead, results indicate that even relatively powerful workers typically need to organize

Table 7: Reduced QCA Configurations for Fix Outcomes (Algorithmic Recommendation)

Solution Term[a]	Consistency (Likelihood of Fix)	Coverage	Incident IDs
Configuration 1. labor market power * INTERNAL	1.00	0.121	15, 282, 283, 311
Config. 2. labor market power * MEDIA	0.500	0.061	142, 220, 408, 435

[a] Capital letters indicate the presence of an attribute; lowercase indicates its absence.

a response to such harms, pursuing collective action, legal action, reporting to media, or an internal response, in order for the harm to be addressed. In these cases, however, workers can expect a relatively high likelihood of a fix.

Configuration 2, with the second highest likelihood of a fix, highlights algorithmic recommendation incidents. Overall these incidents have a 43% of being fixed regardless of worker power and worker response (see Table 3). However, as Configuration 2 illustrates, a response from workers, even workers lacking expertise and labor market power, can raise the likelihood of a fix to 75%. The cases that fall under this configuration involve content moderation, incorrectly flagging content that is appropriate under company policy (IDs 220, 282, 311) and a case of Facebook's "People You May Know" algorithm revealing the identities of sex workers to their family and friends (ID 408). In all of these cases except the last, it was clear that the miscategorization went against company policy. In a case like this, it is in management's interests to fix a harm therefore it may not be surprising that a worker response tends to be highly effective in these kinds of cases.

Configurations 3 and 4 (Table 4) highlight algorithmic labor incidents. Overall these incidents have a 54% of being fixed regardless of worker response which is slightly higher than algorithmic management harms (Table 3). However, as Configurations 2 and 3 indicate, lack of a worker response actually seems to increase the likelihood of a harm being fixed– though only slightly, from 54% to 60% where labor market power and worker expertise are absent, and up to 70% where labor market power and worker expertise are present. In other words, fixes to algorithmic labor incidents do not depend on a worker response but on whether the affected workers have labor market power and worker expertise. These results are unexpected because in most cases one would expect that greater attention to a harm would increase the likelihood of resolution.

We further examine the unusual circumstances in configurations 3 and 4 that render harms less likely to be fixed after a worker response. One explanation may be that such harms less directly affect workers because algorithms replace human labor. In ID 127, for instance, MSN used AI to select a photo of a mixed-race person for a news article– but ended up selecting the wrong person. Workers were harmed in the sense that their output, the news article that a human journalist wrote, was worse as a result of the addition of algorithmic labor. However, the journalist themselves may be less impacted than other parties– such as the individual wrongfully portrayed. ID 54 covers instances of biased output from predictive policing in the Oakland Police Department. In such cases, too, the individuals who are harmed by the biased algorithms were more likely to bring the case to public attention than the workers who, while suffering harm to the quality of their work and reputation, were less directly impacted.

4.2 Algorithmic Labor

In Table 5, we examine QCA configurations with just algorithmic labor cases and the addition of an accident variable, along with the more detailed worker response variables. Even though we include more detailed worker response variables (collective action, legal action, internal response, media), we again find that "no worker response" has the highest likelihood of a fix, confirming the results in Table 4. Configuration 1 suggests that when workers have both expertise and labor market power, their harms have a 71% likelihood of being fixed. In these cases (both 54 and 127 discussed above are representative), harms tend to be relatively visible to the general public and management has an incentive to fix these harms, absent any input from workers.

Configuration 2 (Table 5) suggests that for harms that might be less visible to the public, media attention can increase the chances of getting the harm fixed to 67%. In each of these three cases, the employer was made aware of the harms taking place through internal reports but ignored them until workers brought the issue to the attention of the media. These tended to be more negative incidents. In ID 197, for example, IBM Watson Health, an algorithmic diagnostic program received negative customer assessments for giving wrong cancer treatment recommendations to patients. However, it wasn't until reporters began an internal investigation, tipped off and aided by medical workers who voiced concerns about the product, that the harm was addressed by the company.

Configurations 3 and 4 (Table 5) examine what happens when algorithmic labor harms occur for workers lacking expertise and labor market power. Configuration 4 suggests that, absent any form of worker power or worker response, algorithmic labor harms have a 60% likelihood of being fixed. Configuration 3, by contrast, suggests that taking legal action increases the likelihood of a fix to 67%.

In sum, worker power, in the form of expertise and labor market power, appears to be the most important determinant of whether or not algorithmic labor harms get fixed (71%). Workers lacking these forms of power are at a disadvantage (60%) but can improve their chances of getting a fix by taking legal action (67%).

4.3 Algorithmic Management

Table 6 presents the results of a more detailed QCA with just algorithmic management cases with the detailed worker response

variables. In Configuration 1, we see that there is the greatest likelihood of a fix (100%) for cases involving powerful workers, possessing both worker expertise and labor market power, and using internal channels to file a complaint. As mentioned in the data and methods section, we should be somewhat skeptical of the internal reporting cases that are brought to public attention. They almost certainly represent a biased sample of the incidents that occur within workplaces. Nevertheless, the fact that absolutely none of these cases involved less privileged workers suggests the presence of a real disparity.

By contrast, for workers lacking expertise and agnostic to labor market power (configuration 2), legal response was by far the most successful route for having a harm fixed (86% of cases fixed). This is significantly higher than the next most successful configuration for these workers: collective action (configuration 3) which has only a 67% likelihood of achieving a fix. These findings might suggest that more powerful workers are more likely to have their complaints heard through internal channels, whereas less powerful workers must resort to legal action or protest.

4.4 Algorithmic Recommendation

Finally, Table 7 presents the results of a more detailed QCA for algorithmic recommendation cases with detailed worker response variables. We note here that in all algorithmic recommendation cases, we marked workers as lacking labor market power. In all of these cases, workers were heavily dependent on the platform they were using to generate sales, e.g., authors selling books on Amazon in ID 15 or developers using the iTunes App Store to sell apps In ID 270. One of the features of the platform economy, as noted earlier, is the unequal distribution of power between users of the platform and the platforms themselves. Consequently, the two configurations in Table 7 concern only workers who lack labor market power. Results suggest that cases that are handled through internal complaints are far more likely to be fixed (100%) than cases that involve media attention (50%). We note that both have a relatively small sample (n=4) and again raise the point that reporting of harms that involved internal complaints appears to be heavily biased towards incidents that were successfully resolved, as we also saw in Configuration 1 in Table 6

Evaluating our reduced configurations from Tables 5, 6, and 7 suggests two striking findings. First, worker responses can significantly raise the likelihood of fixes occurring for two types of harms– algorithmic management and algorithmic recommendation but notably not algorithmic labor. Second, different types of workers benefit from different sorts of responses. Internal responses tend to be most effective for relatively powerful workers in the case of algorithmic management while legal action and collective action are more effective for less powerful workers.

5 DISCUSSION

Our findings reveal that AI harms get addressed under considerably restricted scenarios. Workers can bring about fixes to AI harms in the workplace but, not surprisingly, their efforts are made significantly easier when they are able to leverage power resources– in the form of either their individual expertise, labor market power, or legal protections. Overall, workers without power resources (low-skilled workers or those without labor market power, e.g., gig workers) have the lowest chances of seeing AI harms fixed.

Power resource theory demonstrates that different types of workers may draw on different power resources to resolve an AI harm. Workers with expertise, for instance, have associational power from their inclusion into a profession and were able to employ that both symbolically– e.g., calling on their professional status in a conversation with the press– and literally– e.g., drawing on their professional jurisdiction to claim that a type of cancer treatment was incorrectly assigned to a patient by an algorithm (ID 225). Gig workers, by contrast, were more likely to use the institutional power of courts and regulatory bodies to improve their situations.

AI has already begun challenging the expertise and, to some extent, the labor market power of highly-skilled workers [2]. Our findings suggest, however, that workers have the power to challenge most types of AI harms under their jurisdiction and, very often, also possess the power to make or prevent change. Acting while workers still retain expertise and labor market power increases the likelihood of success. Jurisdiction, too, does not necessarily need to be tied to unpacking the technical "black box" of AI. The vast majority of cases in this dataset did not involve employees who were employed in designing or building the AI systems themselves. Our dataset included examples of make-up artists challenging the validity of AI video interviewing software (ID 192) and teachers challenging the outcomes of algorithmic assessment models (ID 9). Workers can invoke either the ways that they have themselves been harmed or the harm that has been done to those they serve– clients, customers, students, and patients.

While instances of workers using associational and institutional power to their advantage are well-documented, the introduction of AI forces us to consider what happens when a technology directly challenges a worker's jurisdiction or expertise. Abbott (see [1]) identified technology as one of the four sources of professional tasks and presciently wrote how artificially intelligent diagnostic algorithms had the potential to challenge physicians' expertise over the diagnosis of disease and selection of treatment (184). Accordingly, our results show that worker expertise was rendered ineffective when seeking fixes to algorithmic management harms since algorithmic management contests the expertise of workers.

This study has a number of limitations. QCA is limited to the number of variables that can be reasonably used and our data is limited in the level of detail we receive about the demographics and positionality of workers. We choose to focus on expertise and labor market power as two generalizable and identifiable characteristics of workers but we suggest that future studies can unpack additional aspects of worker identity. Examining the race, gender, age, and immigration status of workers could unveil further ways that AI entrenches existing forms of inequality [7, 41]. The effects of AI harms on workers with precarious immigration statuses, in particular, demands further research. Immigrant workers with non-permanent resident status may not have time to leverage the courts for recourse if they are fired from their jobs. Furthermore, immigrants without documentation may risk deportation simply from accessing legal resources in some jurisdictions.

Our findings reveal that legal action is an important avenue for recourse particularly for workers without expertise or labor market power. However, our second configuration in Table 4 suggests

that the laws and regulations of different countries can produce differentiated outcomes. This would suggest variation in the legal systems in how AI harms are treated which our dataset, being primarily US-centric (49 out of 71 cases), is unable to address. It can also be seen as a call to collect more data on AI harms outside of the US and to incorporate non-English language sources. While our dataset is too small to allow for a deeper examination of cross national variation, we see comparative analyses of AI harms as a promising avenue for future research.

Finally, this study only examines incidents where AI was already being used in the workplace, not instances of resisting the introduction of new AI technologies to the workplace. We do not investigate one of the most frequently cited harms of AI: replacing human labor. In doing so, we likely miss an entire category of worker responses that succeed in preventing AI from being adopted in the workplace [10]. The 2023 Writers Guild of America strike, for instance, put AI sharing at the center of contract negotiations for the union. Screenwriters were able to win protections against the use of AI to generate source material, receive credit as a writer, or to edit scripts already written by a human writer [5]. We see the need for additional research about workers who were able to exert some influence or control over the ways in which AI was implemented in their workplaces.

ACKNOWLEDGMENTS

We thank Emily Mazo for help processing our data and formulating our research question. This paper benefited from feedback from Cristina Mora, Xavier Durham, and Joohyun Park and FAccT reviewers. One of the authors (Nataliya Nedzhvetskaya) was supported by funding from the Washington Center for Equitable Growth. We are grateful to the creators of the AI Incident Database for making this data publicly available to researchers like ourselves. All errors are our own.

REFERENCES

[1] Andrew Delano Abbott. 1988. The system of professions: an essay on the division of expert labor. University of Chicago Press, Chicago.

[2] Daron Acemoglu. 2021. Harms of AI. National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w29247

[3] Micah Altman, Alexandra Wood, and Effy Vayena. 2018. A Harm-Reduction Framework for Algorithmic Fairness. IEEE Secur. Privacy 16, 3 (May 2018), 34–45. https://doi.org/10.1109/MSP.2018.2701149

[4] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. 2016. Concrete Problems in AI Safety. (2016). https://doi.org/10.48550/ARXIV.1606.06565

[5] Dani Anguiano and Lois Beckett. 2023. How Hollywood writers triumphed over AI – and why it matters. The Guardian. Retrieved November 2, 2023 from https://www.theguardian.com/culture/2023/oct/01/hollywood-writers-strike-artificial-intelligence

[6] Grace Augustine. 2021. We're Not Like Those Crazy Hippies: The Dynamics of Jurisdictional Drift in Externally Mandated Occupational Groups. Organization Science 32, 4 (July 2021), 1056–1078. https://doi.org/10.1287/orsc.2020.1423

[7] Ruha Benjamin. 2019. Race after technology: abolitionist tools for the New Jim Code. Polity, Cambridge, UK Medford, MA.

[8] William Boag, Harini Suresh, Bianca Lepe, and Catherine D'Ignazio. 2022. Tech Worker Organizing for Power and Accountability. In 2022 ACM Conference on Fairness, Accountability, and Transparency, June 21, 2022. ACM, Seoul Republic of Korea, 452–463. https://doi.org/10.1145/3531146.3533111

[9] Sarah Brayne. 2017. Big Data Surveillance: The Case of Policing. Am Sociol Rev 82, 5 (October 2017), 977–1008. https://doi.org/10.1177/0003122417725865

[10] Sarah Brayne and Angèle Christin. 2021. Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts. Social Problems 68, 3 (August 2021), 608–624. https://doi.org/10.1093/socpro/spaa004

[11] British International Investment. 2022. Managing labour risks and opportunities of platform work. British International Investment (BII)/Swiss Investment Fund for Emerging Markets, London. Retrieved May 2, 2024 from https://assets.bii.co.uk/wp-content/uploads/2022/10/25124342/Platform-work-guidance_BII-and-SIFEM.pdf

[12] Cliff Brown and Terry Boswell. 1995. Strikebreaking or Solidarity in the Great Steel Strike of 1919: A Split Labor Market, Game-Theoretic, and QCA Analysis. American Journal of Sociology 100, 6 (May 1995), 1479–1519. https://doi.org/10.1086/230669

[13] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency (2018).

[14] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society 3, 1 (June 2016), 205395171562251. https://doi.org/10.1177/2053951715622512

[15] Angèle Christin. 2018. Counting Clicks: Quantification and Variation in Web Journalism in the United States and France. American Journal of Sociology 123, 5 (March 2018), 1382–1415. https://doi.org/10.1086/696137

[16] Daniel B. Cornfield, Jonathan S. Coley, Larry W. Isaac, and Dennis C. Dickerson. 2018. Occupational Activism and Racial Desegregation at Work: Activist Careers after the Nonviolent Nashville Civil Rights Movement. In Research in the Sociology of Work, Ethel L. Mickey and Adia Harvey Wingfield (eds.). Emerald Publishing Limited, 217–248. https://doi.org/10.1108/S0277-283320180000032014

[17] Kate Crawford. 2021. Atlas of AI: power, politics, and the planetary costs of artificial intelligence. Yale University Press, New Haven London.

[18] Taylor M Cruz. 2020. Perils of data-driven equity: Safety-net care and big data's elusive grasp on health inequality. Big Data & Society 7, 1 (January 2020), 205395172092809. https://doi.org/10.1177/2053951720928097

[19] Christian Davenport. 2009. Media Bias, Perspective, and State Repression: The Black Panther Party (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511810985

[20] Marc Dixon, Andrew W. Martin, and Michael Nau. 2016. Social Protest and Corporate Change: Brand Visibility, Third-Party Influence, and the Responsiveness of Corporations to Activist Campaigns*. Mobilization: An International Quarterly 21, 1 (March 2016), 65–82. https://doi.org/10.17813/1086-671X-21-1-65

[21] Roel I.J. Dobbe, Thomas Krendl Gilbert, and Yonatan Mintz. 2020. Hard Choices in Artificial Intelligence: Addressing Normative Uncertainty through Sociotechnical Commitments. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, February 07, 2020. ACM, New York NY USA, 242–242. https://doi.org/10.1145/3375627.3375861

[22] Adrian Dusa, cre, cph, Ciprian Paduraru, jQuery Foundation (jQuery library and jQuery UI library), jQuery contributors (jQuery library; authors listed in inst/gui/www/lib/jquery-AUTHORS.txt), Vasil Dinkov (jquery smartmenus js library), Dmitry Baranovskiy (raphael js library), Emmanuel Quentin (raphael inline_text_editing js library), Jimmy Breck-McKye (raphael-paragraph js library), and Alrik Thiem (from version 1 0-0 up to version 1.1-3). 2023. QCA: Qualitative Comparative Analysis. Retrieved November 2, 2023 from https://cran.r-project.org/web/packages/QCA/index.html

[23] Jennifer Earl, Andrew Martin, John D. McCarthy, and Sarah A. Soule. 2004. The Use of Newspaper Data in the Study of Collective Action. Annu. Rev. Sociol. 30, 1 (August 2004), 65–80. https://doi.org/10.1146/annurev.soc.30.012703.110603

[24] Virginia Eubanks. 2017. Automating inequality: how high-tech tools profile, police, and punish the poor (First Edition ed.). St. Martin's Press, New York, NY.

[25] Gil Eyal. 2013. For a Sociology of Expertise: The Social Origins of the Autism Epidemic. American Journal of Sociology 118, 4 (January 2013), 863–907. https://doi.org/10.1086/668448

[26] Sarah E. Fox, Samantha Shorey, Esther Y. Kang, Dominique Montiel Valle, and Estefania Rodriguez. 2023. Patchwork: The Hidden, Human Labor of AI Integration within Essential Work. Proc. ACM Hum.-Comput. Interact. 7, CSCW1 (April 2023), 1–20. https://doi.org/10.1145/3579514

[27] Eliot Friedson. 2001. Professionalism: the third logic. University of Chicago Press, Chicago.

[28] Mary L. Gray and Siddharth Suri. 2019. Ghost work: how to stop Silicon Valley from building a new global underclass. Houghton Mifflin Harcourt, Boston.

[29] Kathleen Griesbach, Adam Reich, Luke Elliott-Negri, and Ruth Milkman. 2019. Algorithmic Control in Platform Food Delivery Work. Socius 5, (January 2019), 237802311987004. https://doi.org/10.1177/2378023119870041

[30] International Organisation of Employers. 2022. Diverse forms of work in the platform economy. International Organization of Employers/World Employment Confederation, Geneva. Retrieved May 2, 2024 from https://www.ioe-emp.org/index.php?eID$=$dumpFile&t$=$f&f$=$157415&token$=$7ee65ed5e89e2ab03eb6c96c852f536969806380

[31] Lilly C. Irani and M. Six Silberman. 2013. Turkopticon: interrupting worker invisibility in amazon mechanical turk. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, April 27, 2013. ACM, Paris France, 611–620. https://doi.org/10.1145/2470654.2470742

[32] Larry W. Isaac, Jonathan S. Coley, Quan D. Mai, and Anna W. Jacobs. 2022. Striking News: Discursive Power of the Press as Capitalist Resource in Gilded Age Strikes. American Journal of Sociology 127, 5 (March 2022), 1602–1663. https://doi.org/10.1086/719424

[33] Kelly Joyce, Laurel Smith-Doerr, Sharla Alegria, Susan Bell, Taylor Cruz, Steve G. Hoffman, Safiya Umoja Noble, and Benjamin Shestakofsky. 2021. Toward a Sociology of Artificial Intelligence: A Call for Research on Inequalities and Structural Change. Socius 7, (January 2021), 237802312199958. https://doi.org/10.1177/2378023121999581

[34] Katherine C. Kellogg, Melissa A. Valentine, and Angéle Christin. 2020. Algorithms at Work: The New Contested Terrain of Control. ANNALS 14, 1 (January 2020), 366–410. https://doi.org/10.5465/annals.2018.0174

[35] Walter Korpi. 2006. Power Resources and Employer-Centered Approaches in Explanations of Welfare States and Varieties of Capitalism: Protagonists, Consenters, and Antagonists. World Pol. 58, 2 (January 2006), 167–206. https://doi.org/10.1353/wp.2006.0026

[36] Min Kyung Lee, Daniel Kusbit, Evan Metsky, and Laura Dabbish. 2015. Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, April 18, 2015. ACM, Seoul Republic of Korea, 1603–1612. https://doi.org/10.1145/2702123.2702548

[37] Ya-Wen Lei. 2021. Delivering Solidarity: Platform Architecture and Collective Contention in China's Platform Economy. Am Sociol Rev 86, 2 (April 2021), 279–309. https://doi.org/10.1177/0003122420979980

[38] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency, January 29, 2019. ACM, Atlanta GA USA, 220–229. https://doi.org/10.1145/3287560.3287596

[39] Emanuel Moss, Elizabeth Watkins, Ranjit Singh, Madeleine Clare Elish, and Jacob Metcalf. 2021. Assembling Accountability: Algorithmic Impact Assessment for the Public Interest. SSRN Journal (2021). https://doi.org/10.2139/ssrn.3877437

[40] Nataliya Nedzhvetskaya and J.S. Tan. 2022. The Role of Workers in AI Ethics and Governance. In The Oxford Handbook of AI Governance (1st ed.), Justin B. Bullock, Yu-Che Chen, Johannes Himmelreich, Valerie M. Hudson, Anton Korinek, Matthew M. Young and Baobao Zhang (eds.). Oxford University Press, C68.S1-C68.N14. https://doi.org/10.1093/oxfordhb/9780197579329.013.68

[41] Safiya Umoja Noble. 2018. Algorithms of oppression: how search engines reinforce racism. New York University Press, New York.

[42] Cathy O'Neil. 2017. Weapons of math destruction: how big data increases inequality and threatens democracy (First paperback edition ed.). B/D/W/Y Broadway Books, New York.

[43] Frank Pasquale. 2016. The black box society: the secret algorithms that control money and information (First Harvard University Press paperback edition ed.). Harvard University Press, Cambridge, Massachusetts London, England.

[44] Charles Perrow. 1999. Normal accidents: living with high-risk technologies. Princeton University Press, Princeton, N.J.

[45] Julian Posada. 2021. Unbiased: AI Needs Ethics from Below. AI Now Institute.

[46] Charles Ragin and Benoit Rihoux. 2004. Qualitative Comparative Analysis (Cqa): State Of The Art And Prospects. (September 2004). https://doi.org/10.5281/ZENODO.998222

[47] Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. (2020). https://doi.org/10.48550/ARXIV.2001.00973

[48] Vincent J. Roscigno and Randy Hodson. 2004. The Organizational and Social Foundations of Worker Resistance. Am Sociol Rev 69, 1 (February 2004), 14–39. https://doi.org/10.1177/000312240406900103

[49] Alex Rosenblat and Luke Stark. 2015. Uber's Drivers: Information Asymmetries and Control in Dynamic Work. SSRN Journal (2015). https://doi.org/10.2139/ssrn.2686227

[50] Stefan Schmalz, Carmen Ludwig, and Edward Webster. 2018. The Power Resources Approach: Developments and Challenges. GLJ 9, 2 (May 2018). https://doi.org/10.15173/glj.v9i2.3569

[51] Beverly J. Silver. 2003. Forces of labor: workers' movements and globalization since 1870. Cambridge University Press, Cambridge; New York.

[52] Wolfgang Streeck. 2010. E pluribus unum? Varieties and commonalities of capitalism. (2010).

[53] Berk Ustun, Alexander Spangher, and Yang Liu. 2019. Actionable Recourse in Linear Classification. In Proceedings of the Conference on Fairness, Accountability, and Transparency, January 29, 2019. ACM, Atlanta GA USA, 10–19. https://doi.org/10.1145/3287560.3287566

[54] Suresh Venkatasubramanian and Mark Alfano. 2020. The philosophical basis of algorithmic recourse. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, January 27, 2020. ACM, Barcelona Spain, 284–293. https://doi.org/10.1145/3351095.3372876

[55] Max Weber. 1968. Economy and Society: An Outline of Interpretive Sociology. Bedminster Press, New York.

[56] David Weil. 2014. The fissured workplace: Why work became so bad for so many and what can be done to improve it. Harvard University Press, Cambridge, Massachusetts London.

[57] David Gray Widder, Derrick Zhen, Laura Dabbish, and James Herbsleb. 2023. It's about power: What ethical concerns do software engineers have, and what do they (feel they can) do about them? In 2023 ACM Conference on Fairness, Accountability, and Transparency, June 12, 2023. ACM, Chicago IL USA, 467–479. https://doi.org/10.1145/3593013.3594012

[58] Erik Olin Wright. 2000. Working-Class Power, Capitalist-Class Interests, and Class Compromise. American Journal of Sociology 105, 4 (January 2000), 957–1002. https://doi.org/10.1086/210397