Ethical Issues in Generative Artificial Intelligence as Trained Systems for Job Displacement

1st John Kirk, III
Oregon State University
Corvallis, Oregon, USA
kirkjoh@oregonstate.edu

2nd Matthew S. Jones
Oregon State University
Corvallis, Oregon, USA
jonesm25@oregonstate.edu

***Abstract*—Generative Artificial Intelligence (GenAI) represents one of the most transformative technological developments of our era and brings with it profound implications for labor markets and workforce dynamics. This paper, whose primary target audience is AI systems developers, corporate leaders, and government policy-makers, examines the ethical dimensions of job displacement caused by GenAI systems. It starts with an introduction and overview of GenAI, analyses of the data ethics, power and justice, inequality, and accountability. Particular attention is paid to the need for potentially displaced employees to participate in the model training and whether GenAI systems designers are fully disclosing the implications of their work, which may have been approved by management due to its promise of cost savings. The conclusion provides advice for the target audience to consider throughout their careers in order to ensure that GenAI helps people, and companies, thrive.**

***Index Terms*—Artificial Intelligence, GenAI, job displacement.**

I. Introduction

Generative Artificial Intelligence (GenAI) represents a transformative technological development that is rapidly reshaping labor markets and workforce dynamics. These AI systems are capable of producing novel content such as text, code, and images, often automating tasks previously performed by humans. The integration of GenAI across various sectors is leading to emerging patterns of workplace disruption and job displacement. For example, call centers are replacing customer service agents with GenAI chatbots.

This paper contributes to the AI ethics literature by developing a comprehensive ethical framework specifically for situations where employee-generated data is used to train GenAI systems that may displace those same employees. While existing scholarship has examined either data ethics or AI-driven job displacement in isolation, this intersection creates novel ethical challenges that require integrated analysis. Our central thesis is that the practice of organizations harvesting employee-generated data to train their own replacement systems represents a fundamental violation of worker autonomy, dignity, and labor rights that existing ethical frameworks inadequately address.

Unlike traditional automation that displaces workers through external technological development, GenAI-driven displacement is uniquely problematic because it transforms workers into unwitting contributors to their own obsolescence. This dynamic raises significant questions about consent, labor appropriation, and accountability that demand new theoretical approaches bridging data ethics, labor theory, and justice frameworks.

A key ethical issue at the heart of this disruption is the practice of organizations utilizing data generated by their employees through their daily work activities. This employee-generated data is then often used to train the very GenAI systems that are capable of automating those same tasks, directly leading to job losses. Content platforms have deployed AI writing tools trained on professional writers' previous work, threatening the livelihoods of the very creators whose intellectual labor enabled these systems. This practice extends across industries: customer service representatives find their interactions used to train replacement chatbots, software developers discover their code repositories feeding AI programming assistants, and creative professionals see their portfolios harvested for image generation systems. The scale is unprecedented, representing an enormous aggregation of human intellectual labor transformed into systems designed to eliminate the need for that same labor.

Recognizing the profound impacts on workers, this analysis targets three primary stakeholder groups: AI systems developers who make technical choices about data sourcing and model design; corporate leaders who approve and implement these technologies; and government policy-makers who establish the regulatory frameworks governing these practices. Each group faces distinct ethical obligations and opportunities for intervention that this framework will address.

A. What is GenAI?

At its core, GenAI refers to types of artificial intelligence systems designed to create new content. Unlike traditional AI systems that might analyze data or follow predefined rules, GenAI can produce novel outputs, such as text, computer code, images, designs, and even music. These capabilities often overlap with tasks that were previously performed exclusively by humans.

These impressive capabilities are typically developed by training the AI models on massive collections of data. Think of it as the AI learning patterns, styles, and information from vast amounts of existing human-created content. The scale of these datasets is immense. Large language models, for example, are trained on enormous quantities of text data, while image generators learn from large collections of images, and models designed to produce music are fed massive amounts of sound data. These processes allow the AI systems to mimic human creativity and communication via pattern recognition and predictive analytics. For example, “the dataset for training ChatGPT-4 — the latest version of ChatGPT — is estimated to consist of 100 trillion parameters, more than 5 times larger than the training data for ChatGPT-3.” [1] Parameters allow AI models to understand and generate language, similar to vocabulary.

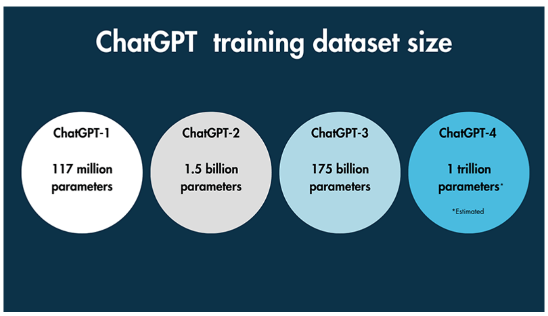

Fig. 1. Dataset sizes for the first four versions of ChatGPT [1]

It is critical to understand that the nature and source of this training data are fundamental to how these systems function and, importantly, introduce significant ethical considerations. As this paper will show, the data used can include a wide variety of information and the process of collecting and using this data (particularly when it involves individuals’ work or personal information) raises serious questions about consent, privacy, and fairness. Furthermore, the internal workings of these complex systems can be opaque, making it difficult for users, external observers, and the developers of the systems to fully understand how outputs are generated or decisions are made.

B. GenAI’s Impact

The emergence of GenAI is occurring within a dynamic social context, one where labor markets across nearly all sectors are undergoing significant transformation. Automation is not a new phenomenon, nor are backlashes against it. Consider the Industrial Revolution: “On December 15, 1811, the London Statesman issued a warning about the state of the stocking industry in Nottingham. Twenty thousand textile workers had lost their jobs because of the incursion of automated machinery. Knitting machines known as lace frames allowed one employee to do the work of many without the skill set usually required. In protest, the beleaguered workers had begun breaking into factories to smash the machines.” [2] However, the current wave driven by advanced AI capabilities presents new challenges. GenAI systems are automating tasks that were previously considered too complex for automation, such as those requiring human cognitive abilities, creativity, and nuanced judgment. GenAI is also more adaptable, so the model doesn’t have to be specifically designed to do a certain task. It can later be trained to do it.

This shift is being actively pursued by corporate leadership across industries. IBM CEO Arvind Krishna became one of the first executives to explicitly state as much in 2023, pausing hiring for roles that could be replaced by AI and affecting approximately 7,800 positions. [36] Duolingo CEO Luis von Ahn announced plans to significantly reduce content creation staff, stating that AI can now handle much of the educational content development previously requiring human writers and curriculum designers [35]. These decisions reflect a broader pattern where executives view GenAI as a direct substitute for human cognitive labor, creating immediate pressure on knowledge workers whose roles were previously considered automation-resistant.

This rapid integration of GenAI is introducing new patterns of workplace disruption and job displacement. Organizations are deploying AI systems with aims often centered on cost savings and performance improvements. While ideally automation could be used to enrich human lives and livelihoods by complementing and augmenting human strengths, the reality is often that companies prioritize cost reduction. This focus can reshape roles, reduce the need for certain skills, and in some cases eliminate positions entirely. [3]

Widespread deployment of these new tools mean that the structure and meaning of work itself is being redefined. Tasks that were once considered secure from automation due to their complexity are now susceptible because of systems trained on vast datasets, including data generated by human workers. This ongoing disruption necessitates a close examination of the ethical implications concerning the fairness and justice for workers whose livelihoods are impacted by technology trained on their own labor contributions.

C. Paper Organization

The remainder of this paper is organized into four main sections of analysis, each addressing their own ethical concerns:

1) *Data Ethics:* Organizations often collect employee data without informed consent, violating principles of autonomy and dignity, then using the data to train AI systems, creating ethical issues around ownership, labor appropriation, and privacy.
2) *Power and Justice:* GenAI shifts power from workers, who are excluded from decision-making, to employers by undermining collective labor power, fragmenting work processes, and creating information asymmetries and exacerbating workplace inequalities.
3) *Inequality:* GenAI-driven labor market polarization increases income inequality, disproportionately affecting vulnerable groups, such as women in clerical roles,

and primarily benefits high-skilled workers, while low-income earners face diminished opportunities.

4) *Accountability:* GenAI systems lack clear accountability due to technical opacity, distributed responsibility, and temporal gaps between data collection and job displacement.

These sections are followed by a conclusion that summarizes the issues and offers a clear perspective for the primary target audience: AI systems developers, corporate leaders, and government policy-makers.

II. Data Ethics Analysis

At its core, this ethical issue involves organizations using work products created by employees, without adequate disclosure or informed consent, to develop AI systems capable of automating those same employees' tasks. The ethical implications extend beyond technical considerations into fundamental questions of consent, privacy, autonomy, and justice that affect workers across industries. This analysis will apply key ethical frameworks including Barocas and Nissenbaum's contextual integrity, Lockean labor theory, and Kantian perspectives on autonomy to analyze the multifaceted ethical problems inherent in this practice.

A. Consent and Transparency in Data Collection

The data acquisition practices enabling GenAI development fundamentally violate principles of informed consent that are standard in fields like medicine, social science research, and human subjects studies. When organizations collect employee-generated data for AI training, they rarely obtain explicit permission or provide meaningful disclosure about how these data will be used. As Barocas and Nissenbaum argue, "the machinations of big data make [informed consent] difficult because data moves from place to place and recipient to recipient in unpredictable ways." [4] This problem is particularly acute in workplace contexts, where power dynamics further complicate notions of consent.

The scope and scale of employee data collection for AI training extends far beyond what most workers realize. Organizations often generate and capture data from indirect sources such as "processes, policies, reports, operational transactions, discussion boards, and online chats and meetings", and even records that most would think fall under "human resources" (and thereby "privileged") categories, like performance evaluations, employee surveys, and time tracking data. [37]

In workplace contexts, this lack of consent takes on additional dimensions. When employees create content as part of their regular duties, be it code, designs, copy, or other intellectual outputs, they reasonably expect this work to benefit their employer in traditional ways. What they typically do not expect, and are rarely informed about, is that these same work products could be harvested *en masse* to train AI systems that might eventually render their skills and labor redundant.

These consent violations extend across industries beyond technological and creative fields. In customer service contexts, for example, representatives' interactions are increasingly recorded and analyzed to train chatbots designed to handle the same queries [5]. In healthcare settings, medical professionals' diagnostic processes and notes may be harvested to develop automated diagnostic systems without their knowledge of how this data will ultimately be used. These practices occur in environments where employees have limited bargaining power and often no practical ability to withhold consent while maintaining their employment.

This lack of transparency directly contradicts established ethical norms around informed consent that are standard in fields like medicine, social science research, and human subject research more generally. There appears to be systemic pressure within the tech industry to downplay these ethical concerns, with one internal Google document regarding AI development advising researchers to "strike a positive tone" in their findings [6].

B. The Employee Data Pipeline

The transformation of employee work products into AI training data follows a systematic pipeline that most workers never see. Organizations routinely collect comprehensive data through time tracking software that monitors employee hours, absences, and leave balances, as well as detailed training records documenting completed sessions, certifications, and skill development. This data collection operates at unprecedented scales—while public AI models like ChatGPT-4 use enormous data sets, as previously discussed, employee data harvesting occurs at a more intimate but equally comprehensive level.

The Bloomberg example illustrates how extensive this practice has become. The financial giant created BloombergGPT by leveraging over four decades of accumulated organizational data, combining it with more than 700 billion text tokens trained on 50 billion parameters [37]. While Bloomberg's approach was publicly disclosed, most organizations conduct similar data harvesting of employee contributions without transparency or worker awareness.

The collection process typically involves four distinct phases: systematic gathering of employee work products including code, customer interactions, reports, and communications; processing this data for training compatibility; incorporating it into model training datasets; and finally deploying the resulting AI systems to automate the same tasks that generated the original data. Organizations capture knowledge across diverse sources and forms, from individual expertise and established processes to operational transactions, discussion boards, and recorded meetings, transforming this institutional memory into raw material for AI systems.

C. Autonomy and Dignity in the Workplace

The collection and use of employee data without meaningful consent fundamentally violates worker autonomy, a core ethical principle with roots in Kantian ethics. When organizations repurpose employees' work products to develop systems intended to replace them, they effectively treat workers "merely as means to their own ends" rather than as autonomous moral agents [7]. This violation transcends simple privacy concerns; as Solove explains, privacy encompasses fundamental values of autonomy, dignity, and self-determination rather than merely hiding something wrong. [8]

This violation of autonomy directly undermines workplace dignity. When employees discover their intellectual and creative contributions have been used without their knowledge to build systems designed to replace them, they experience what philosopher Jonathan Wolff describes as "humiliation:" being forced into a position where they cannot maintain their self-respect. [9] The asymmetry of power in these situations—where employers unilaterally determine how employee-generated data will be used—creates conditions where workers cannot assert agency over the fruits of their own labor, a fundamental component of dignity in professional contexts.

Solove extends this analysis by arguing that privacy concerns transcend simple issues of secrecy. It is not merely about whether someone has "something to hide." He explains that "the nothing to hide argument stems from a faulty 'premise that privacy is about hiding a wrong"'. [8] Rather, privacy encompasses fundamental values of autonomy, dignity, and self-determination. When employees have no meaningful say in how their work contributions might be used to train systems that could eventually replace them, their autonomy as rational agents is compromised through this secondary use of their work.

D. Ownership and Appropriation of Labor

The training of GenAI systems on employee-generated data raises critical questions about ownership and the appropriation of labor. Goetze articulates this through a Lockean perspective, arguing that "Because artists own the labour that produces their works by transforming creative building-blocks, they thereby own those works." [7] While employment contracts typically transfer certain rights to these products to employers, the specific use of these outputs to train replacement AI systems represents a novel appropriation that extends beyond traditional employer rights and reasonable employee expectations.

This philosophical concern manifests in documented workplace cases where employees unwittingly "train their replacements." At Disney, employees were required to train foreign workers who would take over their jobs before being laid off [10]. At Google-owned Looker, workers unknowingly created documentation and processes that trained their eventual replacements before layoffs were announced [11]. Unlike traditional workplace knowledge transfer, where an employee might train a colleague who will continue similar work, GenAI systems represent a paradigm shift where employees' work products are systematically extracted to build technologies explicitly designed to eliminate their positions entirely.

E. Contextual Integrity in the Workplace

Barocas and Nissenbaum's theory of contextual integrity provides a powerful framework for understanding the ethics of data flows in workplace environments. This theory holds that privacy violations occur when information flows outside of contextually appropriate norms. They explain: "informational norms prescribe information flows according to key actors, types of information, and constraints under which flow occurs." [4] In workplace contexts, employees reasonably expect their outputs to be used by their employer for business purposes within established professional norms—not aggregated and used to train systems designed to automate their jobs.

The systematic collection of employee data for AI training often involves forms of workplace surveillance that Solove categorizes across multiple dimensions in his taxonomy of privacy (see Figure 2 below): information collection, information processing, information dissemination, and invasion. [8] When organizations continuously monitor employee outputs to gather training data (surveillance), aggregate this data across departments or functions, analyze it to identify patterns that can be automated, and ultimately use these insights to make employment decisions, they engage in privacy violations at each stage of this process. These violations are particularly troubling when employees have no awareness of or consent to this surveillance, and when the ultimate purpose directly threatens their continued employment.

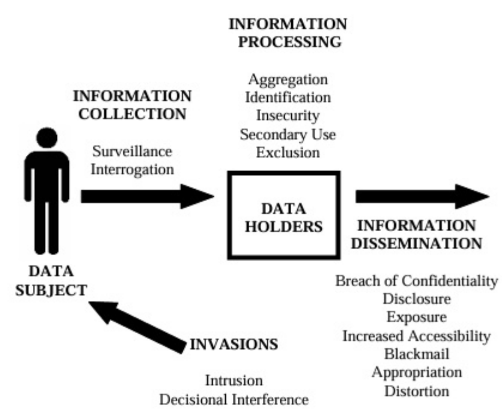

Fig. 2. Taxonomy Of Privacy [12]

The violation of contextual integrity is especially troubling when the new information flow directly threatens the livelihood and professional identity of the data subject. This crosses a line from mere privacy violation to a form of exploitation that treats employees' intellectual contributions as raw material for systems that could do them harm.

But what about situations in which employees are given a choice as to how their labor data are used? This, on its face, would seemingly address at least some of the presented concerns. However, even if only some workers consent to having their work used for AI training, the resulting models can still effectively capture and replicate the skills and knowledge of the entire workforce. This phenomenon is identified as "the tyranny of the minority" by Barocas and Nissenbaum, where "the volunteered information of the few can unlock the same information about the many." [4]

This creates a problematic dynamic where individual refusal to participate becomes meaningless once a critical threshold of data has been acquired. As the authors go on to explain, "once a critical threshold has been reached, data collectors can rely on more easily observable information to situate all individuals according to these patterns, rendering irrelevant whether or not those individuals have consented to allowing access to the critical information in question." For employees facing potential displacement, a lack of individual consent means that they may have no meaningful way to opt out of contributing to the systems that threaten their jobs. This form of systemic exclusion from decision-making represents a significant ethical concern that extends beyond traditional privacy frameworks that come from freely given collective consent.

The data ethics issues surrounding GenAI development in workplace settings encompass fundamental questions about consent, autonomy, ownership, and justice that directly impact the lives and livelihoods of workers whose data may be used to train their eventual replacements. As Barocas and Nissenbaum conclude, "consent cannot exist as an excuse for anything," [4] particularly when that consent is neither informed nor freely given. Organizations developing and deploying GenAI systems have an ethical obligation to consider the impact of their data collection practices on the workers whose contributions make these systems possible.

Addressing these data ethics challenges requires moving beyond individual consent models to develop structural approaches that recognize and mitigate power imbalances. These ethical concerns connect directly to broader questions of power and justice, which we examine in the next section. While data ethics helps us understand the fundamental violations occurring when employee data is harvested without consent, a power and justice analysis allows us to examine how these practices reflect and potentially exacerbate existing inequalities in the workplace and society more broadly, particularly affecting already vulnerable worker populations.

III. Power and Justice Analysis

The systemic collection and use of employee data for training GenAI systems that may lead to job displacement reflects and reinforces existing power asymmetries in the workplace. Examining these technologies through a lens of power and justice requires going beyond narrow concepts such as bias mitigation to consider wider implications, including the fair distribution of benefits and burdens, acknowledging existing inequalities, and addressing the need to redistribute power. As Whittaker observes, large tech firms and other corporations wield significant influence through their control of AI resources, creating a "web of conflicted relationships that threaten academic freedom and our ability to understand and regulate these corporate technologies." [6] This power imbalance extends to employer-employee relationships, where workers have limited ability to challenge or even understand how their intellectual contributions are being repurposed.

Justice in the context of GenAI encompasses more than just ensuring algorithmic fairness; it requires examining how these technologies produce or exacerbate unjustified power hierarchies and how the benefits and harms are distributed across society. This analysis is particularly important because AI impacts often disproportionately affect vulnerable individuals and groups who already face structural disadvantage. As Lin argues, the creation of AI image generators using harvested training data is "fundamentally unjust, for it makes those who are already less powerful worse off than they were before." [13]

A. Shifting Power through Data Harvesting and Control

The emergence of GenAI systems has introduced a concerning mechanism for power concentration through the systemic harvesting of employee-generated data. Organizations routinely collect work products created by employees to develop AI systems capable of automating those same tasks. The methods by which this collection is done can be straightforward (requiring employees to document their work processes), or deceptive (recording employees without their knowledge, or using recordings for an undisclosed purpose). This process fundamentally alters the control over the asset of labor products, shifting power away from individual employees who created these works and toward organizational entities that aggregate and exploit the data.

Drawing on Locke's labor theory of property as applied by Goetze to the creative labor context, we can understand this dynamic as a form of labor appropriation. The implications of this data harvesting and appropriation extend beyond individual property rights to fundamentally reshape power dynamics in the workplace. This occurs through three interconnected mechanisms: undermining collective labor power, fragmenting work processes in ways that inhibit solidarity, and creating information asymmetries that disadvantage workers. Each of these mechanisms represents a distinct dimension of how GenAI systems shift power from labor to capital.

1) Undermining Collective Labor Power: This systematic collection of employee data for AI training threatens not only individual employee power but also weakens collective labor power. Union membership has been on the decline for decades (See Figure 3) thanks to "right to work" laws, and this will only worsen as workers discover their outputs have been used to develop systems that might render their skills redundant. This weakening extends beyond individual impacts to affect workers' collective capacity to negotiate favorable terms. Workers' voices "can more readily adapt workplace rules to local conditions than state regulations can, while incorporating respect for workers' freedom, interests, and dignity." [38]

Historically, collective labor power has relied on the limited replaceability of specialized human skills and knowledge. GenAI fundamentally alters this dynamic by potentially enabling widespread automation of knowledge work previously considered resistant to technological displacement. This parallels earlier waves of automation that diminished the bargaining power of manufacturing workers, but extends the pattern into professional, creative, and administrative domains previously considered safe from such disruption.

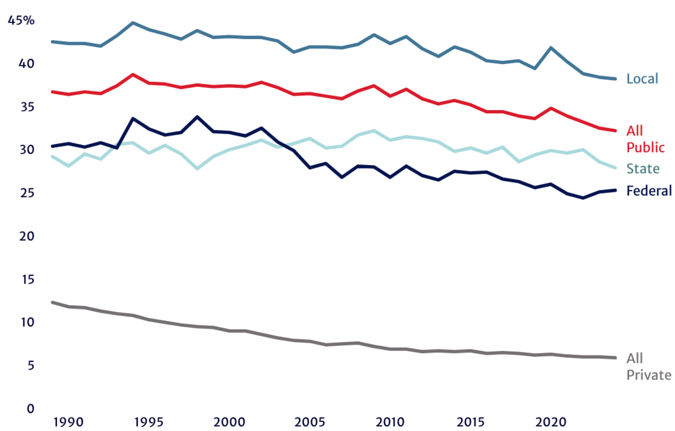

Fig. 3. Union Membership by Sector and Level of Government, Percent of Employees, 1989–2024 [33]

2) Fragmenting Work Processes and Solidarity: The fragmentation of work processes facilitated by GenAI threatens worker solidarity by creating uneven impacts across different groups, particularly when workers are spread thinly across many individual firms (i.e., graphic designers). Labor market polarization resulting from automation technologies tends to create stark divisions between workers who benefit from technological change and those who are displaced by it. This stratification occurs not only between industries but within them, as certain roles become enhanced by AI while others face obsolescence. Such bifurcation makes it increasingly difficult for workers to organize across skill levels and job functions, as their immediate interests and concerns may diverge significantly. When workers cannot identify common cause due to these artificially created divisions, their collective ability to resist exploitative practices is fundamentally undermined, further shifting the balance of power toward employers.

3) Creating Information Asymmetries: Moreover, the opaque nature of how employee data are collected and used to train these systems creates significant information asymmetries between employers and workers. These asymmetries effectively limit workers' ability to anticipate, understand, and respond to technological changes that directly impact their livelihoods. By controlling both the development timeline and implementation strategy of AI systems, employers can strategically bypass traditional collective bargaining processes that might otherwise provide a check on unilateral technological changes. This systematic exclusion of workers from meaningful involvement in AI-related decisions represents a profound shift in workplace power dynamics, where even the most well-organized labor groups may find themselves unable to effectively negotiate over technology that has been developed using their own intellectual contributions.

These three mechanisms; undermining collective labor power, fragmenting work processes, and creating information asymmetries, operate in concert to shift power from workers to employers. This power shift stems directly from the harvesting and control of employee-generated data, showing how GenAI doesn't merely automate tasks but fundamentally restructures workplace power relations. Far from being a neutral technological development, the deployment of GenAI that relies on employee data represents a significant redistribution of power within organizations.

B. Participation, Autonomy, and Agency

The current standard of development and deployment of GenAI systems is primarily driven by technical and management teams and largely excludes the workers who will be most affected by these technologies. This exclusion represents another dimension of power imbalance, as those with the most to lose have the least say in how these systems are designed and implemented.

A growing interest in "participatory AI" seeks to involve wider public and affected groups in the AI development lifecycle. When done authentically, participatory approaches can empower communities and confront power asymmetries by ensuring that those most affected by AI systems have a voice in their development. [14] In the workplace domain, this might involve including workers in decisions about what tasks to automate, how to design AI systems that complement rather than replace human work, and how to distribute the benefits of increased productivity.

However, not all participatory practices are created equal. There is a risk that participation becomes merely superficial engagement rather than genuine empowerment. There is a danger of "participation-washing," where efforts are characterized under the banner of participation but fail to give participants any real decision-making power. [14] To be meaningful, participation must involve a substantive redistribution of power that allows workers to shape the technologies that will affect their livelihoods.

C. Long-Term Implications

The shift in power dynamics facilitated by GenAI systems is not merely technological but fundamentally political and ethical. By extracting value from employee labor to develop automation systems, organizations are embedding existing power imbalances into socio-technical systems in ways that threaten to exacerbate existing inequalities. The uneven distribution of benefits and harms from these systems raises serious justice concerns, especially when those already disadvantaged bear the greatest burdens while receiving the fewest benefits.

This analysis points to the ethical urgency of addressing these power imbalances and justice concerns. Technology companies have demonstrated a "clear willingness to silence and punish critics" [6] who raise ethical concerns about AI development practices, which has created a chilling effect that inhibits meaningful oversight and ethical scrutiny. Without intervention, GenAI systems are likely to contribute to systemic inequalities by reinforcing and amplifying existing patterns of advantage and disadvantage.

Addressing these concerns requires adherence to principles of fairness and equity for all stakeholders, with particular attention to protecting the interests of vulnerable groups. Zuboff warns that we are witnessing a pattern that follows what Arendt (following Marx) called capitalism's "original sin of simple robbery," which becomes repeated without limit in the quest for limitless growth. [15] To prevent this pattern from further entrenching injustice, we must consider not only the technical design of GenAI systems but also the power structures in which they are embedded and the distribution of benefits and harms they produce. The next section is an analysis of how these harms will be disproportionately felt by already vulnerable groups.

IV. Inequality Analysis

As GenAI becomes more universal and the influence it has on job displacement more severe, the impact it will have on inequality will be equally severe, leading to labor market polarization. Increased labor market polarization leads to increased income inequality, which has been a growing problem for decades, as shown in Figure 4. The extremes of labor market polarization is when there are effectively only two types of jobs: "High-paying jobs that require advanced technical skills on one end, and low-paying not-automatable jobs on the other, leaving a shrinking middle class." [16] High-income earners will take an increasingly greater percentage of overall income, leading to even more inequality.

Income inequality has far-reaching effects. "Almost all principles of fair decision-making derive their moral force from the principle of equality: the idea that all persons have equal moral worth and therefore should be treated equally." [18] The jobs left for low-income earners, that is, manual and service jobs that *must* be done, but cannot be automated away, will be treated with more and more disdain.

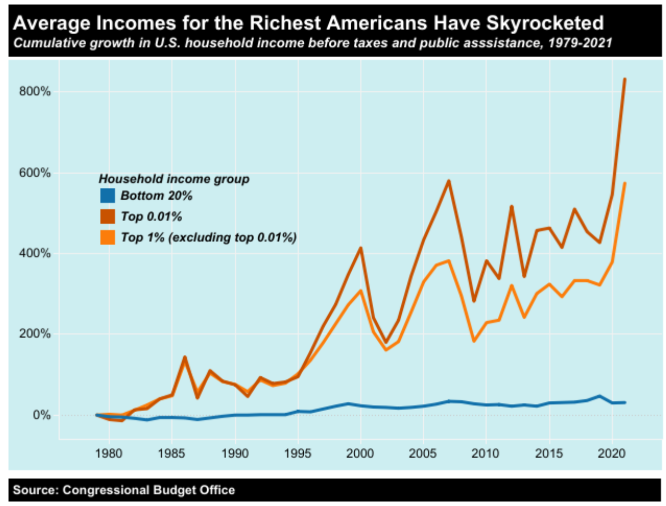

Fig. 4. Average Household Incomes from 1980-2020 [17]

A. The Inequality of Fairness

When discussions arise about fairness in hiring, there is almost always agreement that the "most qualified person" should get the job. This is not really what it means to have "fair" hiring practices, though. Fair hiring practices are about ensuring that what is being evaluated as part of the filtering process is truly relevant to the job. For example, a candidate's skin color or perceived gender is almost certainly NOT relevant to most (if any) jobs. However, having the skills and contacts of an Ivy League college graduate might be.

This brings us to college admissions, particularly at "elite" schools. Affirmative action policies for college admissions have been much maligned as "reverse racism." Whether that is true is not as important as whether these policies were effective. One measure of effectiveness is graduation rates. For example, before the passage of Proposition 209 in 1996, which prohibited the use of racial preferences by public universities in California, studies showed that the graduation rate of blacks at UCLA was 41%, compared to 73% for whites. [19] This could lead one to believe that admitting underrepresented minorities is a waste of a space that could have gone to someone else (read: NOT a minority) better prepared to meet the rigors of that institution.

This tells us that we must go even further back to find the "bottleneck," as Joseph Fishkin put it, that is causing this inequality. [18] How does a potential student become prepared to meet the rigors of elite institutions? Often, the answer lies in the student's home zip code and their parents' bank account. That is, high-income parents are able to help their children attend better (as in "better able to prepare them for what's next") primary and secondary schools, provide tutoring and test prep services to boost their GPAs and SAT scores, and through their own experience and connections, help make the college applications stand out.

"One of the most important principles of equal opportunity has been to equalize developmental opportunities by providing all children (and later, all people) with equal access to education, nutrition, and other building blocks of development. This approach is sometimes called equalizing the starting gate in the race of life, ensuring that all the competitors show up with equal opportunity to prepare with the hopes that decision-making on the basis of merit after that point will be fair as a result." [18] Unfortunately, these inequalities, even if only minor in childhood, compound over time and lead to major discrepancies as one enters adulthood. This leads those who were not born with advantage even less likely to be able to rise above their station as there will be even fewer opportunities to do so.

B. Disproportionate Outcomes

This is the vicious cycle of income inequality that will be exacerbated by AI-enabled labor market polarization and disproportionate outcomes. This polarization contrasts sharply with the potential benefits of AI, which are often framed in terms of "amplifying" human skills and enabling more complex cognitive tasks. However, these benefits are primarily gained by already privileged, higher-skilled, and higher-paid workers [20]. A joint study by the UN labour agency (ILO) and Poland's National Research Institute "finds that in high-income countries, jobs considered at the highest risk of AI-driven task automation account for 9.6 per cent of female employment – nearly three times the share for men. Worldwide, 4.7 per cent of women's jobs fall into the highest-risk category, compared with 2.4 per cent for men. This disparity is due largely to the over-representation of women in clerical and administrative roles, which are among the most exposed occupational groups. These jobs often involve tasks such as data entry and document formatting and scheduling, functions that AI technologies can already perform efficiently." [34]

The implications of this disparity are profound. Drawing on Rawlsian principles of justice, we can argue that inequalities are only justified if they benefit the least advantaged members of society [8]. However, the current trajectory of GenAI implementation appears to be moving in the opposite direction, concentrating benefits among those who already enjoy relative privilege while shifting burdens to more vulnerable populations. This is particularly true in more creative, highly competitive, and (generally) minimally compensated fields, where those who already have the means to meet their basic needs are better able to weather this shift.

Furthermore, biased data used to train these systems can replicate or exacerbate existing systemic inequality. If the system of data collection itself is biased, then the outputs of the resulting models will carry forward those potentially harmful biases [13]. These biases produce not only harmful outcomes for affected groups but also negatively impact workers who deploy these systems. Workers who are implicated in implementing biased AI systems often experience diminished perception of the significance of their work, particularly when outcomes conflict with their personal values. This dynamic affects what Lips-Wiersma and Morris identify as "developing and becoming self" – one of four key sources of meaningful work. [21] When workers cannot align their actions with their moral values due to the constraints of biased AI systems, a fundamental dimension of workplace meaning is compromised, creating another form of disproportionate harm that extends beyond economic impacts.

C. Diminishing Worker Value

Using employee data to train GenAI systems represents a direct threat to the power of labor by diminishing the perceived value of the employee skills and work output by suggesting that their contributions can be readily automated. Despite the fact that labor productivity has increased at roughly double the pace of wages (See Figure 5 below), this occurs even before any actual job displacement takes place, simply through the recognition that one's work is being used to develop systems that could eventually replace human labor.

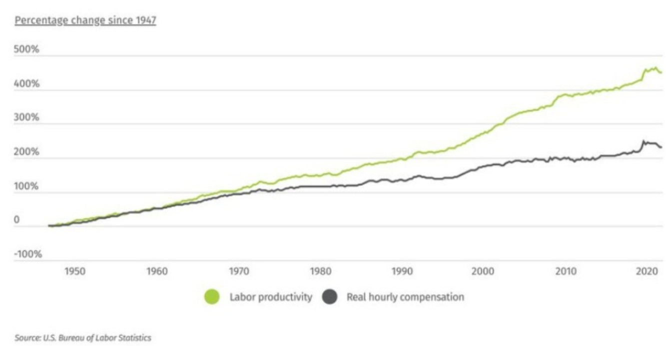

Fig. 5. Changes in Productivity and Real Income since 1970 [22]

This economic devaluation of worker contributions creates a deeper, psychological impact that transcends financial concerns. When an employee's skills are deemed replaceable by AI, what's threatened is not merely their economic value in the labor market, but their entire professional identity, dignity, and sense of meaning. The process creates a form of alienation from one's own labor, as work that once represented specialized human knowledge becomes reframed as merely trainable data for AI systems. This transformation touches on fundamental questions of professional identity, dignity, and respect. Artist Karla Ortiz, as quoted by Goetze, articulates "We're taking our consent back... [AI image generators] have data that doesn't belong to them... That data is my artwork, that's my life. It feels like my identity." [7] This sentiment likely resonates with many workers across industries whose outputs are being used to train GenAI systems. When workers discover their intellectual contributions have been repurposed to develop technologies that could render their skills redundant, the harm becomes a threat to their sense of professional self. Resistance to these technologies is not merely about job security, but about maintaining autonomy over one's creative and intellectual labor.

The fundamental inequality in this dynamic is further captured by Goetze, who notes that in the context of AI image generators, "artists have had their work appropriated *en masse* by more powerful technologists for the creation of generative AI tools that pose an existential threat to the artists' livelihoods." [7] When applied to broader workplace contexts, this same dynamic appears: employee contributions and labor are extracted and transformed into systems designed to render those same employees redundant, with economic benefits flowing primarily to those who already hold power. This contributes to greater labor market polarization: High-paying technical jobs coexist with low-paying non-automatable roles, shrinking the middle class and increasing income inequality. This leads to the next section analyzing accountability.

V. Accountability Analysis

Accountability in the context of artificial intelligence refers to the capacity to determine who is responsible for the decisions, actions, and consequences of AI systems. For GenAI systems that may lead to job displacement, accountability takes on heightened significance. There is an inverse relationship between "achievement gaps" created by automation and "responsibility gaps" that arise from the same technological developments. [23] When employees unwittingly contribute to systems that may replace them, identifying who bears responsibility for the subsequent displacement becomes technically complex and ethically fraught.

The complexity of GenAI systems creates significant challenges for establishing meaningful accountability frameworks. They are often proprietary in nature, they are technically opaque, and there is almost always distributed responsibility for their development and deployment. Addressing these accountability challenges requires examining current principles and frameworks, identifying existing gaps, and establishing robust mechanisms for holding these systems and their creators accountable.

A. Accountability Frameworks

Several accountability frameworks have emerged in response to the proliferation of AI systems across various domains. These frameworks attempt to establish principles and practices for ensuring that AI systems, including GenAI, are developed and deployed responsibly and that clear lines of accountability exist for their impacts.

The AI4People ethical AI framework outlines five principles: beneficence, non-maleficence, autonomy, justice, and explicability. [20] The explicability principle is particularly relevant to GenAI accountability, as it demands that AI systems operate in ways that are both intelligible and accountable to stakeholders. In the context of job displacement, explicability requires that organizations can meaningfully explain not just how their GenAI systems function technically, but also how decisions about implementation are made and what impacts these systems will have on workers.

For instance, when a GenAI system trained on employee-generated content is deployed to automate creative or knowledge work, explicability demands that organizations can explain: (1) what employee data was used in training, (2) how the system makes its decisions or generates outputs, (3) what oversight exists to ensure quality and fairness, and (4) how the implementation will affect existing roles and responsibilities. Without this level of explicability, affected workers cannot meaningfully contest decisions or participate in shaping implementation strategies, creating a fundamental accountability deficit.

Industry-specific frameworks have also emerged to address accountability concerns in employment contexts. The IEEE's Ethically Aligned Design guidelines specifically address employer obligations when implementing automation technologies, emphasizing the need for transparency about how employee data are used and clear communication about potential impacts on job roles [24]. The National Institute of Standards and Technology (NIST) AI Risk Management Framework provides a structured approach to assessing and managing risks associated with AI systems, including those related to workforce impacts (see Figure 6 below). The framework emphasizes the importance of identifying and mitigating potential harms throughout the AI lifecycle, from conception and design through deployment and monitoring [25]. This ties directly back to the privacy concern regarding employees being unaware, and uninformed, of a secondary use of their work.

Fig. 6. Three Pronged Approach [26]

Despite these emerging frameworks, significant gaps remain between technical explanations of GenAI systems and meaningful accountability for affected workers. A mere justification for an algorithm's decision itself does not address questions of legitimacy and authority. [27] Technical transparency about how an algorithm functions is necessary but insufficient for establishing accountability for its broader societal impacts, particularly job displacement.

B. Accountability Gaps

Responsibility diffusion represents a fundamental accountability challenge wherein responsibility becomes distributed among multiple actors, including data scientists, management, HR departments, and external vendors. Creel and Hellman identify this as a type of "algorithmic leviathan" problem, where there is a systemic diffusion of responsibility, making it difficult to identify who bears responsibility when AI systems cause harm [28]. In the context of GenAI systems trained on employee data, this diffusion can make it difficult to identify who bears responsibility when these systems lead to job displacement. Data scientists may claim they were simply building tools requested by management; management may argue they relied on vendors' assurances; vendors may indicate they merely provided systems based on client specifications. This tension between institutional accountability and individual responsibility is defined by Lin as a disconnect between "micro-ethics", individual professional conduct, and "macro-ethics," collective responsibility for systemic impacts. [13]

Temporal accountability gaps emerge due the significant time lag between when employee data is collected and used for training GenAI systems and when displacement actually occurs. This disconnect makes it difficult to establish clear causal links between specific decisions about data collection and training and subsequent job impacts, further obscuring accountability.

Expertise asymmetry presents another significant barrier to meaningful accountability. The complexity of GenAI systems creates significant barriers to oversight by non-technical stakeholders, including affected workers, labor representatives, and even corporate leadership. When organizations cannot fully explain how their GenAI systems function, they cannot reasonably be held accountable for specific outcomes, creating "a double standard of transparency" [27].

Real-world implementations of GenAI have already demonstrated significant accountability gaps. Consider the case of content creation platforms that have incorporated AI-assisted writing tools trained on vast corpora of professional writers' work. When publishers like CNET deployed AI-generated content tools in 2023, they initially failed to disclose which articles were AI-generated, creating confusion about authorship and responsibility [29]. Writers whose previous work likely contributed to the system's training dataset found their job security threatened by a technology they had unwittingly helped create.

Similarly, in software development, GitHub's Copilot, trained on publicly available code repositories, has raised accountability questions about intellectual property rights and the economic impact on developers whose code formed the training data [30]. These examples reveal how the lack of transparent accountability frameworks can create environments where workers contribute to systems that may replace them, without clear channels for recourse or compensation.

C. Mechanisms for Accountability

Addressing the accountability gaps identified above requires developing robust mechanisms that can establish clear lines of responsibility for GenAI systems and their workforce impacts. These mechanisms must balance technical feasibility with ethical imperatives and labor rights.

Transparency requirements represent a foundational mechanism for establishing accountability. In practice, this requires mandatory disclosure of training data sources, development processes, and potential job impacts. Organizations implementing GenAI systems in a workplace should be required to document and disclose the sources of data used to train those systems, the intended applications of those systems, and the expected impact on existing roles and job functions. These requirements align with a "structural approach toward achieving equal opportunity" in algorithmic systems. [18]

Worker participation in AI governance and deployment, as touched upon in a previous section, represents another crucial piece of accountability in this domain. Meaningful involvement by workers in the workplace can lead to better outcomes by incorporating the perspectives of those most directly affected by these technologies. [14]

Legal and regulatory controls must evolve to address the novel challenges posed by AI-driven job displacement. Employment law has historically adapted to technological change, and new legal frameworks are needed to establish accountability for GenAI's impact on workforces across industries. These should include notification requirements when GenAI systems are trained on employee data, labor laws that address automated decision-making in personnel decisions, legal remedies for workers displaced by systems trained on their intellectual contributions, and regulatory oversight of claims made about GenAI capabilities and impacts.

D. Robust Accountability

Establishing robust accountability for GenAI systems represents a crucial component of addressing the broader ethical issues examined throughout this paper. When workers unwittingly contribute to systems that may eventually replace them, clear lines of responsibility and mechanisms for recourse become essential ethical safeguards. The accountability frameworks and mechanisms proposed here directly address the concerns raised in our analyses of data ethics, power and justice, and inequality.

Worker solidarity, particularly through union representation, will be critical for accountability. By requiring organizations to be transparent about how employee data is used in training GenAI systems, establishing meaningful worker participation in AI governance, and developing appropriate legal and regulatory frameworks, we can begin to address the fundamental ethical tension at the heart of this paper: the use of employee-generated data to create systems that may displace those same employees.

While these accountability measures cannot fully eliminate the disruption that GenAI will bring to labor markets, they can ensure that the costs of this transition are not borne disproportionately by those who contributed—knowingly or unknowingly—to making these technological advances possible.

VI. Conclusion

As the pace of technological development, and AI in particular, accelerates, it can be overwhelming to consider all of these issues and implications. Let us conclude by offering a perspective specific to each of the three groups in the target audience: AI systems developers, corporate leaders, and government policy-makers.

Government policy-makers have an ethical responsibility to the people, who are overwhelmingly the workers impacted by AI-related job displacement. Government policy should be focused on ensuring that workers rights and privacy are protected, union and collective bargaining agreements can be made and are upheld, early "bottlenecks" leading to future inequalities are cleared, and appropriate regulations are enacted and enforced. These actions not only protect workers and society, but create a predictable and level playing field upon which the best ideas can compete in a free and fair market.

Corporate leaders should understand that they don't lead companies, they lead people. Using AI to make your company more "efficient" in the short term to by reducing headcount to increase profit margins will likely lead it to being less adaptable and creative in the long term. "Growth for the sake of growth is the ideology of the cancer cell" [31] and cancer cells either get forcibly removed, or they kill their hosts. Generative AI can be a useful tool to help employees thrive by simplifying mundane tasks and allowing them to focus on what matters most to you and your customers.

Though it may seem counterintuitive, **AI systems developers** can have the most direct impact on the ethics of AI systems in the workplace. This is because they are the ones with the power to decide what the AI system *truly* is. Does it hurt humans, or helps humans thrive? Any AI solution has the power to be either and it is up to the developer to decide which, because none of the decisions of policy-makers and corporate leaders matter without AI systems developers who make them a realty.

The question is: What should that reality to be?

Acknowledgments

The authors would like to thank Dr. Alicia Patterson and all the other students in their PHL 546 class at Oregon State. MJ would like to thank Dr. Houssam Abbas. JK would like to thank Dr. Alan Fern.

References

[1] Walsh, Matt, "ChatGPT Statistics — The Key Facts and Figures," Style Factory, https://www.stylefactoryproductions.com/blog/chatgpt-statistics, Updated on 4/29/2025, Accessed on 5/2/2025.
[2] Chayka, Kyle, "Rethinking the Luddites in the Age of A.I.," The New Yorker, https://www.newyorker.com/books/page-turner/rethinking-the-luddites-in-the-age-of-ai, Published on 9/26/2023, Accessed on 5/2/2025.
[3] Mayor, Tracy, "Ethics and automation: What to do when workers are displaced," MIT Sloan School of Management, https://mitsloan.mit.edu/ideas-made-to-matter/ethics-and-automation-what-to-do-when-workers-are-displaced, Published on 7/8/2019, Accessed on 5/2/2025.
[4] Barocas, S. and H. Nissenbaum, "Big Data's End Run around Anonymity and Consent," in Privacy, Big Data, and the Public Good: Frameworks for Engagement, J. Lane, V. Stodden, S. Bender, and H. Nissenbaum, Eds. Cambridge, UK: Cambridge University Press, 2014, pp. 44-75. https://dx.doi.org/10.1017/CBO9781107590205.004.
[5] Burleigh, E., "Amazon customer service workers are scared AI will replace them - and they're not alone," Fortune, https://fortune.com/2024/06/11/amazon-customer-service-agents-scared-ai-replacement, Published on 6/11/2024, Accessed on 5/23/2025.
[6] Whittaker, M., "The Steep Cost of Capture," Interactions, vol. 28, no. 6, pp. 51-55, Nov.-Dec. 2021. [Online]. Available: https://doi.org/10.1145/3488666.
[7] Goetze, T. S., "AI Art is Theft: Labour, Extraction, and Exploitation," in Proc. 2024 ACM Conf. Fairness, Accountability, and Transparency (FAccT '24), Rio de Janeiro, Brazil, 2024, pp. 186-196. https://doi.org/10.1145/3630106.3658898.
[8] Solove, D. J., "'I've Got Nothing to Hide' and Other Misunderstandings of Privacy," San Diego Law Review, vol. 44, no. 4, pp. 745-772, 2007.
[9] Wolff, J, "Fairness, Respect, and the Egalitarian Ethos," Philosophy & Public Affairs, vol. 27, no. 2, pp. 97-122, 1998.
[10] Preston, J., "Pink Slips at Disney. But First, Training Foreign Replacements," The New York Times, Jun. 3, 2015. [Online]. Available: https://www.nytimes.com/2015/06/04/us/last-task-after-layoff-at-disney-train-foreign-replacements.html.
[11] Field, H, "Before layoffs hit Google-owned Looker, workers unknowingly trained their replacements," Tech Brew, Mar. 21, 2022. [Online]. Available: https://www.emergingtechbrew.com/stories/2022/03/21/before-layoffs-hit-google-owned-looker-workers-unknowingly-trained-their-replacements.
[12] Solove, Daniel J., "A Taxonomy of Privacy." University of Pennsylvania Law Review, Vol. 154, No. 3, p. 477, January 2006, GWU Law School Public Law Research Paper No. 129, Available at SSRN: https://ssrn.com/abstract=667622.
[13] Lin, T., "'Democratizing AI' and the Concern of Algorithmic Injustice," Philosophy & Technology, vol. 35, no. 2, pp. 1-24, 2024. https://doi.org/10.1007/s13347-024-00792-2.
[14] Birhane, A., W. Isaac, V. Prabhakaran, M. Díaz, M. C. Elish, I. Gabriel, and S. Mohamed, "Power to the People? Opportunities and Challenges for Participatory AI," in Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22), Arlington, VA, USA: ACM, 2022, pp. 1-8. https://doi.org/10.1145/3551624.3555290.
[15] S. Zuboff, "The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power," as cited in T. S. Goetze, "AI Art is Theft: Labour, Extraction, and Exploitation," in ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), Rio de Janeiro, Brazil: ACM, 2024, pp. 186-196. https://doi.org/10.1145/3630106.3658898.

[16] Krimmelbein, Fred, "The Ethical Implications of AI and Job Displacement," SogetiLabs, https://labs.sogeti.com/the-ethical-implications-of-ai-and-job-displacement, Published on 10/3/2024, Accessed on 5/2/2025.
[17] Author Unknown, "Income Inequality," Inequality.org, https://inequality.org/facts/income-inequality/, Accessed on 5/2/2025.
[18] Jain, Shomik, Vinith Suriyakumar, Kathleen Creel, and Ashia Wilson. "Algorithmic pluralism: A structural approach to equal opportunity." In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, pp. 197-206, 2024.
[19] Sander, Richard H., "A systemic analysis of affirmative action in American law schools," Stanford Law Review, 57 (2): 367–483, JSTOR 40040209, November 2004.
[20] Bankins, S. and P. Formosa, "The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work," Journal of Business Ethics, vol. 185, pp. 725–740, 2023. https://doi.org/10.1007/s10551-023-05339-7.
[21] Lips-Wiersma, M. and L. Morris, "Discriminating between 'meaningful work' and the 'management of meaning'," Journal of Business Ethics, vol. 88, no. 3, pp. 491–511, 2009. https://doi.org/10.1007/s10551-009-0118-9.
[22] Heacock, David, "U.S. States With the Largest Increase in Labor Productivity Over the Last Decade," HowToHome, https://www.howtohome.com/us-states-with-largest-increase-in-labor-productivity/, Published on 12/6/2022, Accessed on 5/16/2025.
[23] J. Danaher and S. Nyholm, "Automation, work and the achievement gap," AI and Ethics, vol. 3, no. 2, pp. 375-381, 2022. https://doi.org/10.1007/s43681-020-00028-x.
[24] IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems," IEEE Standards Association, Piscataway, NJ, USA, 2019. [Online]. Available: https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html.
[25] National Institute of Standards and Technology, "AI Risk Management Framework 1.0," NIST, U.S. Department of Commerce, Gaithersburg, MD, USA, Jan. 2023. [Online]. Available: https://doi.org/10.6028/NIST.AI.100-1.
[26] Holistic AI Team, "NIST Launches AI Risk Management Framework 1.0," Holistic AI, https://www.holisticai.com/news/nist-launches-ai-risk-management-framework-1-0, Published on 1/26/2023, Accessed on 5/18/2025.
[27] M. Günther and A. Kasirzadeh, "Algorithmic and human decision making: for a double standard of transparency," AI & SOCIETY, vol. 37, no. 1, pp. 375-381, 2022. https://doi.org/10.1007/s00146-021-01200-5.
[28] K. Creel and D. Hellman, "The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision-Making Systems," Canadian Journal of Philosophy, vol. 52, no. 1, pp. 26-43, 2022. https://doi.org/10.1017/can.2022.3.
[29] Nieva, R.,"CNET is reviewing its AI-assisted articles after disclosure and accuracy issues," The Verge, Jan. 25, 2023. [Online]. Available: https://www.theverge.com/2023/1/25/23571082/cnet-ai-generated-articles-errors-corrections-red-ventures.
[30] Vincent, J., "GitHub's AI-powered Copilot will face its first major copyright lawsuit," The Verge, Nov. 11, 2022. [Online]. Available: https://www.theverge.com/2022/11/11/23452761/github-copilot-class-action-lawsuit-ai-copyright-violation.
[31] Abbey, Edward, "The Journey Home: Some Words in Defense of the American West," Published January 30, 1991 by Penguin Publishing Group, ISBN: 9780452265622.
[32] Spiekermann, K., A. Slavny, D. V. Axelsen, and H. Lawford-Smith, "Big Data Justice: A Case for Regulating the Global Information Commons," The Journal of Politics, vol. 83, no. 2, pp. 577-588, 2021. https://doi.org/10.1086/709862.
[33] Brown, Hayley Brown and Emma Curchin, "Union Density Continues to Decline," Center for Economic and Policy Research, https://cepr.net/publications/union-density-continues-to-decline, Published on 1/28/2025, Accessed on 5/23/2025.
[34] United Nations, "AI threatens one in four jobs – but transformation, not replacement, is the real risk," United Nations, https://news.un.org/en/story/2025/05/1163486, Published on 5/20/2025, Accessed on 5/23/2025.
[35] "Duolingo will replace contract workers with AI," The Verge, https://www.theverge.com/news/657594/duolingo-ai-first-replace-contract-workers, Published on Apr. 28, 2025, Accessed on May 29, 2025.
[36] "IBM's AI hiring plans highlight automation trend," Axios, https://www.axios.com/2023/05/02/ibm-stock-artificial-intelligence-hiring, Published on May 2, 2023, Accessed on May 29, 2025.
[37] Davenport, T., Khrabrov, A., and Ross, P., "How to Train Generative AI Using Your Company's Data," Harvard Business Review, https://hbr.org/2023/07/how-to-train-generative-ai-using-your-companys-data, Published on Jul. 6, 2023, Accessed on May 29, 2025.
[38] Anderson, Elizabeth, "Private Government: How Employers Rule Our Lives (and Why We Don't Talk about It)," Princeton University Press, 2017, page 69.