International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

Impact of Unauthorized Data Practices on U.S. Social Inequalities

Minh Ngan Pham[1], Beverly Grace Clapano Oblina[2]
[1]Student, Vinschool Central Park, Ho Chi Minh City, Vietnam, 700000
[2]AP Seminar, Academic Writing, & ESL Teacher, ESL Standard Department, Vinschool Central Park, Ho Chi Minh City, Vietnam, 700000
Email address: beverlygrace90210@gmail.com

Abstract—*Unauthorized data collection and algorithmic decision-making in the United States have intensified systemic discrimination, exacerbating social inequalities for marginalized communities. This paper examines how biased data practices—ranging from social disparities, healthcare to law enforcement—disproportionately harm vulnerable groups. Key findings reveal that predictive algorithms trained on historically skewed data replicate and amplify societal prejudices. For instance, healthcare algorithms misjudge Black patients' health risks due to cost-based metrics linked to systemic barriers, while law enforcement tools falsely label Black individuals as higher-risk, perpetuating harsher judicial outcomes. Additionally, corporations commodify personal data, prioritizing profit over equity, as seen in Google's Project Nightingale, which exploited health records without consent. Policymaking gaps, such as the absence of federal privacy laws in the U.S., further enable exploitative practices. The analysis underscores the dual responsibility of developers and regulators: developers must integrate equity-centered design and diverse perspectives to mitigate biases, while policymakers need robust, inclusive regulations to curb unauthorized data use. The study concludes that addressing these issues requires transparency, community-driven solutions, and systemic reforms to ensure data technologies advance social justice and equity rather than deepen existing disparities.*

Keywords— *Unauthorized data practices; social inequalities; algorithmic bias; systemic discrimination.*

I. Introduction

In late 2019, a whistleblower revealed that Google's machine-learning research programme, Project Nightingale, had secret access to health records of 50 million Americans without their consent (Ebeling, 2021). The data breach raised major concerns about data privacy and exploitation, bringing attention to a broader issue in our data-driven society: the unauthorized collection, sharing and usage of data in the US. While data-driven technologies promise efficiency, they often perpetuate systemic biases, worsening social inequalities. While policymakers deal with regulations, developers also face challenges and responsibilities. Although some argue sensitive and personal data can help mitigate these dangers, reality shows that these systems often harm vulnerable groups, highlighting the need to re-evaluate the situation and propose appropriate solutions.

II. Social Disparities and Impact on Marginalized Groups

Over the course of time, predictive algorithms and autonomous technologies have perpetuated discrimination and social inequalities in many ways. Algorithms and autonomous technologies can be understood as sets of rules or instructions that computers follow to process data. Some papers claimed that these technologies escalate the exploitation of data from vulnerable individuals. For instance, people who are less capable economically, such as elderly gamblers, are targeted with high-rate instant credit advertisements or urged to take out risky loans (Favaretto et al., 2019). This explains how some data-driven services intended to assist the poor may inadvertently push them further into poverty. On top of that, it creates misconceptions and wrong assumptions that further contribute to inequalities, such as images of higher positions like CEOs are often pictures of males and less of females, or Black-sounding names trigger ads for jail records more (Obermeyer et al., 2019).

In healthcare, where data can determine life-and-death situations, disparities also happen. Obermeyer et al. (2019) found that Black patients assigned the same health risk score as White patients were actually sicker, with 26.3% more chronic illnesses at the 97th percentile. This stems from algorithms using medical expenses to measure health risks, which is a flawed metric as Black patients historically incur lower costs due to systemic barriers like unequal access to care, socioeconomic factors, discrimination or less trust in the healthcare system. Another study shows that certain racial and ethnic groups, specifically, Black/African Americans and Latinxs are 5.6 and 4.3 times more likely to die from COVID-19, at much higher rates than non-Latinx Whites, raising concerns about fairness and equality in healthcare (Sabatello et al., 2020). This highlights how marginalized communities face greater health risks and that unauthorized data practices may overlook their needs.

Similarly, in law enforcement, ProPublica's online article found that Black citizens were assigned higher risk than White people even with the same level of crime and were nearly twice as likely to be misclassified as high-risk for future violence (Angwin et al., 2016). As a result, Black defendants face harsher outcomes due to flawed algorithmic assessments, along with the nature of this data being not publicly accessible, it is even harder to audit fairness. The impact of these practices extends to public policy. Garrett and Decoteau (2023) show how COVID-19 resources are distributed in favor of wealthier places while neglecting those who are most in need. These examples show that when algorithms are built on biased data, they actually reinforce inequalities rather than addressing them. Vulnerable groups are often treated as data subjects without control over their data, leading to unfairness where companies benefit while they face risks, also called datafied marginalization, further creating a cycle of data colonialism (Ruijer et al., 2022).

III. Technology Developers' Role in Algorithms and Biases

The role of technology developers is pivotal in shaping how data practices impact society. Clinical decision-making technologies often replicate inequalities because they learn from historically biased data. This is because collected data can be divided into demographic factors, such as race, socioeconomic status, or religion, which are connected to societal disadvantages (Challen & Danon, 2023). Developers designing these systems often unknowingly encode bias into algorithms. Algorithmic bias is intensified when external experts analyze data without understanding historical or local contexts, leading to decisions that unintentionally encode systemic prejudices into technologies. Machine-learning models, trained on historical data influenced by prejudices, replicate and amplify discriminatory patterns. Alongside, throughout the process of collecting and analyzing data, information is grouped into sets, labeled, or identified as patterns, which may omit and overlook crucial details, resulting in underrepresentation or overrepresentation of statistics. On the other hand, many are also looking for solutions. In particular, Favaretto et al. (2019) propose incorporating varied perspectives into the design for these algorithms so they can detect biases instead. Techniques like synthetic datasets and privacy-preserving methods also help examine algorithms without compromising users' privacy (Bekkum & Zuiderveen Borgesius, 2023).

IV. Data Commodification

Back to the incident mentioned above, data breaches can happen without the participator's consent as data collected from citizens might be used for many other unknown purposes. This is known as function creep—a progressive expansion of the usage of technologies beyond their initial purpose. Ebeling (2021) argues that patient data, often harvested under the pretense of medical innovation, becomes a corporate commodity. The difference in power dynamics in data commodification ensures that marginalized communities continue to suffer from the risks of data misuse, while corporations monopolize the profits, creating a modern form of digital exploitation. For example, Google's Project Nightingale transformed intimate health details into profit-generating assets, sidelining patient rights in favor of corporate gain. Additionally, Hoffmann (2019) criticizes the tendency to prioritize efficiency and profit over fairness, as algorithms maximize performance metrics without considering different social contexts. Hence, it may unintentionally harm marginalized groups, widening disparities between them and privileged individuals. This implies how big companies focus on profits over equity, making vulnerable groups continue to suffer.

V. Regulations and Policy Makers

Policymakers face both challenges of having to protect individual's privacy while guaranteeing data practices do not perpetuate discrimination. For example, the General Data Protection Regulation (GDPR) in Europe restricts the use of sensitive data, such as race or ethnicity, without explicit consent. While this can protect privacy, it also hinders the ability to examine bias in algorithms (Bekkum & Zuiderveen Borgesius, 2023). Moreover, the lack of federal data privacy laws in the U.S. leaves marginalized groups exposed. Unlike the EU's GDPR, which mandates transparency, U.S. corporations like Meta and Google operate in a regulatory vacuum—profiting from data practices that disproportionately harm low-income and minority communities. Corporations like Google and Facebook profit from data extracted without proper oversight, and these data are often used in ways that amplify inequities (Ebeling, 2021). Although some states have executed data privacy regulations, this patchwork approach lacks the consistency to actually address complex issues. Moreover, efforts to regulate data practices, such as Europe's proposed AI Act, have shown promise by encouraging the investigation of high-risk AI systems. Yet, many criticized that those efforts are not enough to protect vulnerable populations (Bekkum & Zuiderveen Borgesius, 2023). Thus, policymakers must balance these demands to propose more sufficient regulations that satisfy both privacy and equity. Without examination, discriminatory practices will still exist.

VI. Potential Solutions to Mitigate Discrimination

There are changes to be made within the industry, starting off with the process of decision-making. Obermeyer et al. (2019) suggest new labels need a deep understanding of the field as well as the process of identifying and extracting different data elements, due to careful choices for labels that allow us to use algorithmic predictions while reducing the risks. Some researchers argue that expanding access to sensitive data—such as detailed socioeconomic histories—could contextualize algorithmic decisions and reduce disparities. However, this approach risks further exploitation without stringent safeguards. Yet, this solution poses new risks because this perspective overlooks the moral aspect of unregulated data practices. For that reason, ensuring fairness and transparency in algorithmic design does not hinder progress but strengthens it by building trust. As Favaretto et al. (2019) emphasize, innovation cannot come at the cost of marginalizing vulnerable populations. Above all, governments should strengthen policies to regulate existing data practices, ensure that unauthorized or illegal practices hold the same consequences, and not just be strict in some places like the EU or Canada. Additionally, solutions for systemic biases in data-driven industries are to check judgmental assumptions in data analytics and develop algorithms that account for equity and diversity. As Ruijer et al. (2022) suggest, empowering marginalized communities to co-design data systems can disrupt the cycle of exploitation, changing them from passive data subjects to active representatives of equity. By including diverse stakeholders in the design process, we can identify potential biases early on, reducing the risk of harm (Favaretto

Minh Ngan Pham and Beverly Grace Clapano Oblina, "Impact of Unauthorized Data Practices on U.S. Social Inequalities," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 7, Issue 10, pp. 11-13, 2025.

et al., 2019).

VII. Conclusion

In summary, data analysis is a powerful tool that can help us understand and improve public services. However, unauthorized data practices in the US have reinforced systemic discrimination and harmed marginalized communities through the use of biased algorithms, data commodification and imbalanced, poor regulations. As data-driven technologies continue to have a fundamental position in society, it is necessary to ensure that the process is inclusive and advances social justice and equity. Addressing fairness and accountability is not only a technical hurdle but a moral duty. Without systemic improvements, data-driven technologies will continue to mirror—and magnify—the very biases they claim to resolve. The path forward demands transparency, fair design, and policies that prioritize human dignity over profit.

References

[1]. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[2]. Bekkum, M., & Zuiderveen Borgesius, F. (2023). Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception? Computer Law & Security Review, 48, 105770. https://doi.org/10.1016/j.clsr.2022.105770

[3]. Challen, R., & Danon, L. (2023). Clinical decision-making and algorithmic inequality. BMJ Quality & Safety, 32(9), 495–497. https://doi.org/10.1136/bmjqs-2022-015874

[4]. Ebeling, M. F. E. (2021). Mining the data oceans, profiting on the margins. Global Policy, 12(S6), 85–89. https://doi.org/10.1111/1758-5899.12887

[5]. Favaretto, M., De Clercq, E., & Elger, B. S. (2019). Big Data and discrimination: Perils, promises and solutions. Journal of Big Data, 6(12), 1–27. https://doi.org/10.1186/s40537-019-0177-4

[6]. Garrett, C. L., & Decoteau, C. L. (2023). Data citizenship: Quantifying structural racism in COVID-19 and beyond. Big Data & Society, 10(2). https://doi.org/10.1177/20539517231213821

[7]. Hoffmann, A. L. (2019). Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. Information, Communication & Society, 22(7), 900-915. https://doi.org/10.1080/1369118X.2019.1573912

[8]. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

[9]. Ruijer, E., Porumbescu, G., Porter, R. P., & Piotrowski, S. (2023). Social equity in the data era: A systematic literature review of data-driven public service research. Public Administration Review, 83(2), 316–332. https://doi.org/10.1111/puar.13585

[10]. Sabatello, M. (2020). Structural racism in the COVID-19 pandemic: Moving forward. American Journal of Bioethics, 20(7), 9–15. https://doi.org/10.1080/15265161.2020.1764138