

Report week 2 - Power & Inequality

Keywords: civil justice, power, inequality, AI takeover

1 Introduction

While currently one of the most common concerns focuses on the potential dangers of AI surpassing human intelligence, some researchers, such as Meredith Whittaker, argue that these risks are largely hypothetical [2]. Specifically, the notion of AI achieving consciousness just diverts attention from more immediate and pressing threats associated with the corporations that control these technologies. This leads us to the first issue that we identified in our readings, which is the consolidation of immense corporate power within the tech industry, particularly regarding the deployment and control of AI technologies. The concentration of corporate power within the tech industry, especially in the field of AI, has already started a controversial debate in both the popular press and the scientific community. This can have great ramifications, with the potential for a dystopian future where a handful of entities dictate not just technological innovation but also societal structures. The risk lies in unchecked decision-making by the tech giants, potentially infringing on privacy, manipulating information, and shaping public discourse to suit their interests. On one hand, some supporters argue that centralised power facilitates more efficient development and deployment of cutting-edge technologies, fostering economic growth and innovation. On the other hand, sceptics fear a future where monopolistic control suppresses the competition, limits user autonomy, and increases social inequalities. The controversy surrounding this issue depends on striking a balance between exploiting the advantages of AI for societal progress while mitigating the risks associated with concentrated corporate power.

The second issue, deeply rooted in historical misuse of data as delineated by Catherine D'Ignazio and Lauren Klein in **Feminist AI**[5], spotlights the contentious debate surrounding AI and racial bias. This debate is fueled by AI's propensity to perpetuate long-standing prejudices, a concern that resonates across various social and ethical spectrums. On one side of the debate are those who argue that AI, with its biased datasets, is inherently flawed in recognizing and respecting the diversity of human experiences. They point to the high rates of misidentification in facial recognition technologies¹ among people of color, particularly Black individuals, as evidence of AI's deep-seated racial bias. This group calls for stringent oversight and reformation of data collection practices to mitigate AI's discriminatory impacts. While acknowledging AI's limitations, with proper checks and balances, it could potentially offer more objective decisions than humans in areas like law enforcement and job recruitment. Yet, it is important to acknowledge its limitations and see it as an impairment tool with regards to inequality and wrongful biases.

2 Corporate Power Consolidation in AI

2.1 The Popular View

To further explore the multifaceted landscape of AI's societal impact, let's now delve into a popular view that draws from a range of reputable press sources.

¹Throughout the rest of the document, we will FRT when referring to facial recognition or facial recognition technologies

The article by Naomi Klein in **The Guardian**[11] raises important questions and concerns about the current trajectory of generative AI and tech companies. The author challenges the utopian claims made by tech CEOs about the transformative benefits of generative AI. She argues that such claims, such as ending poverty, curing diseases, and solving climate change, are “hallucinations” and divert attention from the actual problems. She points out that the wealthiest companies are monopolising human knowledge, using it to create proprietary products without permission or consent, which she argues should be considered illegal. Klein emphasises that the motivations behind AI development are rooted in profit and shareholder interests rather than altruistic goals.

Moreover, in one of her articles for **Medium**[1], Emily M. Bender identifies several shortcomings in the voluntary commitments made by seven tech companies (Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI) regarding “safe, secure, and trustworthy AI.” She highlights the absence of a commitment to detailed documentation of training data and models. This documentation is crucial for understanding and mitigating potential harms associated with AI systems. The lack of commitment to data and model documentation raises concerns about transparency, accountability, and the ability to address biases. The author also advocates for policymakers to seek input from experts in the field who can provide insights into the real harms of AI, emphasising the need for diverse perspectives, including scholars, community activists, and journalists, rather than relying solely on input from the companies.

The **MIT Technology Review** series on AI colonialism[6] highlights the parallels between the impact of artificial intelligence and colonial history. While acknowledging that the AI industry does not seek to capture physical land or engage in mass-scale slavery, it argues that it employs more insidious means to enrich the wealthy and powerful at the expense of the poor. The series explores various aspects of AI colonialism, examining case studies from different regions. For example, in Venezuela, AI data-labelling firms exploit cheap labour amid an economic crisis, creating a new form of labour exploitation. The overarching theme is that AI is impoverishing communities and countries that lack a say in its development, much like those already impoverished by historical colonial empires. A new concept of “decolonial AI” is introduced as a way to shift power dynamics from Silicon Valley back to the people.

2.2 Comparison against Scientific Literature

In this part we discuss the current scientific views on the issues raised in the popular press regarding corporate power consolidation in AI and examine how the perspective from the scientific literature aligns or differs from the media.

While the popular press raises valid concerns, the scientific literature provides a more comprehensive understanding of the issues. The scientific community not only validates concerns but also contributes valuable insights into the technical, ethical, and global dimensions of the challenges. The first paper by Michael Hilb[7] discusses the monopolisation of intelligence capabilities as a significant ethical concern. It notes that the ability to create intelligence is akin to acquiring a new form of power. Corporate boards are highlighted as responsible for ensuring that AI technologies are not misused, emphasising the ethical implications of concentrated intelligence capabilities. The author also discusses the ethical consideration of data ownership and the benefits derived from economic value generated by data. He acknowledges that data is a crucial asset and raises questions about who benefits from the economic value created by data.

Furthermore, the proposal presented in the next paper by Mihailis E. Diamantis [4] aligns with the broad discussions encountered in the popular press by addressing the evolving landscape of corporate decision-making in the age of automation. The author argues that current legal frameworks may struggle to address corporate liability when algorithms, rather than hu-

man employees, are responsible for certain actions. The extended mind thesis [4], drawn from contemporary philosophy and cognitive science, is introduced as a potential solution to legal challenges posed by increasing automation in corporations. The argument suggests that corporations using algorithms to fulfil employee roles should be treated as having the same mental states as those relying on human employees. This proposal, while maintaining a minimalist approach that requires minimal changes to existing law, addresses the potential gaps in corporate liability as algorithms play an increasingly prominent role in decision-making processes.

In examining the scientific literature, it becomes evident that the concerns raised in the media are not merely speculative but have a basis in ethical considerations related to the concentration of intelligence capabilities. The ethical considerations of data ownership and economic benefits in Hilb's paper parallel the legal argument that corporations should be held liable for decisions and actions facilitated by algorithms.

3 AI's Racial Bias

As such, the discussion is not just about differing views on the facts or predictions of AI's capabilities. The crux of the argument centers around the ethical responsibility of AI development and the societal implications of deploying AI systems without addressing their inherent biases. The controversy thus pivots on whether to restrain AI's integration in crucial sectors until it is deemed equitable or to progressively refine AI within these sectors, balancing risks and benefits. The discussion was sparked in numerous press articles.

3.1 The Popular View

The article by Kashmir Hill in **The New York Times**[9] tells the harrowing story of Porcha Woodruff, a 32-year-old pregnant nursing student, was wrongfully arrested for robbery and carjacking due to a false FRT match. The incident highlights concerns about the reliability of AI in law enforcement and judicial contexts. Despite her pregnancy and lack of evidence, Woodruff experienced distress and health issues due to the arrest. This case highlights the significant risks associated with AI technologies, especially FRT, such as false accusations and wrongful arrests, which disproportionately impact minorities. The same author reported 3 years earlier, for the same publication[8] the case of Nijeer Parks, a 33-year-old Black man from New Jersey, was wrongfully arrested due to a poor FRT match, raising concerns about the ethical use of AI in law enforcement. Parks was accused of shoplifting and assaulting a police officer, but was identified by the software despite being 30 miles away. He spent 10 days in jail and spent \$5,000 on legal fees.

The articles above also shed light on the broader implications of this technology. The Detroit Police Department, which averages 125 FRT searches a year, primarily on Black men, exemplifies the racial biases inherent in the technology. Experts and critics of the technology argue that such cases expose its weaknesses and the dangers posed to innocent people. The toll of such technology on individuals and communities, as demonstrated by Woodruff's traumatic experience[9], underscores the urgent need for more responsible and ethical implementation of AI technologies. Sony researchers, in an article from **Wired**[3], have highlighted a significant issue in the current methods of AI testing, which primarily focuses on a scale from light to dark skin tones, neglecting the diversity of skin hues. This oversight leads to biases in AI systems, affecting their accuracy and fairness, especially for individuals from East Asia, South Asia, Latin America, Middle East or African Americans.

Looking towards future scenarios, there is an expectation that AI will continue to be used in law enforcement with scenarios expected to include both the potential for improved crime-solving capabilities but also the risk of increased instances of misidentifying and bias. It is

no surprise that this approach for including FRT, could offer benefits in law enforcement, such as improving the efficiency of criminal investigations. Namely, the benefit lie on processing large volumes of data and potentially identifying suspects more. However, this case, along with others, raises critical risks associated with AI, particularly concerning racial biases and the reliability of the technology. The concerns raised by the cases above-mentioned are profound, questioning the reliability and ethical implications of AI in critical areas like justice.

The concerns raised by Parks’[8] case are multi-faceted. They include the reliability of AI in critical decision-making processes, ethical considerations in deploying such technologies without sufficient accuracy, and the potential for exacerbating racial inequalities in the criminal justice system. Additionally, the case challenges the defense often made by law enforcement that FRT is merely a clue and not the sole basis for arrests, as Parks’ experience contradicts this claim.

3.2 Comparison against Scientific Literature

In contrast, the paper **Understanding bias in facial recognition technologies** by David Leslie[12] delves into the critical issue of bias within FRT, while Pauline Kim’s paper **AI and Inequality**[10] takes a more nuanced and detailed exploration of how AI can exacerbate existing inequalities and create new forms of disparity.

The perspectives presented in popular media about biases in FRT are largely grounded in reality. The two papers corroborate these concerns through discussing various instances and studies which have demonstrated the presence of racial, gender, and age biases in these technologies. These biases are not speculative but are well-documented issues. For example, the paper by David Leslie[12] references studies showing that some FRT systems have higher error rates for women and people of color. It delves into the technical aspects and underlying causes of biases in FRT, providing a broader and more systematic understanding of the issue. Conversely, Kim[10] analyzes how AI can be integrated, through FRT, in various sectors like employment, healthcare, and law enforcement, as well as the benefits and pitfalls of this integration. This approach is more comprehensive than typical press coverage, which usually highlights individual AI failures or successes without fully exploring the underlying systemic issues.

The issues outlined in both papers are current and relevant for today, and not merely speculative concerns for the future. The biases in FRT are existing problems that have been observed and reported in various studies. Leslie’s[12] paper consolidates these findings, indicating that these are not distant issues but rather immediate challenges that need addressing, while Kim’s[10] analysis extends to the potential long-term societal impacts of AI. The latter illustrates that the issues of AI and inequality are not speculative concerns for the distant future but are pressing matters that are shaping our society, as we speak.

Leslie’s[12] paper indicates that the scientific community is not only aware of these issues but actively engaged in researching them. The paper itself is a testament to the scholarly efforts to document and analyze biases in FRT. Researchers in AI and ethics have been vocal about these problems, publishing studies that expose the limitations and biases of current FRT systems. One study that strongly supports this movement is **A phenomenological perspective on AI ethical failures: The case of facial recognition technology**[13] from the Journal *AI & Society*, written by Wen & Holweg. The study, as well as the previously mentioned ones showcase a clear and growing concern within the academic community. This awareness, as later emphasized in Kim’s paper[10] is not just in terms of technological biases (as often reported in the press) but extends to broader socio-economic impacts. Both works suggests that the academic community is looking at AI not just as a technical or ethical challenge but also as a socio-economic one.

The scientific community is actively contributing to both identifying and solving these issues. While the popular press plays a crucial role in raising public and policy awareness about the immediate impacts of AI, academic works like Kim's[10] provide a deeper understanding of the systemic nature of these challenges. This includes conducting research to understand the nature and extent of biases in FRT, developing more inclusive and fair algorithms, and advocating for ethical standards and regulations in AI technologies. Leslie's[12] paper not only synthesizes existing research on the topic but also discusses potential pathways for mitigating biases, such as improving data diversity and implementing ethical guidelines for AI development. In addition, Yuni Wen and Matthias Holweg's paper[13] also makes a significant contribution by employing a phenomenological approach to examine ethical failures in facial recognition technology (FRT). This perspective enriches the understanding of AI's impact on human experience, going beyond technical aspects to consider the profound personal and societal implications of AI technologies. Wen and Holweg's study complements the works of Kim and Leslie by providing a unique lens to view AI ethics, emphasizing the experiential and human-centric aspects of technology use and its consequences. This research further underscores the importance of holistic approaches in addressing AI ethics, combining technical, systemic, and human-focused analyses.

4 Conclusion

In conclusion, the discussion around AI's impact on society, particularly in the realms of corporate power consolidation and racial bias, reveals a complex interplay between popular media narratives and scientific literature. While journalistic reports like those by Naomi Klein, Emily M. Bender, and the MIT Technology Review series shed light on immediate, real-world implications and ethical concerns of AI, academic research provides a deeper, systemic understanding of these issues. Papers by Michael Hilb, Mihailis E. Diamantis, Pauline Kim, and others offer nuanced insights into the monopolization of intelligence, legal challenges in corporate AI decision-making, inequalities perpetuated by AI, biases in facial recognition technology, and the human experience affected by AI ethical failures. Together, these diverse perspectives highlight the urgency of addressing AI's societal impact, advocating for responsible development and deployment of AI technologies, and emphasizing the need for comprehensive approaches to tackle the ethical, technical, and socio-economic challenges posed by AI.

References

- [1] E. M. Bender. "Ensuring Safe, Secure, and Trustworthy AI": What those seven companies avoided committing to. *Medium*, 7 2023.
- [2] W. Chan. Researcher Meredith Whittaker says AI's biggest risk isn't 'consciousness'—it's the corporations that control them. *Fast Company*, 5 2023.
- [3] P. Dave. AI algorithms are biased against skin with yellow hues. 10 2023.
- [4] M. E. Diamantis. The Extended Corporate Mind: When Corporations Use AI to Break the Law. *North Carolina Law Review*, 98:893–931, 12 2019.
- [5] C. D'Ignazio and L. Klein. Feminist ai. *Feminist AI*, 2023.
- [6] K. Hao. Artificial intelligence is creating a new colonial world order. *MIT Technology Review*, 4 2022.
- [7] M. Hilb. Toward artificial governance? The role of artificial intelligence in shaping the future of corporate governance. *Journal of Management and Governance*, 24:851–870, 7 2020.
- [8] K. Hill. Another arrest, and jail time, due to a bad facial recognition match. *The New York Times*, 2020.
- [9] K. Hill. Eight months pregnant and arrested after false facial recognition match. *The New York Times*, 2023.
- [10] P. Kim. AI and inequality, 9 2021.
- [11] N. Klein. AI machines aren't 'hallucinating'. But their makers are. *The Guardian*, 5 2023.
- [12] D. Leslie. Understanding bias in facial recognition technologies, 2020.

- [13] Y. Wen and M. Holweg. A Phenomenological perspective on AI ethical failures: the case of facial recognition technology. *AI SOCIETY*, 4 2023.