**Navigating the Ethical Terrain of AI: Lessons from Amazon's Recruitment System Failure**

Klaus Schwab (2015) described the Fourth Industrial Revolution as an era where the boundaries between the physical, digital, and biological spheres are increasingly blurred through the fusion of advanced technologies. This revolution is characterised by the growing utilisation of technologies such as artificial intelligence (AI), cloud computing, robotics, 3D printing, the Internet of Things (IoT), and advanced wireless technologies, among others (Xu et al., 2018). One of the most heralded aspects of this revolution is the potential of AI to automate complex processes, enhance efficiency, and unlock new avenues of innovation. However, the adoption of AI also introduces significant challenges and ethical implications, as evidenced by the failure of Amazon's AI-driven recruitment tool in 2015.

Amazon's project on AI-driven recruitment aimed to optimise the search for top talent by reviewing job applicants' resumes and ranking them with a system similar to product ratings on their e-commerce platform. This initiative was part of a broader effort to automate and optimise operations across the company. However, the tool developed by Amazon encountered a significant setback when it was discovered to exhibit bias against female candidates. This bias stemmed from the AI's training on a decade's worth of resumes submitted to Amazon, which predominantly came from men, reflecting the gender imbalance in the tech industry. As a result, the AI learned to favor male candidates over female candidates, even penalising resumes that included words like "women’s" or those from all-women's colleges (Dastin, 2018).

The discovery of this bias led to Amazon's decision to eventually disband the project, acknowledging the limitations and potential discriminatory effects of their AI recruitment tool. This failure highlights several key implications for the Fourth Industrial Revolution and the use of AI in the workplace:

1. Ethical AI Use and Bias Mitigation: It highlights the importance of addressing biases in AI algorithms, particularly those trained on historical data that may reflect existing prejudices. There is need for organisations to implement robust bias detection and mitigation strategies to ensure AI tools do not perpetuate discrimination.
2. Transparency and Accountability: There is a need for greater transparency in how AI systems make decisions, especially in critical applications like recruitment. Organisations should be accountable for the outcomes of their AI systems, with mechanisms in place to review and address potential issues.
3. Regulatory and Ethical Frameworks: This incident emphasises the urgency of developing comprehensive regulatory and ethical frameworks governing AI use.

This can be seen in frameworks like the European Union’s AI Act which aims to make sure that AI systems used in the EU are safe, transparent, traceable, non-discriminatory and environmentally friendly (European Parliament, 2023). This serves as a foundational step towards establishing a legal structure that can adapt to the rapidly evolving landscape of AI technologies.

1. Human Oversight: The Amazon case illustrates the indispensability of human oversight in AI applications. AI systems should be viewed as tools to assist human decision-making, not replace it. The EU’s AI act reinforces the need for AI systems to be overseen by people, rather than by automation, this is often refered to as the ‘human-in-the-loop’ approach(Yakimova & Ojamo, 2023). In this approach, decisions made by AI systems can be reviewed and overridden by human operators. This ensures that the ongoing human engagement corrects and guides AI systems towards ethical, unbiased and beneficial outcomes.
2. Diversity in AI development: Finally, the failure of the recruitment system highlights the critical need for diversity among teams developing AI systems. A more diverse team is more likely to identify and prevent biases in AI algorithms. EU’s AI act does not explicitly address diversity in AI development but insists that AI developers should conduct risk assessment to ensure their systems are fair, transparent, and non-discriminatory (Zync, 2023). This requirement indirectly promotes the idea that diversity in AI development teams is essential for identifying potential risks and biases in AI systems. By bringing together individuals from various backgrounds, perspectives, and experiences, teams are better equipped to foresee and mitigate issues that may not be apparent to a more homogenous group.

The failure of Amazon’s AI recruitment system in 2015 highlights the dangers of deploying AI without proper oversight and ethical considerations. It emphasises the significance of regulatory frameworks like the EU’s AI act, which advocate for safety, transparency, and fairness. Although Amazon encountered setbacks, they have since learned from their mistakes and implemented a more cautious approach. They now use a simplified version of the recruitment engine for tasks such as eliminating duplicate candidate profiles and promoting diversity (Castleberry et al., 2023) .

**References**

Castleberry, C., Hastings, J., & Palacios, A. (2023, June 6). How Amazon leverages AI and ML to enhance the hiring experience for candidates. Available from: <https://www.aboutamazon.com/news/workplace/how-amazon-leverages-ai-and-ml-to-enhance-the-hiring-experience-for-candidates> [Accessed 3 April 2024]

Dastin, J. (2018, October 11). Amazon scraps secret AI recruiting tool that showed bias against women. Available from: <https://www.reuters.com/article/idUSKCN1MK0AG/> [Accessed 29 March 2024]

European Parliament. (2023, December 19). EU AI Act: first regulation on artificial intelligence. Available from: <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=Parliament’s%20priority%20is%20to%20make,automation%2C%20to%20prevent%20harmful%20outcomes> [Accessed 31 March 2024]

Klaus Schwab. (2015, December 12). The Fourth Industrial Revolution.

Xu, M., David, J. M., & Kim, S. H. (2018). The Fourth Industrial Revolution: Opportunities and Challenges. International Journal of Financial Research, 9(2), 90. Available from: <https://doi.org/10.5430/ijfr.v9n2p90> [Accessed 29 March 2024]

Yakimova, Y., & Ojamo, J. (2023, December 9). Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI. European Parliament. Available from: <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai> [Accessed 1 April 2024]

Zync. (2023, July 7). The EU AI Act: The Future of Artificial Intelligence. Cyber Security Insights. Available from: <https://www.linkedin.com/pulse/eu-ai-act-future-artificial-intelligence-zyncgroup/> [Accessed 1 April 2024]