Data Science Ethics -Cases of discrimination

Dana Assad
1211452@student.birzeit.edu

Rawan Amer
1210295@student.birzeit.edu

Introduction

➢ ***Def of Data science:***

Data science encompasses the analysis of large datasets to identify patterns and derive meaningful insights. These datasets are drawn from different sources and often in different formats or formats that require processing to generate a format suitable for the next step of building predictive models using machine learning techniques. It is helpful in developing the smart models to develop intelligent systems and helps in taking the strategic business decisions.

The applications of data science are vast and impactful:
digital companies like Amazon and Facebook use data science to predict consumer behavior for advertising purposes; leading pharma firms like Pfizer for drug efficacy and use of IBM Watson for diagnosing diseases and other financial institutions like MasterCard to use data science in detecting fraud incidences in banking and other transactions for increased security and efficiency.

➢ ***Def of data discrimination:***

Is defined as the ability to detect differences between two sets of data based on specific criteria, similar to how individuals detect differences between lights in chromatic discrimination studies.

What Happened?

Big data and analytics and AI have changed the ways of operations by bringing data analysis and automation to industries. Nevertheless, these technologies have been largely coupled with ethically questionable features especially on the aspect of discrimination.

➢ Several high-profile cases have highlighted these systemic issues:

a. The different categories that have been impacted by algorithmic bias include patient care delivery in healthcare. Research has also shown that statistical models of health care adopted in patient risk profiling and resource distribution contained higher discrepancies in lessening the health needs of black people than fellow white counterparts with comparable health status, thus worsening the pre-existent racial discrimination in the health care systems.

b. Algorithmic bias has also made its way into influencing the criminal justice system. Studies of risk-assessment measures adopted for both sentencing and parole practices, demonstrate racial implications with the algorithms predicting higher re-offending rates among black individuals even when other characteristics are accounted for. This has raised critical questions on the entrenchment of systematic racism through automated systems and programmers.

c. Hiring Algorithms: Since 2014, an AI research team aimed to create a machine learning recruitment system to automate talent searches by analyzing resumes. The goal was to identify the top five candidates from 100 applicants. However, by 2015, Amazon found that the system was biased against gender, particularly in software development roles, due to the male-dominated resume data from the past decade. Despite algorithm adjustments to minimize bias, concerns remained about potential discrimination through word patterns in resumes. Experts like Nihar Shah highlight ongoing challenges in ensuring system fairness and interpretability.

d. Gender Shades Study: Joy s 'Buolamwini and Timnit Gebru Gender Shades study showed that there was a high level of inequity where the performance of facial recognition technology differed considerably depending on the demographic of the person. The study showed that these systems performed the worst when it came to skinned individuals and -darker women, with the worst rates recorded in cases where the subjects were both black and female. The revelation of serious issues in the technology Gender Shades by published the study, it shows that gender classification systems from major companies were less accurate darker skin women than lighter for skin men.

Classifier	Metric	All	F	M	Darker	Lighter	DF	DM	LF	LM
MSFT	TPR(%)	93.7	89.3	97.4	87.1	99.3	79.2	94.0	98.3	**100**
	Error Rate(%)	6.3	10.7	2.6	12.9	0.7	**20.8**	6.0	1.7	0.0
	PPV (%)	93.7	96.5	91.7	87.1	99.3	92.1	83.7	**100**	98.7
	FPR (%)	6.3	2.6	10.7	12.9	0.7	6.0	**20.8**	0.0	1.7
Face++	TPR(%)	90.0	78.7	99.3	83.5	95.3	65.5	**99.3**	90.2	99.2
	Error Rate(%)	10.0	21.3	0.7	16.5	4.7	**34.5**	0.7	9.8	0.8
	PPV (%)	90.0	98.9	85.1	83.5	95.3	98.8	76.6	**98.9**	92.9
	FPR (%)	10.0	0.7	21.3	16.5	4.7	0.7	**34.5**	0.8	9.8
IBM	TPR(%)	87.9	79.7	94.4	77.6	96.8	65.3	88.0	92.9	**99.7**
	Error Rate(%)	12.1	20.3	5.6	22.4	3.2	**34.7**	12.0	7.1	0.3
	PPV (%)	87.9	92.1	85.2	77.6	96.8	82.3	74.8	**99.6**	94.8
	FPR (%)	12.1	5.6	20.3	22.4	3.2	12.0	**34.7**	0.3	7.1

Table 4: Gender classification performance as measured by the positive predictive value (PPV), error rate (1-TPR), true positive rate (TPR), and false positive rate (FPR) of the 3 evaluated commercial classifiers on the PPB dataset. All classifiers have the highest error rates for darker-skinned females (ranging from 20.8% for Microsoft to 34.7% for IBM).

Table 1: Gender Classification

Table 1 shows the gender classification results of Microsoft, Face++, and IBM on the Pilot Parliaments Benchmark (PPB) dataset, which has serious biases. All systems provided better identification accuracy for males and lighter skin, with darker skin females being the least identified having error rates ranging 20.8% and 34.7%. This underlines the need for diversity in training data sets and fairness in AI design and construction.

Who Is Responsible?

Responsibility for decisions generated by AI systems is distributed among several parties, reflecting the complexity of the issue. This can be summarized in the following key points:

- Developers: Developers must ensure training datasets reflect diverse populations and test algorithms for biases. They are responsible for biased or incorrect decisions stemming from the systems they create.

- Companies: Companies implementing AI must enforce ethical guidelines neglecting fairness leads to discrimination and prioritization of profits.

- Users: Users should use these technologies carefully and are responsible for knowing how they operate.

In summary, Responsibility for ethical concerns is shared among developers, users and Companies, requiring collaboration.

Ethical Analysis:

Evaluate ethical implications using three ethical frameworks.

1. *Kantianism:*

According to Kantian ethics, people should be treated as ends in and of themselves, which implies they should not just be utilized as a means to an end but also for their inherent value. Because discriminatory algorithms are biased, they reduce humans to data points and deny them equal opportunity, which is a violation of this fundamental principle. Applying this to the examples:

- **Gender Shades study** demonstrated how biased facial recognition systems were less accurate when identifying darker-skinned women, treating them as less valuable in technological interactions. This diminishes their intrinsic worth, as they are being reduced to inaccurate data points, violating Kantian ethics.

- **Amazon's AI recruitment tool** systematically disadvantaged female candidates by prioritizing resumes from a male-dominated pool. This not only denied women equal opportunities but also undermined their dignity by treating them as a means to achieve recruitment goals rather than recognizing their inherent value as individuals

Kantian ethics deems these practices unethical, undermining individual worth and equality.

2. *Act Utilitarianism*

Act Utilitarianism evaluates actions by their consequences, aiming to maximize overall happiness and minimize harm. Applying this to the examples:

- Amazon’s AI Recruitment Tool: The tool was biased against female candidates, causing harm by systematically reducing their opportunities. While it improved efficiency for some, the harm to women outweighs the benefits. Therefore, from an act utilitarian perspective, it is unethical.

- Gender Shades Study: The study showed facial recognition technology was less accurate for darker-skinned and female individuals, leading to discrimination. The negative impact on marginalized groups outweighs the benefits to others. This makes the technology unethical from an act utilitarian viewpoint.

In both cases, the harm caused by discrimination to underrepresented groups outweighs the benefits, making these algorithms unethical based on act utilitarianism.

3. *Rule Utilitarianism:*

Rule utilitarianism advocates for regulations that maximize well-being, as demonstrated by the Gender Shades research and Amazon's AI hiring tool.

- Amazon's AI Recruitment Tool: Rules could be established to use diverse and inclusive datasets, and conduct fairness audits during algorithm development to avoid bias against women or any other group. This promotes equality and reduces discrimination.

- Gender Shades Study: Rules should ensure the use of diverse datasets that reflect all demographic groups, including darker-skinned individuals and women, as well as conducting fairness reviews during system design to ensure no bias against any group.

AI can enhance outcomes and reduce harm by emphasizing diversity and ongoing fairness assessments, aligned with Rule Utilitarianism.

Addressing Ethical Issues in Data Science

In order to resolve discrimination-related ethical concerns in data science, the following steps should be taken:

i. **Improve data diversity**, make sure that the datasets used to train AI models include a variety of age groups, genders, and ethnicities. To get inclusive statistics, work with communities and organizations.

ii. **Develop Fair and Open Algorithms**: When training algorithms, minimize biases by utilizing fairness-aware machine learning approaches. Conduct routine audits of AI systems to find and address discriminatory trends, guaranteeing openness and responsibility.

iii. **Increase Team Diversity**: Assemble a broad group of developers that will contribute a range of viewpoints to the creation and application of AI systems. Encourage sociologists and ethicists to be involved in AI initiatives in order to solve ethical issues.

iv. **Create Ethical Standards and Guidelines**: Create industry norms for moral AI behavior with an emphasis on accountability, transparency, and justice. Legislation and supervision should be used by governments and regulatory agencies to enforce compliance.

v. **Increase Community Involvement**: Hold conversations about ethical AI with stakeholders from academia, business, and civil society. Encourage public education initiatives to inform people about the possible drawbacks and advantages of artificial intelligence.

Conclusion:

The issues of data discrimination in ethical concerns of AI are challenge that needs to be addressed by developers, organizations and society. If we want to create fair Artificial Intelligence that are safeguarding human dignity and obeying social justice standards – we must work these values into development processes. Ethical AI is not a mere technology problem but a problem of ethics which will demand the efforts of all actors. With data science becoming more prevalent in society, its moral principles must make technology’s benefit to all society.

BS These questions can be answered in a way that enables considerations for the greater good and the potential of data science with all citizens in mind:

References:

[1]:Carnegie Mellon University. (n.d.). *Amazon scraps secret AI recruiting engine that showed Biases against women - machine learning - CMU - Carnegie Mellon university*. Machine Learning | Carnegie Mellon University. Retrieved January 12, 2025.

[2]:Buolamwini, J., & Gebru, T. (2018). *Gender shades: Intersectional accuracy disparities in commercial gender classification*. 77–91.

[3]:*Unit 6. Ethical Issues in Data Science PDF*. (n.d.). Scribd. Retrieved January 12, 2025

[4]:What is data science? (2024, December 17). *Ibm.com*.

[5]:Analytics, E. I. (2024, February 14). *Ethical considerations in Data Science and AI - emerging India analytics*. Medium.

[6]:(N.d.). Sciencedirect.com. Retrieved January 12, 2025, from