Irrationality of Algorithmic Discrimination: An Agent-Based Modelling Approach

ST 313 - Ethics For Data Science
London School of Economics
Isabel Mora, Zita Szeker, Baris Yazici

1 Introduction

In the current digital era, data-driven decisions and algorithmic predictions impact every aspect of life, and algorithmic fairness has risen as a primary concern among various actors affected by technological predictions. Yet, in the evolving landscape of algorithmic development, the achievement of fairness often collides with the practical challenges faced by companies seeking to sell their models or use the algorithm in question to maximise their profit. The cost of fairness in algorithmic decision-making is a complex subject that entails the consideration of the interests of the company using or selling the technology and the well-being of those affected by the algorithm. Companies, driven by the inherent need to prioritise profit, often deem the cost of fairness too high, and thus will opt to consider fairness in a way that will minimise the overall cost when designing algorithms. This can lead to discriminatory practices, as companies often conclude, based on the most powerful stakeholder's utility definition, that the cost of discrimination is lower than the cost of fairness.

Motivated by the heightened concerns and discussions surrounding algorithmic fairness, this project will delve into the question of whether discrimination is actually a rational strategy for companies through analysing if short-term profits obtained through discriminatory practices imply long-term benefits for companies and society overall. Such an analysis considers utility functions for the companies and the people affected by the algorithm, as well as a precise definition for cost of fairness and cost of discrimination. These terms have ambiguous definitions shaped by subjective interests, and no single interpretation fits every specific case concerning algorithmic decision-making. Thus, in order to address the motivating question and contribute meaningful insights to the topic of algorithmic fairness, the paper will explore a specific storyline and use agent-based modelling to simulate scenarios that resemble interactions between companies and individuals affected by predictive algorithms.

The goal of this project is to provide a comprehensive understanding of the complex relationships between algorithmic decision-making, discrimination and societal well-being. By considering the rationality of discriminatory practices in the pursuit of short-term profits, we aim to provide insights into the trade-offs and ethical considerations faced by companies when employing algorithms. This discussion is highly relevant for society, as it addresses the broader effects of algorithms on individual and societal prosperity, which will become increasingly important as technology continues to evolve and impact every aspect of human lives.

2 Theoretical Background

This section includes research on cost of discrimination, utility functions and agent-based modelling, all of which are integral to the development of our storyline and subsequent analysis.

2.1 Cost of Discrimination

The cost of discrimination in algorithmic decision-making involves the multifaceted impact on both individuals affected by the algorithms and the broader societal structure. Discrimination can appear in various forms, such as biased outcomes based on race, gender, or socioeconomic status. Regardless of whether the systematic bias embedded within algorithms was intentional or not, its negative effects are significant and thus should not be overlooked. For some examples, the cost of discriminatory processes can be more easily understood, such as for instance the loss of productivity and therefore economic output due to discrimination based on gender or other sensitive variables in workplaces. Or another example could be the future benefits lost from the exclusion of people living with disabilities in education. In the case of algorithms, the interpretation heavily depends on the chosen definition of fairness, but in general the cost of fairness or nondiscrimination can be interpreted as the loss of accuracy when choosing a fair algorithm over an unfair one. It should be noted that with many stakeholders in such situations, facing different utility functions, it is inevitable to have several fairness definitions. The article *Fairness,* *utility and power in algorithmic decision-making* states that one of the key limitations of the leading notions of fairness is that they "legitimise and perpetuate inequalities justified by 'merit' both within and between groups. The focus on 'merit' – a measure promoting the decision-maker's objective – reinforces, rather than questions, the legitimacy of the status quo" [15]. This means that when the managers in a company are faced with a decision, they might choose a fairness definition that maximises their own utility levels. However, as utility is often unknown, one can use merit as a prediction of future utility levels. This focus on merit can, as stated above, legitimise inequalities and can act to support the decision-makers goals. This project aims to extend this observation to the profit of companies, as firms might knowingly or unknowingly engage in discriminatory practices on the basis of higher short-term profits.

2.2 Utility Functions

In order to understand how companies and individuals are affected by predictive algorithms, one needs a comprehensive measure reflecting the overall satisfaction or pleasure that consumers, or in this case companies, receive from consuming a good or accessing a service. In the field of economics, utility functions are commonly used for this purpose. This section of the paper explores the economic theory behind utility functions, highlighting some common forms of utility functions and their potential disadvantages. Many forms of utility functions exist in economic literature, and each form is tailored to better understand the mechanisms behind human behaviour in a specific situation. Therefore an overview of some common utility functions will help determine an appropriate form of utility functions to use in the simulation for this project.

Utility functions measure consumers' preferences over a set of goods and services in units of utils [1]. These functions are used to better understand human behaviour and preferences, as consumers/companies are assumed to act rationally when they are faced with decisions, meaning they make choices that maximise their utility levels. To make utility functions more realistic, it is often assumed that they are subject to diminishing marginal utility, meaning that as consumption increases, the additional utility derived from each additional good decreases. In economics literature, two types of utility are usually distinguished: ordinal and cardinal utility. Ordinal utility aims to rank the choices by preference, but does not consider by how much the consumer prefers one choice to another. On the other hand, cardinal utility indicates the degree to which one choice is preferred to another by the consumer [1]. It is important to mention that it is challenging to assign an exact and accurate numerical value to satisfaction, thus one can interpret utility functions as models aiming to approximate the true utility levels of consumers.

Building on the general definition of utility functions, this section will consider the most common and applicable forms of utility functions. There are many forms and models of utility functions; some are used in specific situations with several assumptions about the background and behaviour. For this analysis we will only focus on the most frequently used functions, which still cover most areas of economic literature, thus this analysis will still give a comprehensive overview of the possible function forms. We thus analyse 5 main forms of utility functions:

1. **Additive utility** uses a straightforward method of taking the sum of utilities from a set of items. One important constraint of this method is that it assumes the independence of the considered goods, which largely reduces the scope where this function can be used [8]. There are many alterations of additive utility, such as weakly additive utility, subadditive utility and weighted combinations of the functions [12].

2. The **Cobb-Douglas utility function** is based on the well-known Cobb-Douglas production function, describing how output is related to capital, labour and technology. For two goods, the Cobb-Douglas utility function can be represented as $U = X^{\alpha}Y^{\beta}$, where X and Y refer to the goods and α and β are utility elasticities adding up to 1. For a small number of goods, this function works well, however, there might be an issue when applied to a large number of inputs [16].

3. **Linear utility** is described by the function $u(x_1, x_2, \ldots, x_m) = w_1x_1 + w_2x_2 + \cdots + w_mx_m$, where w represents the subjective preferences of consumers. A drawback of this approach is its number of assumptions, such as strictly monotone preferences, straight-line indifference curves and substitute goods [11]

4. The **von Neumann-Morgenstern utility** uses probabilities to calculate the expected utility of consumers and they are assumed to make decisions that maximise the expected utility. This function is mostly applicable for lotteries and decision-making under risk. It is commonly shown as an S-shaped utility function, incorporating risk aversion when considering losses [3].

5. **Quasilinear preferences** take the form $u(x) = x_1 + v(x_2, \ldots, x_L)$ and are useful since they are not subject to the wealth effect [17]. To incorporate diminishing marginal utility, one usually uses concave functions, such

as $log(x)$ or $\sqrt{x}$ in this case.

Overall, one can see that many of the forms of utility functions are specific to certain situations and/or constrained by several assumptions. The form that is the least affected by these restrictions is the quasilinear utility. This form is straightforward to use, and it reflects an important property of utility, namely diminishing marginal utility. Therefore, the quasilinear utility function will be chosen for the simulation in this paper.

2.3 Agent-Based Modelling

Agent-based modelling is a methodology used to build formal models of real-world systems that are made up by individual units which repeatedly interact among themselves and/or with their environment. In this type of simulation, the target system is represented as a collection of autonomous agents, each possessing distinct characteristics, decision-making processes, and rules for interacting with other agents and their environment. These agents navigate the simulated environment based on their programmed behaviours, and their interactions lead to emergent patterns and system-level outcomes. This approach contrasts with other techniques, such as equation-based modelling, where entities of the target system are represented through average properties or single representative agents. In an agent-based model, the individual units of the real-world system to be modelled and their interactions are explicitly and individually represented in the model, and no additional assumptions are introduced. This makes it a flexible technique that can be tailored to fit varying levels of complexity [4].

Agent-based modelling is particularly valuable for capturing the heterogeneity and dynamic nature of real-world systems, allowing researchers to explore how the micro-level interactions of individual agents contribute to macro-level phenomena. Thus, this technique is widely used in sociology, economics and qualitative research [2]. Additionally, due to its bottom-up construction, agent-based modelling is considered a useful technique for reproducing social phenomena and understanding the casual structures underpinning observable macro-level behaviour [14]. This technique has been used successfully, for example, to explore behavioural dynamics affecting climate [10], reproduce and understand the phenomenon of bullying [14], and understand and model crime patterns [13]. Overall, although agent-based models can lead to complex simulations with unnecessary details, it is a technique with extensive variations that has been implemented in the social sciences with a great amount of success [6].

For this project, we use agent-based modelling to understand the macro level behaviour of a system by making micro level assumptions about cash flow between individuals and companies, and the amount of utility these flows provide. Specifically, we are interested in the accumulated wealth of individuals relative to the neighbourhood they live in, and how we explore this phenomenon will be carefully explained when we present the setup, metrics, and results of the simulations.

In order to build an agent-based model in Python, we use the open-source Mesa library [5], which provides a framework for creating the four essential components of an agent-based model: agents (individuals and companies, in our case), attributes and actions (wealth of agents, giving money, etc), schedule (the time component of the model), and grid (the spatial component, which are neighbourhoods in our model). Mesa also provides a direct framework for running the model many times and for collecting and visualising data from the simulation. Although there are multiple implementations of agent-based modelling in Python, and additional specialised environments such as NetLogo [18], we decided to build our simulation using Mesa due to its user-friendly design, flexibility in defining agent behaviours, built-in visualisation tools, and seamless integration with other Python libraries.

3 Case Study: Amazon Same-Day Delivery

This section presents Amazon's Prime Free Same-Day Delivery as an example of the complex relationships between algorithmic decision-making, discrimination and company profitability.

Developed and promoted as a means to provide instant gratification to costumers and gain a competitive advantage over physical stores, the service, in cities where it was initially available, offered Prime members same-day delivery of more than a million products for no extra fee on orders over $35. However, a detailed analysis of the service carried out by Bloomberg in 2016 [9] revealed troubling disparities, particularly along racial lines.

Eleven months after it was implemented, the service included 27 metropolitan areas and provided broad coverage in the majority of them. However, in six major same-day delivery cities, a Bloomberg analysis found that predominantly black ZIP codes are excluded to varying degrees, presenting a concerning pattern of racial disparity. In cities like Atlanta, Chicago, Dallas, and Washington, which have struggled with historical racial segregation and economic inequality, black citizens are about half as likely to live in neighbourhoods with access to Amazon same-day delivery as white residents. Likewise, in New York City the service is available in predominantly white areas, while excluding the Bronx and certain majority-black neighbourhoods in Queens. In Boston, three ZIP codes covering the primarily black neighbourhood of Roxbury are not included, while the surrounding neighbourhoods on all sides maintain eligibility [9]. Figure 1 shows maps, divided by neighbourhood, of the six major cities analysed by the article. The article also goes on to analyse each city independently in greater detail with additional graphs, which revealed that factors such as crime rates, income, and neighbourhood demographics also contribute to the complex web of exclusions.

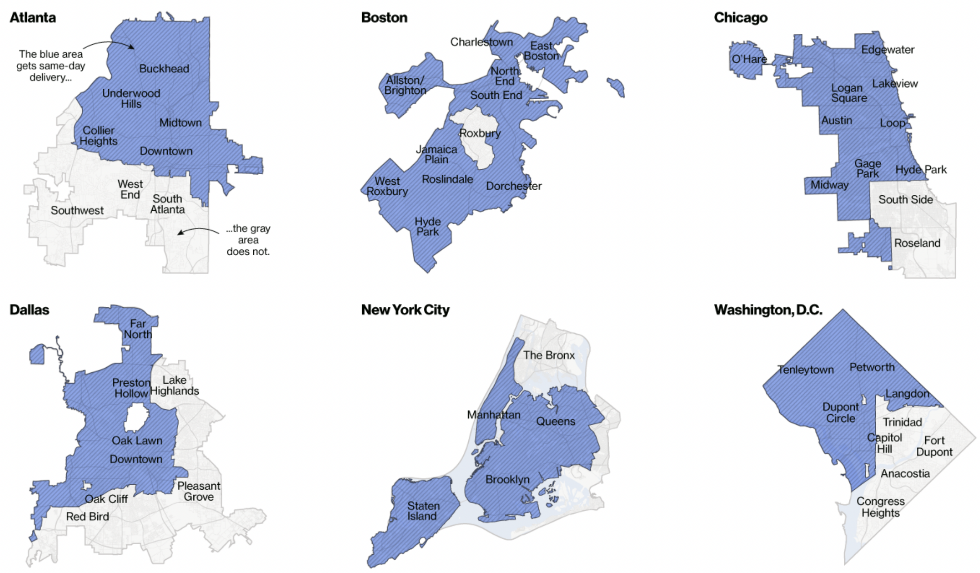

Figure 1: Amazon's Prime Free Same-Day Delivery Coverage in Six Major US Cities [9]

Amazon denied making delivery decisions based on race, emphasizing a focus on serving as many people as possible. The company claimed that ethnic composition is not considered in drawing up delivery maps, as their only goal was to prioritize ZIP codes with a high concentration of Prime members. From a standpoint of cost-effectiveness and efficiency, prioritizing areas with a higher number of current paying members for a new product makes logical sense. However, in cities where these paying members are primarily located in predominantly white neighborhoods, relying solely on data-driven calculations that focus on numbers rather than people can reinforce inequality in access to retail services [9].

The analysis of Amazon's Prime Free Same-Day Delivery was carried out in 2016, and since then the service has evolved and improved. However, the case study still serves to demonstrate how companies using data-driven decisions tend to prioritise profits, potentially perpetuating discriminatory practices. Even when not overtly intentional, the pursuit of efficiency and financial gains can inadvertently contribute to inequalities. Thus, this case study highlights the importance of questioning the pursuit of short-term profits in exchange for an increasingly unequal society.

4 Simulation Storyline

Inspired by cases like Amazon's same-day delivery algorithm outlined in the previous section, we present a fictional storyline to provide context for the simulation that follows. The narrative story provides a tangible backdrop that grounds the simulated scenarios in real-world relevance, making the outcomes more relatable and insightful.

Our storyline is centred in the city of Velvet Vale, where the private company CityPulse Technologies stands at the forefront of urban innovation. As the biggest tech company in the region, CityPulse specialises in cutting-edge solutions designed to optimise various aspects of urban life. The company's main tool is their signature Harmony Algorithm, a predictive analytics tool that aims to enhance the quality of life for Velvet Vale's residents. This algorithm dynamically optimises traffic flow, reduces energy consumption, and tailors multiple public services to individual preferences, fostering a uniquely personalised urban environment.

Although the company has been crucial in fostering Velvet Vale's economic growth and has positioned the city as a leader in urban innovation, CityPulse Technologies is now facing backlash from investors and stakeholders amidst a steady decline of the company's profit in the last year. Seeking to regain profitability and restore investor confidence, the company has decided to introduce the Economic Prosperity Module (EPM) to upgrade its Harmony Algorithm. This new feature analyses socioeconomic indicators such as income levels, employment status, and educational backgrounds in order to identify areas and residents considered key contributors to Velvet Vale's economic growth. CityPulse will then increase public services in these areas and provide the identified individuals with tailored incentives in order to boost the city's economic activity (and thus the company's profits).

The company is aware of the ethical implications of a raw implementation of the EPM, as the module will inevitably target high-income areas and individuals for the introduction of new public services and incentives. The module has been perfectly designed to increase CityPulse's profits in the immediate short term, yet all citizens will not benefit equally from the new algorithm. For example, Maria, an uneducated single mother living in the city's periphery, will witness a decline in crucial public services that were tailored to address the unique needs of her community. On the contrary, Robert, a successful entrepreneur living in an affluent neighbourhood, will experience a significant boost in his already high standard of living.

The simulation that follows will explore two scenarios arising from contrasting decisions regarding the implementation of the new algorithm. In the first scenario, CityPulse will opt for discrimination and prioritise short-term profits by not incurring the cost of creating a fair algorithm. In the second scenario, a policy intervention by the local government will force CityPulse to recognise the ethical concerns of the original definition of the EPM and they will thus invest resources into creating a new module that ensures public services and incentives are distributed more equitably across all neighbourhoods. This action will inevitably lead to less profit in the short term in comparison to the first scenario.

Through the analysis of the company's profits, the citizens' wealth and the city's overall equality, we seek to analyse the effect of CityPulses's choices on societal well-being and their own long-term profits. This exploration aims to discern whether discrimination, as a strategic choice, is a rational decision. The specific set-up of the simulation is described in the section that follows.

5 Simulation Study

Building on the storyline detailed previously, this section presents the formal structure of the agent-based model developed to simulate the interaction between the private company CityPulse Technologies and Velvet Vale's citizens. We will analyse two simulations: one where the company continuously discriminates and the other where the discrimination generated by the company is prevented. As a conclusion, we will suggest that it is irrational for the company to discriminate, as this will inevitability lead to a decline in their profits in the long term and a very unequal society.

Our agent-based model was developed using the Python library Mesa [5], and our analysis relies on specific metrics and assumptions that will be detailed in the following subsections. The actual code for the simulation can be found in the two Python Notebooks attached to the submission.

5.1 Metrics

We first define a set of metrics that will be calculated throughout the simulation and used as comparison indicators between the two scenarios.

Gini Coefficient The Gini coefficient is a measure of income inequality widely used in economics and sociology. It quantifies the extent to which the distribution of income or wealth among individuals or households deviates from perfect equality. The coefficient is expressed as a value between 0 and 1, where a higher value implies greater inequality [7]. For our simulation, we will use the Gini Coefficient as the indicator of the city's wealth inequality.

Suppose x_{ik} is the wealth of the i-th agent at time step k, and N is the total number of agents in the simulated closed system, then the Gini coefficient at each time step k can be expressed as [5]:

$$G_k = 1 + \frac{1}{N} - 2B_k \tag{1}$$

Where B_k is given by:

$$B_k = \frac{\sum_{i=1}^{N} (x_{ik} \cdot (N - i))}{N \cdot \sum_{i=1}^{N} x_{ik}}.$$

Profit and Marginal Profit: In microeconomics, marginal profit refers to the profit made by a company with each additional unit produced. For our simulation, however, we define marginal profit as the increase in total profit of the company from time step k to $k + 1$. We refer to the overall sum of the net earnings of the company from each time step as the profit.

Wealth: Every agent has some amount of money at each time step k, which is defined to be equal to their wealth. That is, we use the terms 'money' and 'wealth' interchangeably.

Life Quality: Every agent has a life quality measure starting at 0. Since we are interested in the dynamics of the company's provided service and people's wealth, we define an agent's life quality as the net life quality increase caused by the service of the company. We increase the life quality of the agent every time the company provides service to that agent.

5.2 Assumptions

We list the main underlying assumptions of the simulation below.

1. **Agents have diminishing marginal utility**. This is a standard way of looking at utility, as argued in Section 2.2. Note that we use the terms utility and life quality interchangeably.
2. **The agent's life quality has a causal relationship with the agent's wealth**, and the relationship holds in both directions.
3. **The company's marginal profit has a causal relationship with the wealth distribution in the area where the company is operating**. This is reasonable because if there is a huge wealth distribution inequality, then fewer people will have access to the service, which means the company is missing out on the opportunity of earning more. One could argue that the company could make reasonable profit by targeting only those who are on the high income side of the wealth distribution, but those people might no longer desire/need the service due to either (I) diminishing marginal utility or (II) their relatively infrequent need of the service.
4. **The marginal cost of the company is constant for each additional unit service provided**. This is not a realistic assumption, as the company technically provides different services that can have different costs depending on their purpose and the people they serve. However, this assumption is necessary to prevent an overly complicated simulation, and it doesn't change the outcome of our study.
5. **The expected long term profit is a determinant of a company's value**, which is why companies have an incentive to maximise both short term and long term profits.

5.3 Setting

We simulated a closed system of 100 agents distributed throughout 10 neighbourhoods and a company that provides service to agents. At each step we keep track of the metrics outlined in section 5.1.

Neighbourhoods: We have high-income and low-income neighbourhoods. We denote the initial wealth distribution of the low-income neighbourhoods as W_l and of the high-income neighbourhoods as W_h. W_l and W_h follow different normal distributions:

$$W_l \sim \mathcal{N}(4, 2) \ \text{ and } \ W_h \sim \mathcal{N}(7, 2).$$

Agents: Each agent has an initial wealth, which is determined by the neighbourhood they belong to. We define three basic operations for these agents so that they interact with each other. Agents can give money to each other (if they have money), in which case their wealth decreases by 1 unit. Consequently, agents can receive money, which increases their wealth by the square root of 1 unit multiplied by $(1 + \text{life quality of agent at that step})$. This choice was made in order to reflect the compounding effect of an agent's life quality on their earnings. Agents can also move along the grid, which means they do not only interact with agents that share the same income characteristics. We keep track of the wealth of each agent in every step.

The Company: The company provides service to agents. The action of providing service is defined using three basic operations. With each service provided the company makes some profit, which is a function of both the value of the service and the Gini coefficient. If the current Gini coefficient is relatively higher, the company's profit for that step is relatively lower. Providing service also has a constant unit cost for the company, which was a simplifying assumption made. With each service provided, the life quality of the agents receiving the service increases by the square root of their life quality at step $k-1$, which reinforces the fact that agents are assumed to have diminishing marginal utility. We keep track of the profit and cost of the company at every step.

5.4 Results

We now present and analyse the results obtained from the two simulated scenarios. It must be noted that the units are arbitrary and magnitudes for metrics may change based on the run of the simulation and are not representative of real cases, which is why we only interpret the relative patterns of the curves.

5.4.1 First Scenario: Discrimination

This is the scenario where the company discriminates based on location. That is, we set up the rules of the simulation so that the company only provides service to agents belonging to a high-income neighbourhood. From Figure 2, we see that, as the simulation progresses, the wealth of individuals who live in high-income neighbourhoods (non-discriminated) increases, whereas the wealth of individuals in low-income neighbourhood is almost constant. This affects the income distribution of the city and increases the Gini coefficient, as observed in Figure 4. As expected, as the Gini coefficient increases and the wealth inequality exacerbates, we observe a decrease in the company's marginal profit and a flattening out of the cumulative profit (see Figure 3). As the time steps are arbitrary, we can think of the first ~ 50 steps as the short-term. We see that discriminating based on area does not hurt the short-term profits of the company, but greatly hurts the long-term profits. This means there is no cost of discrimination for the company in the short term, but there is a positive cost of discrimination in the long term.

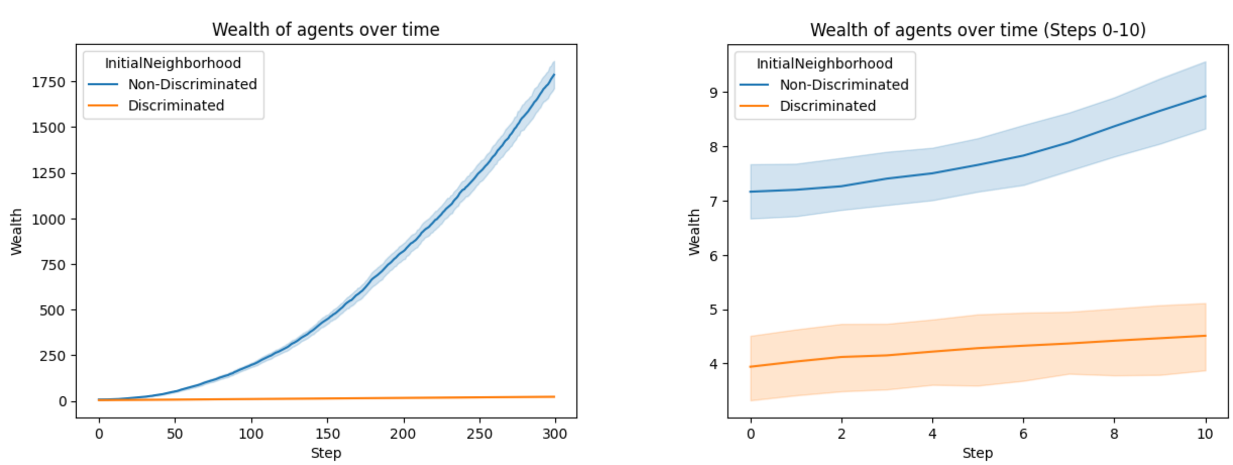

Figure 2: Wealth of Agents in Discrimination Scenario.

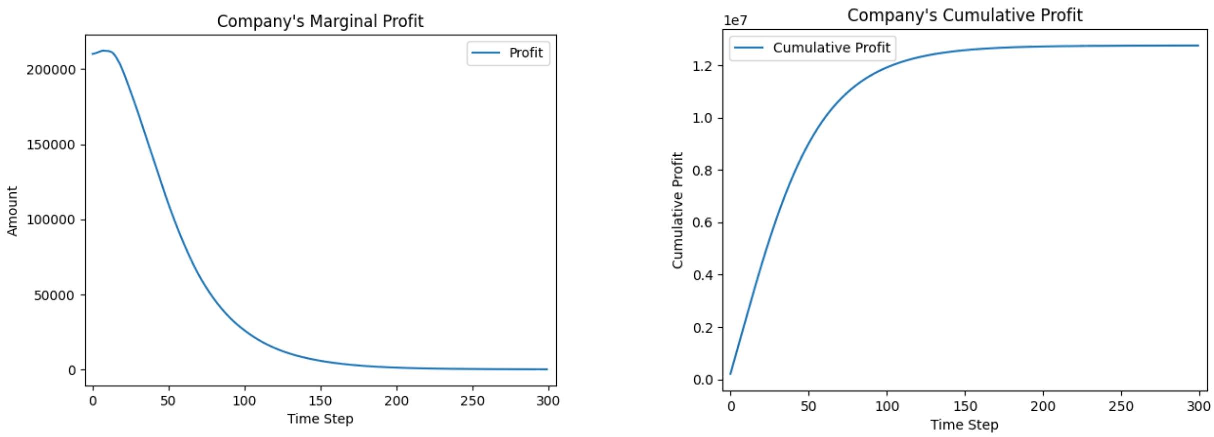

Figure 3: Company's Cumulative and Marginal Profit in Discrimination Scenario.

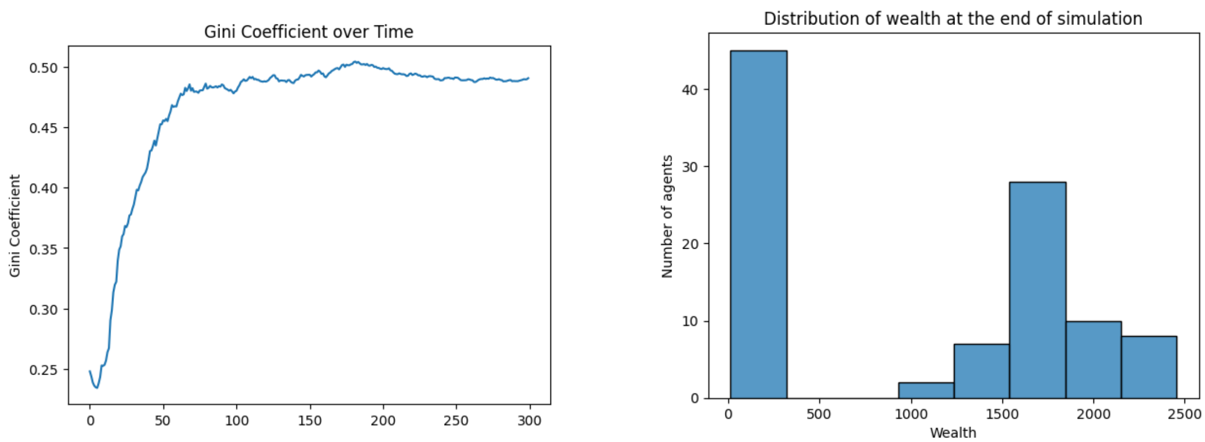

Figure 4: Gini Coefficient for Steps $0 - 300$ and Distribution of Wealth at Step 300 in Discrimination Scenario.

5.4.2 Second Scenario: Policy Intervention to Prevent Discrimination

In this scenario, the company discriminates based on location until a time step k. We set up the simulation so that the company only provides service to high-income neighbourhoods until time step $k = 200$. At this point, a government policy intervention aimed at reducing discrimination forces the company to start providing service to every neighbourhood.

As shown by Figure 5, the difference between agents' wealth in low-income (discriminated) and high-income (non-discriminated) neighbourhoods increases until step 200 and then stays constant. This is expected because we assume that company's service increases agents' life quality, which in turn increases their wealth. The sudden change in the wealth distribution can be seen in Figure 7, since the Gini coefficient starts decreasing after step 200. The impact of a policy shift would probably not be immediately observable in real life scenarios, but our time frames are arbitrary, so we can conclude that the Gini coefficient will eventually start decreasing. We also note that the Gini coefficient at the end of the simulation has fallen below its value at step 0. This suggests that the multiplier effect of the service provided is higher in the low income areas, since the Gini coefficient would flatten out after step 200 if the multiplier effect was the same for every agent. This makes sense considering that agents have diminishing marginal utility, meaning the service is more impactful when an agent's wealth is relatively lower. Perhaps the most interesting finding of this scenario is that the marginal profits of the company keep decreasing for some time after the policy intervention at step 200 but then eventually starts increasing. This suggests that making non-discriminating decisions eventually pays off for the company, even though the marginal profit might not immediately react to the change. Overall, we can conclude that algorithmic discrimination is irrational from both a profit and a social welfare perspective, if the assumption that the wealth inequality interacts with companies' sales holds.

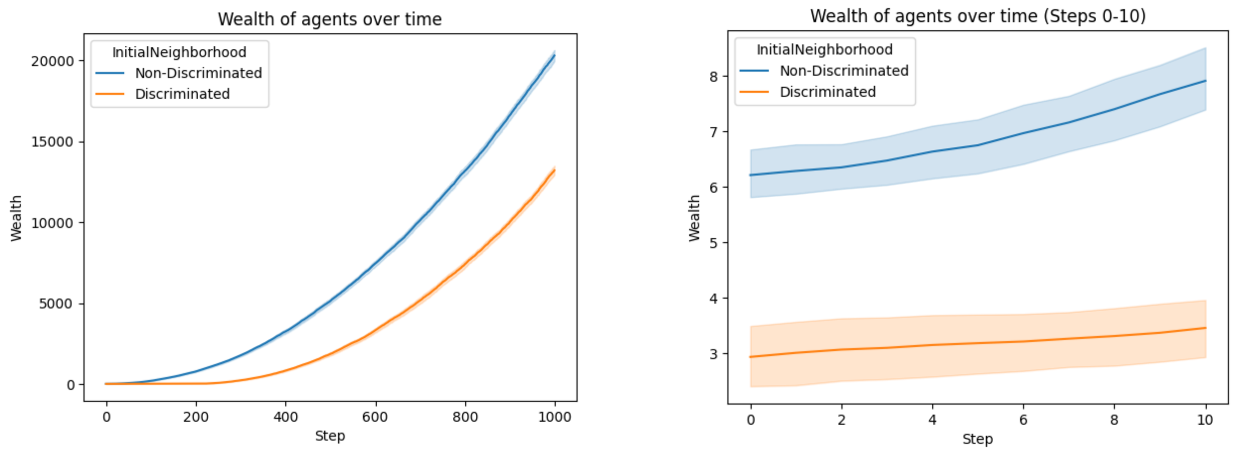

Figure 5: Wealth of Agents in Policy Intervention Scenario.

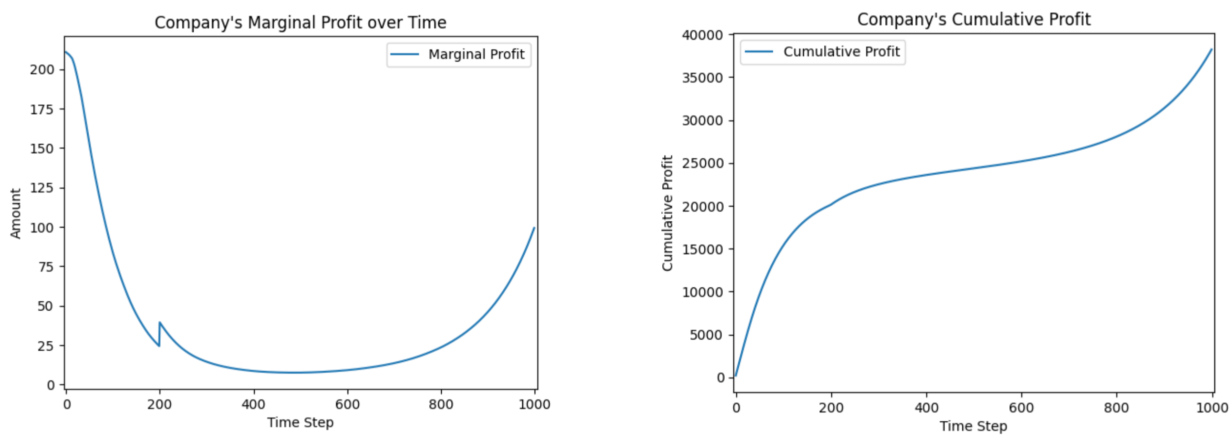

Figure 6: Company's Cumulative and Marginal Profit in Policy Intervention Scenario.

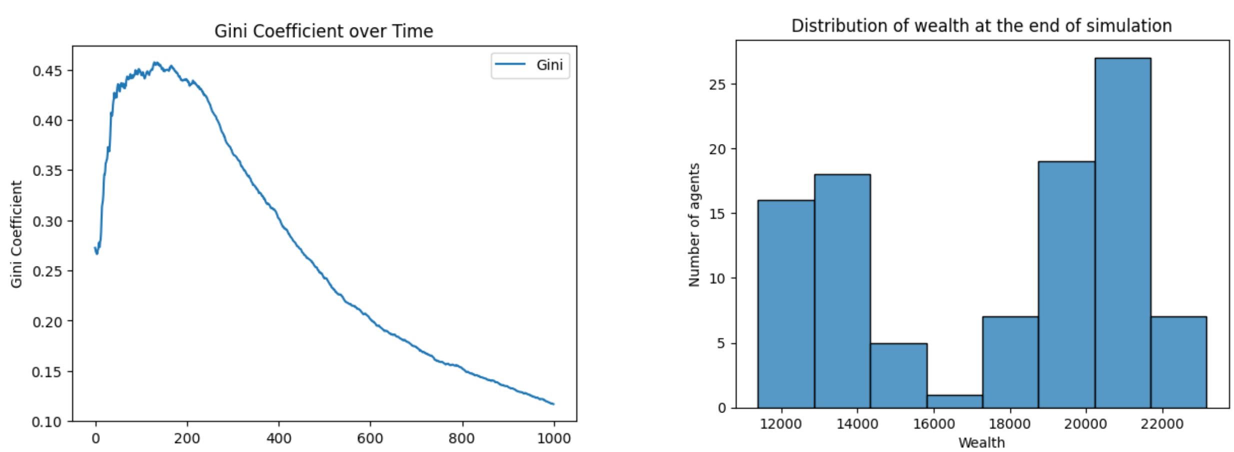

Figure 7: Gini Coefficient for Steps $0-1000$ and Distribution of Wealth at Step 1000 in Policy Intervention Scenario.

5.5 Limitations and Suggestions

There are several points to consider for future studies regarding agent-based modelling of algorithmic discrimination and its societal impacts. Initially, we could have different types of high and low income neighbourhoods in order to further explore the interaction between the company's service and the initial average wealth. Similarly, the simulation could be enhanced by considering other types of discrimination and measuring societal well-being through additional indicators and metrics. This could lead to context-specific results relevant for particular case studies. Additionally, instead of assuming that the wealth inequality directly interacts with the company's profits, a future simulation could be setup that makes assumptions only about individuals' decision-making patterns. This would loosen our assumption of a direct causal effect, and lead to a more refined exploration of the intricate dynamics involved in algorithmic decision-making. Finally, we suggest defining a more dynamic structure for the company based on economies of scale to make the case more realistic.

As detailed in section 2.3, agent-based modelling is a flexible technique that can be tailored to meet various degrees of complexity and specificity. Thus, the simulation outlined in this project can be considered as a starting point, and future work can be done to refine assumptions, introduce additional parameters and agents, and model context-specific scenarios.

6 Conclusion

Overall, our agent-based simulations offer insights into the dynamics of algorithmic decision-making, particularly in the context of discriminatory practices. The first scenario, where the company discriminates based on location, illustrates the temptation of short-term profits. Initial gains are apparent, but the resulting wealth disparity and rising Gini coefficient underscore the hidden long-term costs. It becomes evident that, while discrimination may seem financially rational initially, its effects on societal well-being eventually translate into tangible long-term costs, mirrored in the company's diminishing marginal profits. Conversely, the second scenario, where the policy intervention limits the company's discriminatory practices, exhibits a decrease in the Gini coefficient and eventual increasing trend in the company's profit. This demonstrates that ethical considerations in algorithmic decision-making can lead to positive outcomes for both societal well-being and long-term profitability, and that the two are not mutually exclusive. Eventually, the cost of fairness, represented by a decline in short-term profits, is offsetted by an increase in long-term profits and wealth distribution. Thus, the overarching conclusion obtained from our agent-based model is that discrimination is not only ethically wrong, but it is also an irrational strategy to implement from a profitability perspective.

The simulation, though based on a specific fictional storyline, mirrors real-world challenges, and its results serve as a valuable resource to provide informed commentary and suggestions regarding algorithmic decision-making. Primarily, it's evident that companies relying on data-driven decision-making should embrace a comprehensive strategy that integrates both fairness and profitability. That is, companies should go beyond conventional measures centred around profit, and should actively asses the societal impact of their products. This implies the acknowledgement of a reinforcing relationship between profitability and societal well-being. Most importantly, companies must stop viewing ethical practices and fairness considerations as obstacles for financial success. As demonstrated by the simulation, companies prosper when societal conditions improve. By acknowledging the irrationality of discrimination companies are not merely aligning with ethical considerations; they are strategically positioning themselves to thrive in a landscape where societal well-being and profitability are interconnected.

Ultimately, algorithmic decision-making will impact every aspect of people's lives and it will come to define the very functioning of society. The companies implementing these technological advancements not only have a responsibility to ensure that their algorithms adhere to ethical standards, but also have the opportunity to utilise algorithmic fairness to catalyse societal well-being and their own financial success. All that being said, the role of individuals must not be disregarded, as they have the power to hold companies accountable and ensure the creation of a future where algorithmic decision-making aligns with collective values and well-being for all.

References

[1] Andriy Blokhin. 2021. Utility function definition, example, and calculation. Accessed: 2024-02-20. https://www.investopedia.com/ask/answers/072915/what-utility-function-and-how-it-calculated.asp.

[2] Edmund Chattoe. 2017. Agent-based modeling. (Jan. 2017). https://www.researchgate.net/publication/318495179_Agent-Based_Modeling.

[3] Raymond Dacey. 2003. The s-shaped utility function. *Synthese*, 135, 2, 243–272. Retrieved Feb. 20, 2024 from http://www.jstor.org/stable/20117365.

[4] Izquierdo et al. 2023. Introduction to agent-based modeling. Accessed: 2024-03-01. (2023). https://math.libretexts.org/@go/page/123822.

[5] Kazil J. et al. 2020. Utilizing python for agent-based modeling: the mesa framework. In *Social, Cultural, and Behavioral Modeling*. et al. Thomson R, (Ed.) Springer International Publishing, Cham, 308–317. ISBN: 978-3-030-61255-9. https://mesa.readthedocs.io/en/stable/.

[6] O'Sullivan et al. 2012. Agent-based models – because they're worth it? In *Agent-Based Models of Geographical Systems*. Springer Netherlands, Dordrecht, 109–123. DOI: https://doi.org/10.1007/978-90-481-8927-4_6.

[7] Joe Hasell. 2023. Measuring inequality: what is the gini coefficient? *Our World in Data*. https://ourworldindata.org/what-is-the-gini-coefficient.

[8] H. S. Houthakker. 1960. Additive preferences. *Econometrica*, 28, 2, 244–257. Retrieved Feb. 20, 2024 from http://www.jstor.org/stable/1907719.

[9] David Ingold and Spencer Soper. 2016. Amazon doesn't consider the race of its customers. should it? (Apr. 2016). https://www.bloomberg.com/graphics/2016-amazon-same-day/.

[10] Wander Jager. 2021. Using agent-based modelling to explore behavioural dynamics affecting our climate. 42, 133–139. DOI: https://doi.org/10.1016/j.copsyc.2021.06.024.

[11] Jaffray Jean-Yves. 1989. Linear utility theory for belief functions. *Operations Research Letters*, 8, 107–112. DOI: https://doi.org/10.1016/0167-6377(89)90010-2.

[12] Bruce W. Lamar. 2009. Min-additive utility functions. Retrieved Feb. 20, 2024 from https://www.mitre.org/sites/default/files/pdf/09_0383.pdf.

[13] Malleson N. 2012. Using agent-based models to simulate crime. In *Agent-Based Models of Geographical Systems*. Springer Netherlands, Dordrecht, 411–434. DOI: https://doi.org/10.1007/978-90-481-8927-4_19.

[14] Research Outreach. 2023. Using agent-based modelling to understand social phenomena. (Nov. 2023). https://researchoutreach.org/articles/using-agent-based-modelling-understand-social-phenomena/.

[15] Abebe R. and Kasy M. 2021. Fairness, equality, and power in algorithmic decision-making. DOI: https://dl.acm.org/doi/pdf/10.1145/3442188.3445919.

[16] Oxford Reference. [n. d.] Cobb–douglas function. Accessed: 2024-02-20. https://www.oxfordreference.com/display/10.1093/oi/authority.20110803095620528.

[17] The CORE Econ Team. 2023. Quasi-linear preferences. Accessed: 2024-02-20. https://www.core-econ.org/the-economy/v1/book/text/leibniz-05-04-01.html.

[18] Wilensky U. 1999. Netlogo. Accessed: 2024-02-20. https://ccl.northwestern.edu/netlogo/.