Grau en Dret

Treball de fi de Grau (21067/22747)

Curs acadèmic 2020-2021

DISCRIMINATION IN ARTIFICIAL INTELLIGENCE

Cèlia Roig Salvat

183820

Tutora del treball: Mireia Artigot Golobardes

Universitat Pompeu Fabra Barcelona

DECLARACIÓ D'AUTORIA I ORIGINALITAT

Jo, Cèlia Roig Salvat, certifico que el present treball no ha estat presentat per a l'avaluació de cap altra assignatura, ja sigui en part o en la seva totalitat. Certifico també que el seu contingut és original i que en sóc la única autora, no incloent cap material anteriorment publicat o escrit per altres persones llevat d'aquells casos indicats al llarg del text.

Com a autora de la memòria original d'aquest Treball Fi de Grau autoritzo la UPF a dipositar-la i publicar-la a l'e-Repositori: Repositori Digital de la UPF, http://repositori.upf.edu, o en qualsevol altra plataforma digital creada per o participada per la Universitat, d'accés obert per Internet. Aquesta autorització té caràcter indefinit, gratuït i no exclusiu, és a dir, sóc lliure de publicar-la en qualsevol altre lloc.

Cèlia Roig Salvat

Barcelona, 28 de maig de 2021

Abstract

Artificial Intelligence and algorithms have unlimited prospects to improve human lives but, at the same time, they pose unfathomed risks for human rights protection. Some of these risks have already manifested, as several discrimination claims, especially based on gender and race, have been brought to courts and media, questioning the platform economies using these technologies, and several experimental studies have shown their discriminatory patterns. It is relevant to examine the application of AI to this business model as the network effects it generates can exacerbate the discriminatory potential. This study analyzes how structural discriminations have perpetuated in Uber, one of the companies with more complaints, and explores the reasons behind this situation in order to determine whether the origin is vertical discrimination, due to the configuration of AI exclusively; the aggregation of horizontal discrimination by its users the algorithm enables, or a combination of both. It has been essential to understand the roots of discrimination in AI in order to develop an accurate account of relevant characteristics of the regulatory responses, self-regulation and legislation in the US and the EU, and determine whether they provide good prevention or an adequate response to the issue.

Index

1. Introduction

New technologies present a myriad of new opportunities for society, but also challenges for legislators in order to ensure that all existing rights before the disruption are respected after its application. One of the latest innovations has been Artificial Intelligence (or AI). It presents many legal issues as it has several applications in different sectors which pose new risks that cannot be addressed by traditional legislation. A business model that is closely linked to AI is that of sharing platform economies, as data processing and user behavior prediction are essential to its success, for which algorithms and AI are necessary. Some of the issues have already been addressed by data protection regulations that were passed with the increase in processing capacity of the internet, but they do not cover all possible AI risks.

The protection of human rights is at the forefront of these risks, as well as ensuring legal certainty for operators. Human rights violations, more specifically, discrimination, in platform economies has been pointed out by human rights activists and users who have been harmed by it. Price differences, delay in the service or even exclusion from the market are some of the denounced practices or outcomes of the use of AI in those platforms.

The aim of this project is to determine which is the discriminatory potential of AI and algorithms in its application to platform economies. The novelty of discriminatory practices in AI is that they are more intangible, as discrimination happens without the affected individual knowing who he or she was discriminated by. We try to determine whether they are only perpetuating existing discriminatory practices, they are magnifying them, they are creating new ones, or a combination of these. Moreover, we try to ascertain which dimension of the discrimination arises from horizontal discrimination and which from vertical discrimination. We do so through the analysis of discrimination found in the ride-sharing platform Uber, specifically focusing on gender and race discrimination, the two most salient forms of discrimination. Regulators should find it essential to find the roots of the discrimination in order to provide adequate prevention mechanisms and adequate responsibility structures in response of human rights violations.

Moreover, we study the impact of these discriminatory outcomes, as platform economies can be accessed worldwide, so the effects of discrimination have a potential to reach a lot of people. Because of the importance of the problem, algorithmic bias and discrimination coming from AI has been studied in depth by many scholars, whose research we use as a basis for our analysis.

Furthermore, we try to give some company-specific solutions to the issues and determine whether the legislative solutions given to the AI risks in the US and the EU are really effective in preventing discrimination or giving an adequate response to it. This analysis of legislation is especially relevant because the EU legislative initiative being studied is the proposal on the Artificial Intelligence Act, which was presented on April 2021, so it has not been analyzed in depth by many publications. The US text is the Algorithmic Accountability Act of 2019, which should be further developed by other regulations targeting specific sectors, but this has not been done yet. Through the comparison of self-regulation and legislation and taking in mind the ongoing debate of which is best we analyze which provides the most adequate response for this case.

The methodology used is, therefore, a review and analysis of existing literature on discrimination at a theoretical level, of case studies on discrimination in Uber and of US and EU legislation on the matter. This study starts with a theoretical framework that aims to define discrimination in different contexts, both online and offline. Later we analyze how this discrimination is presented at Uber, focusing on gender and racial discrimination. We try to determine the best policy solutions the company could adopt, taking into account that compliance with human rights is paramount. Afterwards, an analysis of legislative texts is carried out, centered on US and EU solutions, which are slightly different because of their political and legal traditions regarding the prevalence of consumers' or businesses' interest. Finally, we try to debate which approach is better for this specific case, and provide overall conclusions.

2. Theoretical framework

Discrimination is present in all society despite being outlawed by international documents protecting human rights such as the UN Charter, the European Convention on Human Rights and the International Bill on Human Rights (comprised by the Universal Declaration of Human Rights, the International Covenant on Economic, Social and Cultural Rights and the International Covenant on Civil and Political Rights). These texts, many other international organizations' binding norms and national laws do not define specifically what the action of discrimination itself is, but give a *numerus apertus* list of different grounds of discrimination, which generally include gender[1], race, religion and age, among others. As a result, the determination of whether a conduct amounts to discrimination will depend on whether the person given different and worse treatment than others in the same context presents one (or multiple) of the grounds of discrimination. Andrew Altman (2020) differentiates between the moralized and non-moralized concept of discrimination, saying that discrimination can just be a statement of comparison (in the non-moralized sense) or a statement of wrongfulness of a certain act, policy or practice, a wrongfulness that is tied to the membership of the person to a group defined by one of the grounds of discrimination.

In this essay we will focus on discrimination based on race and gender, the most salient forms that we find in the particular platform economy we are analyzing. Discrimination can arise from only one of these characteristics or from a combination with each other or with other grounds of discrimination, which is considered intersectional discrimination. Sexism and racism have been structurally protected throughout history and embedded into law, resulting in long outstanding injustice, as women and people of color have been systematically underrepresented and presented with disproportionately worse conditions to develop themselves in society. It is not the aim of this study to prove or explain the scope of these types of systemic discrimination, as it has been analyzed in depth by many scholars and publications, but it is to analyze some of the new ways it can present itself, taking its ingrainedness in society as a given.

Depending on the relationship between the discriminated individual and the discriminator, discrimination can have horizontal or vertical effects. It is horizontal when it affects private relationships between equal-powered parties, so between natural persons or between private enterprises at the same level (for example, a company does not contract with another one based on the race of their owners). It is vertical when there is an imbalance of power between the two parties, with the discriminator holding a position of power with respect to the natural person being discriminated. The position of power is held in most cases by private administration with respect to citizens and companies with respect to consumers.

---

[1] Instead of “gender”, “sex” is used in many human rights texts, such as the UDHR or the ECHR. However, in this study, “gender” is used because it is a more inclusive term and biases, discrimination and stereotypes affect the social construct (gender) and not only the biological attributes (sex). Moreover, there is an international shift towards the protection of gender as a ground of discrimination. The Yogyakarta Principles in 2006 already provided guidance on that, and now gender is the element considered for many equality policies (for example, it is the 5th UN SDG).

Finally, discrimination, regardless of the ground it is based on and whether it comes from a natural person or an organization or institution, can be direct and indirect (Altman, 2020). It is direct when discrimination is explicit (in the formulation or in the justification of the action) and intentional. It is indirect when it has comparatively worse effects on a group defined by a ground of discrimination despite not having discriminatory motives or not specifically seeking a discriminatory result. The European Court of Justice only accepts three justifications (European Union. Agency for Fundamental Rights, 2018) for direct discrimination: exceptions on the basis of age, genuine occupational requirements and religious institutions. As for indirect discrimination, it is only permitted by law when it has an objective justification, based on a case-by-case analysis of the legitimacy of the aim and the proportionality (appropriateness and necessity) of the treatment. The Human Rights Committee of the United Nations similarly requires objective and reasonable criteria in order not to consider a certain policy discriminatory (Altman, 2020).

In order to analyze which can be the possible origins of AI and algorithm discrimination we first need to understand where discrimination from humans can arise, and in a broad sense, we can distinguish between discrimination arising from bias or consciously. Firstly, a bias, according to the Cambridge Dictionary is "the action of supporting or opposing a particular person or thing in an unfair way, because of allowing personal opinions to influence your judgment". Biases alter rational decision making and can lead to discrimination, as the personal opinions regarding race, gender identity, sexuality, age, religion, etc. of others is what leads to a different treatment. Human biases have been thoroughly studied in psychology and applied to the study of decision making in other fields, such as behavioral economics and political sciences. The origin of such biases has been attributed to constraints on the design of the mind or other irrationalities as well as evolutionary adaptations that can arise for three reasons: as heuristics or shortcuts, as artifacts and as tools for error management (Haselton et al., 2015, p. 969). For example, as heuristics, biases can lead to people choosing a male driver rather than a female driver based on stereotype and not on real performance that leads to rational decision making, resulting in discrimination. Regarding error management bias, Haselton et al (2015, p. 976) identify as a relevant source of discrimination the bias towards false positives in assessing cues of disease as it "may lie at the root of many forms of stigmatization and prejudice, including racism, ageism, homophobia, and anti-fat prejudice". The social exchange bias can lead to increased cooperation due to fear of ostracism (and thus may reduce the effect of discriminatory biases) but it still is not enough to outweigh all other biases and result in no discrimination. This social exchange bias is relevant in platform economies, where scores and reviews' discriminatory tendencies are partially outweighed by fear of ostracism.

However, as mentioned before, discrimination does not only come from bias, but it can also be a conscious choice. Many public policies have been deemed directly or indirectly discriminatory towards a protected group. For example, establishing different periods of maternity and paternity leave is indirectly discriminatory towards women as it burdens them with the care of children. In *Sejdić and Finci v. Bosnia and Herzegovina*, the European Court of Human Rights found that the electoral law determining that only those affiliated to Bosniac, Serb or Croat was discriminatory on grounds of race because "[d]iscrimination on account of a person's ethnic origin is a form of racial discrimination". This conscious decision does not only concern governments, but also applies to private companies (for example, when they have a gender pay gap or they do not hire somebody based on race) and discrimination from individuals.

In these definitions we have given on discrimination, we have focused on discrimination coming from human actions, but the choices stemming from algorithms and Artificial Intelligence, despite their apparent impartiality and rational logic can also be discriminatory. Even when they are configured to be neutral, they mirror human biases or prejudices present in data (Gordon Clausen, 2020), which in turn are a result of human prejudice and conscious discrimination themselves, if we consider that data is a compilation of human choices. We will further develop the origins of discrimination in AI later, but it is necessary to point out that many cases of discrimination from algorithms have already been detected and have become news, such as the ProPublica investigation "Machine Bias" on the racially biased US system used to decide pretrial release or incarceration; the Amazon sexist hiring algorithm (Dastin, 2018); or the Uber and other ride hailing apps discrimination cases we are further studying in this paper.

In order to better understand how discrimination in AI and algorithms works it is first paramount to give some general definitions of some technological concepts that will be repeated throughout this study. An algorithm is a computational procedure used to solve a problem and it produces an automated decision. Artificial intelligence or AI, on the other hand, is the capacity of machines to make autonomous decisions and learn without explicit programming, simulating human intelligence and decision-making capacity (but it still does not have the conscience of human decision making). There are two main branches in AI (Barnett et al., 2017): machine learning and knowledge-based systems. Machine learning, which has the ability to learn without explicit programming, can infer a solution from the new data it receives. It is possible thanks to deep learning (Hargrave, 2021), a function that imitates the workings human data processing and pattern creation processes for decision making. Knowledge-based systems have the ability to reason, and their knowledge is represented as ontologies or rules rather than via code.

Mireille Hildebrandt (2020, p. 76) points out that it is essential to recognize the agency that these technologies have, defined as the "ability to perceive an environment in terms of actionability, coupled with the ability to act on the world". However, despite having agency and being able to learn without being explicitly programmed to achieve a certain result, ultimately these algorithms and AI rely on code, which is designed by humans. In order to establish proper regulations and limitations to these technologies, a good understanding of this is essential to determine responsibility in cases such as the ones studied in this project, where the algorithm and AI present biases or result in discrimination.

AI, while apparently being neutral and rational, cannot escape the objection of being discriminatory. This discrimination can come from, first of all, the biases present in the people who code, which are translated into the algorithm, and therefore algorithms follow a decision structure that ends up reproducing the biases these people have. Mullainathan (2019) talks about label bias, which stands for the effect that determining one objective or another for the function of an algorithm has on the final result. This means that the framing of the algorithm itself can be biased. This research is in line with findings of Cowgill et al.'s (2020, p. 22-23) study that points out that engineers' demographics play a role in the bias that is found in algorithms, despite not being its main source. More homogenic demographics (in terms of race, gender, age, educational background…) lead to more like-minded individuals, who are less likely to think of the several implications that has determining a certain objective in a certain way.

Secondly, it can come from algorithmic prejudice (Datta, 2021). Even if the factor that leads to discrimination is not specifically built in the algorithm (for example, gender is not an input), all the other information that the algorithm feeds off of has characteristics that are correlated to or can infer the gender, race, age…, so the result ends up being discriminatory because, on aggregate, the information it is given can point to the discriminatory characteristic. This is because historical power imbalances affecting protected groups have current effects on some indicators of said protected groups, such as education level or neighborhood where they live. Aside from the characteristics of the information, biases in AI can also come from lack of complete data due to the selected inputs, as the model is based on a non-complete version of reality (Kantarci, 2021).

Thirdly, from negative legacy (Datta, 2021): The AI, despite having been corrected for such biases and presenting an objective code, in the case of machine learning, learns from data which represents society, and, as society's patterns of conduct are biased, AI will be so as well and can help exacerbate the discrimination resulting from the bias. People using the algorithm also have discriminatory tendencies, whose choices end up getting built in the data the AI learns from.

This becomes heightened in the case of platform economies that use AI as the bases of their services, as they present all types of biases. First of all, the algorithms can reproduce the bias of the person or people who designed them or learn from data that includes discrimination, and users of the algorithm, with the choices they make, incorporate further bias and discrimination in the AI system. Moreover, algorithms and Artificial Intelligence sometimes even discriminate in ways people would not, as researcher Latanya Sweeney (2013, p. 34) found when studying the difference in Google Ads appearing when searching traditionally white versus traditionally black names.

As Mullainathan (2019) argues, the algorithm is not the only discrimination source, but it signals that previous discrimination exists. Then the issue becomes how to hold algorithms and AI to the same standards that companies' and governments' policies are held to. Why should a company's or government's hiring process be consistent with human rights if it doesn't use algorithms and not be required the same if it decides to rely on them or AI for selection? And the same can be said for all other aspects of operation, which points to asking for their liability when they use AI as well.

A step further than requiring AI not to be discriminatory and be respectful of human rights is asking for algorithmic transparency. This is easier to ask from public administration uses of AI and algorithms, as Andrés Boix (2020, p. 262) does, by saying that algorithms become regulations used for decision making and, as all other decision-making processes expressed in regulations of the public administration, they must be held to transparency standards and duly published. On the argumentation, O'Neill (2017), in *Weapons of Mass Destruction*, thinks that systems that have an impact in our lives should be open and available to the population. Some other authors advocate for a more limited transparency, which would consist of disclosing the fact that a decision is made by an algorithm, but not necessarily the characteristics of the algorithm itself.

When private companies are concerned, the debate on transparency, and how much of it there should be, becomes a balance of two interests: the protection of human rights when there are potential violations on one side and the protection of intellectual property and trust in companies processes on the other (Hosanagar & Jair, 2018). There are methods of accountability other than transparency, such as external audits, and some say that transparency should be limited to cases that result in serious problems because of the problems it can cause. Moreover, transparency does not solve all issues, as Hosanagar and Jair (2018) explain, because machine learning presents very little code and makes decisions based on the data it is fed, so analyzing a decision becomes a complex analysis of large amounts of data. Furthermore, we need to consider that programmers have already developed tools for testing biases in AI before implementing it, in order to correct for possible discriminatory outcomes. A solution prior to transparency that would not compromise business secrets and intellectual property could be establishing mandatory tests for AI biases.

Still, in companies, such as Uber, whose scope and reach is larger than even a government's and, in consequence, its possible impact on fundamental rights is also huge, we must ask ourselves to which extent protection of human rights prevails over the intellectual property rights of a company. The network effects[2] that characterize sharing-economy platforms generate an increasing aggregate effect for all the inputs the platforms receive, thus increase the discrimination potential. In ride hailing companies, we can find positive (the more drivers there are, the more likely a passenger is to get a ride, but also the more passengers there are, the more likely it is for drivers to get income) and negative network effects (at a specific time, too many passengers will lead to longer wait times and increased prices), but they are overall positive, as shows the growth of the company.

The aggregate effect can even lead to barriers of entry of certain groups of people, as they may self-exclude from the community when discovering such discrimination is an issue (both becoming a driver or using the service) if they can feel that their chance of success is lower. When multiple infractions of human rights have been pointed out for a same company or business model (platform economies, and specifically, ride hailing apps), the argument for regulation becomes stronger.

---

[2] A network effect is the economic phenomenon by which the value of a good or service is conditioned by other's preferences. It is positive when the more users or buyers the service or product has, the more utility or value a user derives from it, so the more valuable it is overall. It is negative when the value of the product or service decreases when more people use it.

3. Discrimination in Uber's technology

Despite the fact that platform economies are said to be more democratic and inclusive than traditional business models, they also present challenges regarding the perpetuation of gender (Schoenbaum, 2016) and race (Edelman & Luca, 2014, p. 8) discrimination, as we can see in Uber, which has been accused and found guilty of several discrimination claims. Moreover, there are several empirical studies that have shown, mostly in US cities, that gender and race (especially towards African Americans) discrimination does take place in ride hailing apps.

Before entering into more detail, we would like to point out why do we consider the choice of sector and company to be a relevant issue. First of all, ride hailing companies are considered to be sharing economy platforms, which are based on trust. In traditional companies, trust in the company, and therefore reputation, is one of the key drivers of sales. This trust, in the case of sharing economies, relies not only on the company's reputation, but also (and more heavily) on members of the community's intimacy (Schoenbaum, 2016, p. 1029), which highlights their identity, including race and gender, to reduce the transaction risk. In the ride hailing sector, this trust is placed between drivers and customers, and, as they are strangers, it is channeled through showing some of the counterpart's characteristics and through the rating system, which indicates the trust the other members of the community placed on a certain individual. This focuses the transaction on a person's traits in two ways: first by showing the name (and sometimes even image) in order to reduce the feeling of transacting with a complete stranger about whom you do not have any references, and secondly by aggregating other people's rating of that stranger, a decision which has, of course, been informed by the person's traits and therefore been influenced by bias and discrimination. This individualization of the counterpart increases the prominence of biases and discrimination both on rating (which produces an aggregate discriminatory effect) and the mere acceptance or cancellation of the ride at an individual level after seeing or inferring from the name or picture the other person's gender or race. All of this results in an increase of salience of horizontal discrimination, which is the one occurring between users. Moreover, as the companies make decisions based on discriminatory data knowingly, they perpetrate vertical discrimination.

Second of all, the company choice is relevant due to Uber's presence in the market. Despite several bans in European countries or cities, Uber is present in more than 10,000 cities in the world (Uber, n.d.), and some of these European bans are temporary. This ubiquity produces relevant network effects, thus creating a high discriminatory potential and a threat to human rights.

Some of the gender and race discrimination claims that have already surfaced are related to their corporate policies, but many are tied to the rating system Uber uses to give preference to drivers and determine whether a driver can continue providing services for Uber or a consumer can use the platform. This rating system, where drivers and customers rate each other on a 5-point basis, is built in the algorithm that determines which drivers get more rides and repercusses on how long customers have to wait as well. The user's score also determines whether he or she can participate in the market, as low scores lead to account deactivation (in the case of drivers it is a score lower than 4.6), so it is also used by the company as a mechanism of quality control (Rosenblat et al. 2017, p. 3 and 17). We focus on its main service, which is providing car rides to passengers, and not on other ones such as food delivery, motorcycle rental or freight transportation.

Uber's algorithm predicts supply and demand and determines fares accordingly. It is not transparent, in fact, drivers are suing Uber to bring to light how it works (Daws, 2020), but, on general terms we try to explain how the system is set up: a customer, through the Uber app in which he or she has a profile, requests a ride at a certain geographical location for a certain destination and at a certain time (immediate pick up or at a scheduled time, which is available only in some areas).Then, the algorithm assigns the ride to a specific driver taking into account factors such as distance from the pick-up point and ratings. If that driver cancels, another one is assigned that ride.

Due to the importance of the algorithm and the app, the Court of Justice of the European Union ruled in its decision on C-434/14, that the drivers would not be in conditions of providing any service were it not for the app, so Uber is not a mere enabler of the service, but has an underlying activity of passenger transportation, which it organizes with decisive influence on the vehicles' quality, the drivers conduct and the setting of the price. As such, Uber is an employer of its drivers (which is what was relevant in that ruling), and we may add that, therefore, the company has a responsibility on the transaction and its potential human rights violations.

This part is focused on gender and racial discrimination arising from the algorithm exclusively due to the limited scope of the project, however, other violations of human rights can be found and are prone to appear with the increased presence of AI, such is the case of self-driving cars, that can lead to harm to humans, violating the rights to life and dignity. Moreover, Uber has been brought to courts for other types of discrimination, which we do not cover because they are not related to the algorithm itself but due to company policy, for example, on grounds of disability for having no (or little) accessibility in its vehicles for people with disabilities (Ciechalski, 2017).

Uber's technology has been found to discriminate against both users and employees, and we focus on both groups in the study of gender and racial discrimination. We have considered such violations separately, but it is necessary to highlight the intersectional approach as well, which explains social relations and the appearance of discrimination towards a person through the layering of his or her different identity markers (in which race and gender are included). It is necessary to study how discrimination takes place in order to provide effective regulatory solutions and know whether they can be applied to other industries in AI.

3.1. Gender discrimination

Issues with gender discrimination are present in ride hailing inner workings in other ways than the safety concerns and presence of sexual harassment that are tied to the intimacy of the transaction (which are outside of the scope of this paper), and they affect both women consumers and drivers.

Non-compliance with stereotypes is at the root of the discrimination in some cases, as women passengers have complained about having lower ratings due to not engaging in conversation. Women are perceived as being more talkative than men, and in the rating system scenario, they have denounced (Waters, 2018) that they are punished with lower points when they do not want to talk to the driver, a behavior that does not affect male counterparts' score. This lower score as a customer means lower prioritization of the ride to drivers, therefore having to wait longer for rides.

A similar phenomenon has been called out by female drivers, who have been given very low scores after rejecting unwanted advances or flirtatious comments from male passengers or not accepting their friendship requests on Facebook (Lee, 2019). Low scores, in the case of drivers, can lead to a reduction in their benefits or even to a job loss, because it is company policy to deactivate accounts with a score lower than 4.6, as in principle, it indicates poor service.

Without the algorithm and the rating system this type of discrimination that can ultimately lead to women passengers' and drivers' removal of the market would not exist. The sexist behavior would still persist, as it does in traditional taxi service and in all areas of society, but it would not have this double effect on women of first having to experience the discriminatory behavior and later being subject to penalization. This is one example of new types of discrimination that can arise with the use of algorithms and AI, and that also applies to race discrimination, as we will show.

Another gender discrimination found in Uber is that female passengers are taken for apparently longer, more expensive rides. In an experiment, Ge et al. (2020, p. 8) found that Uber drivers start the trip before picking up the female passenger or end the trip a while after having dropped her off, which explains why, with the same pick up and drop off destinations as males, female passengers have a record of a longer and more expensive drive than the male participants in the experiment. This is, of course, enabled by the use of apps to make the payment, as the app intermediates between the driver and the customer when making the money transaction, which does not need to happen face to face as with traditional taxi. This allows drivers to set the end of the trip after the trip has actually ended. This is another example of discrimination that did not exist and that has appeared with the use of AI. It is not because of the objective function of the algorithm at all, but a translation of human discrimination that is made easier with the use of the app.

3.2. Racial discrimination

The rating system also has a discriminatory outcome when we consider race factors, especially on drivers. The high threshold for deactivating drivers' accounts is sensitive to race, as passengers are likely to discriminate on the basis of race when rating the drivers, and Uber is aware of it. The European Court of Justice ruled in favor of Uber being considered an employer, so this implies that their firing system is based on discriminatory indicators, however, the labor implications of account deactivation are out of the scope of this paper.

Uber says it uses ratings to protect drivers from low scoring customers, but this also incorporates racial discrimination for passengers because drivers are also likely to discriminate against them, as drivers are members of society in which there is a structural race discrimination problem. This structural racism has been shown for online marketplaces by several studies in which lower offers and lower prices are received by black sellers (Doleac & Stein, 2010; Pope & Sydnor, 2011; Ayres et al., 2015). For individual passengers, the discrimination arising from low ratings may have less impact than on individual drivers, as drivers lose an important source or all of their income; however, on aggregate, the discrimination presents important effects for passengers and for drivers.

Racial discrimination towards customers is present in ride hailing apps despite presenting less discrimination with respect to traditional taxi, as shows the study conducted in LA by Anne E Brown (2019, p. 10), focused on waiting times and cancellation rates, which decrease compared to non-internet solutions. The author points to three factors that explain this difference: "cashless payment, driver and passenger ratings and instant reporting that increases driver accountability". This shows that, overall, AI and algorithms may help reduce discrimination with respect to traditional decision-making outcomes by reducing transaction costs, but they are still perpetuating it and even creating new forms that may offset the reduction on traditional discrimination.

Other studies find significant race discrimination enabled by the technology. Ge et al (2020, p.7) noticed that the possibility of drivers accepting a ride and then cancelling it when seeing the name of the passenger[3] more than doubles in Uber for Black passengers (with a "black-sounding" name) compared to the average. This is a significant finding because it implies that the driver would have otherwise accepted to provide the service but due to the (inferred) race of the passenger, they cancel, which shows very clearly a direct horizontal discriminatory practice. This leads to longer wait times as, until a ride is in fact accepted, there is a higher chance of there having been a previous cancellation of a driver that was nearer (as Uber assigns the ride to the driver that is closest to the pick-up point), so the driver that actually provides the service is further away and this prolongs the overall process.

Another type of racial discrimination is related to pricing. Uber's algorithm determines fares according to supply and demand, among many other factors, which are unknown and may explain the differences found in some studies. A study by Pandey and Caliskan (2021) of over 100 million sample trips found that a higher price per mile was charged if the pickup or destination points are predominantly non-white and low-income neighborhoods. They found this can be explained by trip duration but not by lower demand, which would provide a good justification for a higher price. This is a case of algorithmic prejudice as, while the passenger's race may not be included in the algorithm as a factor[4], environmental factors that are indeed inputs in the algorithm indicate race due to historical discriminatory policies. This happens with neighborhood, that serves as a race

[3] In Uber, the information on name is shown only once the driver has accepted the ride.
[4] We can assume it is not, despite the Uber algorithm not being public, because if race was a factor that led to a higher fare it would be direct discrimination, and we highly doubt that is the case due to the current climate of corporate social responsibility in tech companies.

indicator due to segregation, which was legalized with Jim Crow in the US since 1877, lasted during decades and has long lasting effects up to this day. In Europe, the European Commission – Joint Research Center (2019) has found an increase in segregation during the last 20 years, with race being one of the factors that is considered among others such as income, education and gender.

4. Regulatory solutions

Given the prominence of discrimination in AI, there is a debate on whether it is necessary to legislate on it and its use by companies or self-regulation by those same companies is enough. We believe the best solution is a combination of both, due to the complexity of these technologies. Companies have an insight on which specific changes are needed to redress violations of human rights the platforms may cause, which the policy makers cannot synthesize and generalize enough to be effective. Nonetheless, public guidelines on best practices, strong testing and accountability requirements are necessary to provide protection to users and incentives for companies to develop robust compliance programs. In what follows we analyze the regulatory techniques each type of solution can provide and determine their efficacy in solving the problems AI poses.

4.1. Self-regulation: Company-specific solutions

The use of algorithms and AI in Uber and other ride hailing apps is perpetuating race and gender discrimination, as it creates new forms in which it can appear and does not make traditional forms of discrimination disappear. Even if in some cases these platforms have less discriminatory results in comparison to traditional methods, they can lead to overall worse effects, because the scale in which platform economies are operating is much larger than traditional companies. The reach of the companies, combined with the use of algorithms that feed from large quantities of data from all over the world results in a magnification of the effect of the discrimination, both horizontal and vertical, because a discriminatory behavior does not only affect that particular situation but also future situations that take the first one into account to make a prediction. This phenomenon would fall under the category of negative legacy discrimination of algorithms that was presented in the theoretical framework, as it uses data with discriminatory patterns to make predictions.

Companies should take a proactive approach and they should strive to have the utmost respect for human rights in line with their importance as an actor in society. They should not be able to shield themselves from responsibility by saying that most of the discrimination happens horizontally and they are not liable for what users input to their platform, as companies, through processing this input from users, end up magnifying the discriminatory effect. Moreover, they are aware that their algorithms produce those results, and they are all not stemming from horizontal relationships, but also appear vertically, such as the discrimination in pricing. Therefore, they should be responsible of violations of human rights that are happening through their channel and because of it.

To address the three main discrimination areas from a self-regulatory point of view, we propose company-specific solutions for each problem Uber faces that balance the protection of human rights and non-discrimination with the respect for intellectual property rights.

Uber has already changed the way information appears on the platform to reduce race-indicating or gender indicating features; now the name of the passenger only appears after the ride has been accepted and the picture does not even appear. However, the problem has not been solved as, after the acceptance of the ride, cancellations become an issue which could be addressed by the elimination of even names, so that users only got to know the identity of the other party at the pickup location. Then, cancellation costs would be higher, as more time and effort would have been invested for that specific transaction to occur. In order to have little or no effect on the efficiency of pickups, some mechanisms could be established. Passengers can identify the driver through the license plate number when he or she arrives at the agreed spot, as they can already do now. For crowded pickup spots, in order to have a fast recognition that can work bilaterally, a technique which was only tested could be fully implemented: Uber launched an experiment consisting of installing a small color changing fluorescent at the front of the car and allowing the passenger to choose a color through the app while waiting, so when the car arrives, the fluorescent lights up with the color the passenger chose and he or she is able to identify it more easily. If this is not deemed a viable solution, we call on companies to increase funding on R&D, as discrimination is a human rights violation and companies' should be the ones investing in new methods that allow them to be competitive in the market while respecting human rights.

The de-personification of profiles may lead to a decrease in the interpersonal trust, which, as we said, is one of the basis of peer-to-peer economies, but it can be corrected by focusing the trust on the rating rather than on gender and race indicating characteristics. If we adopt these kinds of solutions, discrimination in the rating system must be addressed. Some authors have pointed to increasing the number of questions that are asked after a low score as a correcting mechanism of this low score for both drivers and passengers. We do not believe that is an optimal solution, because quickness is a relevant factor in the competitivity of an app, and a lengthy evaluation may lead to abandonment of the rating process overall and a switch to other companies, as it is very easy to download a competitor's app. Instead, a solution could be creating an adjustment mechanism within the algorithm that detects which users emit biased ratings by studying the rating pattern of said user and then correcting the ratings accordingly in order to have a more accurate measure of performance. In order to do this, Uber should collect demographics data of its users (gender, race, age…) for internal use to infer the biases that a person presents.

As for racial discrimination in pricing, a corrective mechanism could also be applied. It could start with an audit to detect which are the inputs of the algorithm that create the difference in pricing and give them a different weigh to have a proactive approach on reducing discrimination. An explicit objective of nondiscrimination should be set up in order to correct the discrimination that arose as a result of the price determination by the algorithm. This correction is what Mullainathan and others call the "disparate benefits from improved prediction", which consist of the relative improvement due to detecting biases in AI and correcting them (Manyika et al., 2019).

For gender discrimination in pricing, as the discrimination stems from human action (drivers establishing longer routes), a double-check system in which both the driver and the passenger confirm the start and end of the route could be put in place.

4.2. Public intervention: Legal Solutions

Private sector incentives may not be strong enough to foster change, so public involvement with legislation drafting is essential. We focus now on which solutions are on the table regarding AI and algorithmic bias and discrimination, both in the European Union and the United States. We analyze whether they sufficiently protect human rights and respond correctly to the challenges that AI poses in the interpretation and application of human rights and anti-discrimination legislation to determine which solution is better for the case we are studying and which are their shortcomings.

We start with the response in the United States, which is relevant because it is where many platform economy companies are created and have the headquarters where AI decisions are taken. The latest relevant legislation on the topic is the Algorithmic Accountability Act of 2019. It gives the Federal Trade Commission (FTC) power to issue regulations which require entities that treat personal information to conduct impact assessments of highly sensitive automated decision systems. These regulations would work together with other broader data protection legislation in the US, which is also applicable to these companies to protect privacy. There is no federal legislation that encompasses all the principal rules on data protection, rather there are several federal and state legislations that specify guidelines depending on the sector.

The regulations by the FTC under the Algorithmic Accountability Act would have a federal reach and should require companies to study the algorithms they use, if possible, with consultation of external third parties; identify bias in the new and existing systems they use and fix any discrimination or bias they find. These studies do not have to be made public, so transparency is not a requirement but an option.

This legislation identifies high-risk systems, differentiating between information systems and automated decision system, in order to establish assessment requirements. Uber's system would qualify as a high-risk automated decision system because it can be qualified under section 2(7)[5] as it is capable of contributing to biased or discriminatory decisions impacting consumers, which we have previously shown; as it can facilitate human decision making by predicting movements and behavior that can significantly impact consumers, and as it involves personal information about the identity of consumers. It is a very consumer-centric Act because it confers regulatory powers to the FTC, whose mission is to protect consumers, however, in Uber's case it would also imply protection for drivers, as the algorithm that would be audited affects both.

Aside from falling under the high-risk systems category that is singled out by the Act, there are further requirements. The companies that are affected by this legislation are those that have an average revenue of more than 50 million US dollars during the last three years or control information of more than one million users or devices, so Uber qualifies because it well passes both numbers, as in 2020, the revenue in the US and Canada Region was 6.8 billion $ (Iqbal, 2021)

---

[5] Section 2(7) of the Algorithmic Accountability Act of 2019 can be found in Annex I.

and it had 93 million users. These size requirements are set up in order to tackle the largest firms, which has been criticized by some that say that all companies using AI are susceptible of affecting consumer rights (New, 2019); however, as we have mentioned before, the largest companies are the ones that are more potentially harmful due to the network effects they generate which increase aggregate effects, so it is in line with human rights protection that they are held to a higher standard. This does not imply that size of the company exonerates the smaller firms from complying with non-discrimination regulations, however, taking into account the complexity of these audits and the novelty of the technologies, targeting first the largest companies is a natural first step towards eliminating algorithmic bias and discrimination.

The European Union's latest initiative mirrors some of what we have highlighted of the Algorithmic Accountability Act but is much more detailed and presents some differences with it. The regulatory framework proposal on an Artificial Intelligence Act was presented on April 21st, 2021, prompted by the recognition of the insufficiency of the current legislation with respect to new risks created by Artificial Intelligence (European Commission, 2021). In it, the Commission tries to promote trust in AI by protecting fundamental rights in it while at the same time fostering innovation and bringing Europe's economy to the forefront. It does so through three inter-related legal initiatives: a general framework for AI to address fundamental rights and safety risks, EU rules to address liability concerns related to AI and a revision of some safety legislation. The latter two have not been presented yet, but the framework centered around fundamental rights has and it is the one of interest to this project.

Paired with this proposal, the General Data Protection Regulation also has to be taken into account as the basis of accountability for data collection and treatment by the companies. Processing data can only be done tied to the identity of the subject when it is necessary, which should be taken into account when implementing the oversight measures the proposal contemplates. Moreover, in its article 22.1, the GDPR protects individuals from decisions taken by automated processing that have significant impact on their lives, giving them the right not to be subject to such decision.

After having analyzed the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), it is relevant to point out that the Commission wants to especially target high risk AI cases, which will require a thorough evaluation before being used in the market. The Commission responds to such high-risk cases with a proportionate and risk-based approach. Non-high-risk cases are not subject to such strict requirements and will only be subject to minimal transparency, and unacceptable risk systems are completely banned (article 5 of the Act). The Commission has deviated from the Algorithmic Accountability Act in that it does not consider size of the company that is using the AI to be important, but rather the risk it poses according to its definition of risk established in Annex III[6] of the Proposal. The proposal specifically states that it aims to protect "non-discrimination (Article 21 of the Charter) and equality between women and men (Article 23 of the Charter)" in a proportionate manner, and it balances these rights with rights of companies. The proposal limits freedom to conduct business only in those situations where there is an overriding public interest, in which the Explanatory Memorandum to the proposal includes consumer protection. Moreover, transparency obligations are also balanced with intellectual property rights, so that companies only have to share the minimum information for people to be able to exercise their rights.

Uber's algorithm could be classified as a high-risk one as per Annex III article 4.b)'s classification of high-risk systems, referred to by article 6.2 of the Proposal. This classification deems AI used for making work-related managerial and contractual decisions as high risk. The European Court of Justice's judgement from 20 December 2017 on Case C-434/15, *Asociación Profesional Élite Taxi v Uber Systems Spain SL*, which ruled Uber was a transportation provider, not a mere intermediary, is relevant because it determines that the company is an employer, not an enabler of transactions, so Annex 3, article 4.b) can be applied to the company without hesitation. Consumer-related AI is not protected by this high-risk classification, but in Uber's case the system that decides labor relations is the same that affects consumer discrimination, so when the requirements of the Commission's proposal for high-risk are met for employees, they are likely to have a collateral improvement for consumers' standards. Moreover, consumers are protected by the transparency obligation of Title IV, which compels providers to inform natural persons "that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use".

The implications for high-risk AI systems is the compliance with requirements set in Title III Chapter 2, which include a continuously updated risk management system with testing procedures, the presentation of technical documentation before putting the product in the market, record-

[6] Annex III's article 4 on high-risk classification, of the EU Proposal on an Artificial Intelligence Act can be found in Annex II.

keeping to be able to track possible incidents, transparency to users[7] so that they are able to interpret the outcome of using said AI system, human oversight either before using the system and putting it to the market or built into the system when possible, and the highest standards of robustness, accuracy and cybersecurity. As one of the included risks to take into account in risk management is the violation of fundamental rights, the European Union's approach for Uber would imply both testing before putting AI in the market (for example, in case changes are introduced) and transparency for users afterwards, which needs to include the information specified in article 13.

In case a company does not comply with the requirements of the Regulation or provides incorrect, incomplete or misleading information, it is subject, according to article 71, to fines that can vary, depending on the article breached, between ten and thirty million euros or 2% and 6% of total worldwide annual turnover, whichever is higher. These are substantial fines, which should prompt companies to ensure the algorithms they are using do not generate discrimination or any other human rights violations.

We can see that both legal norms tackle AI discrimination, but they do so from a different perspective. In the Uber case, the EU proposal would impose high-risk obligations due to workers' protection (as it qualifies AI affecting labor relations as high-risk) and the US Algorithmic Accountability Act would do so due to consumers' protection. This does not imply that the EU leaves consumers unprotected, as it imposes some requirements and codes of conduct as well for non high-risk AI, transparency requirements for AI that interacts with natural persons, and one of its main aims is the protection of human rights.

Both legislations take into account protection from high-risk AI in a proportionate manner in relation to freedom to conduct business and the protection of intellectual property. Affections to these rights are justified by public interest and they are expressed in the form of required tests and audits to ensure non-violation of human rights and voluntary transparency in the case of the US or strict tests and risk management systems but very limited mandatory transparency in the case of the EU (which is achieved with a combination of the Proposal of AI Act and GDPR).

---

[7] Users are defined by article 3 of the EU Proposal on an Artificial Intelligence Act as natural or legal persons using AI systems, except when AI is used in a personal non-professional activity, so it does not target consumers.

However, none of the regulatory solutions provides a concrete definition of what AI discrimination can be or which are its roots in the algorithm. It is up to the providers to conduct the tests of whether there is discrimination, and third parties' collaboration is only encouraged (in the case of the US) and can be opted out of (in the case of the EU). Therefore, the providers themselves determine which is their standard of risk elimination and they are given a wide margin of appreciation. While the complexity of these systems justifies in part this self-testing, because establishing standard parameters would not lead to a satisfactory result, it still leaves room for new forms of algorithmic discrimination to appear and not be detected until harm has already been done.

The AI environment is fast paced and, together with its capacity to process large amounts of data, scalability and the network effects it creates with some uses (such as Uber's), its discriminatory potential is far-reaching and unknown. In none of the Acts we see a specific solution for the different roots of AI discrimination we presented in the theoretical framework, so the legislation does not answer which approach would be best for discrimination arising from apparently neutral algorithms which can pass tests and audits that are conducted with partial data.

Algorithmic transparency is not the solution either. It has many proponents, as we explain in the theoretical framework, however, many critical voices have started appearing and calling out the problems it presents. New and Castro (2018, p. 10) say that calling for transparency and explainability of decisions algorithms make in order to eliminate biases, thus holding AI to a higher standard than human decision making would be unreasonable; however, we believe that, precisely due to the fact that biases and discriminatory patterns in AI can be detected through a study of these patterns and have a fixable nature, unlike human decision making, we should strive to eliminate them in order to respect fundamental human rights. Our response is not a call for transparency, because it does not inherently eliminate biases, but for accountability at the root of the problem.

Other criticisms to transparency, ones we share, are that in most cases transparency itself is not beneficial or necessary to solve the underlying discrimination problem, and it leads to presenting very difficult, complex or costly to interpret information, which most people cannot do and control organizations would not have enough resources to do. For example, by analyzing how the Uber algorithm works it would probably be difficult to determine that the 4.6/5 rate cutoff is discriminatory, and it is only analyzing the outcome when we can say there is discrimination, so transparency in itself is not always helpful. The requirements for implementing a measure (transparency) that affects a right (intellectual property and proprietary information rights) are having proportionality between the intrusion and the level of protection of another right, and for the measure taken to be the least restrictive and most effective of all possible solutions, which would not be the case, as effectiveness is not ensured. The key is not to rule out transparency overall, but to demand information to be presented in understandable ways so that it is useful to users. For example, information could be useful to Uber's drivers, so they can understand why their ratings are lower. However, depending on the way this information is presented, it may lead bad players to game the system, so the company has to be mindful of what is shared and how.

If we take all this into account, transparency may not be the most adequate response to algorithmic discrimination, because in many cases it does not solve the problem due to interpretation difficulties and there are less restrictive and more effective solutions, such as accountability mechanisms. Transparency implies that the inputs and the decision process is known to people, but it does not necessarily mean that the outcome will be non-discriminatory. Accountability, instead, makes sure that the provider is responsible for the outcome (Kossow et al., 2021).

Algorithmic accountability consists of ensuring that harms can be "assessed, controlled and redressed" and that "laws that apply to human decisions can be effectively applied to algorithmic decisions" (New and Castro, 2018). This is precisely what the audits and identification of biases in the Algorithmic Accountability Act in the US and the evaluation required in the Proposal for the Artificial Intelligence Act in the EU do, although they combine this solution with some transparency. Desai and Kroll (2017, p. 9-10) point out flaws of the testing solution, as many different inputs to the algorithm would need to be tested and auditing can only test a small part of them, so the difficulty is transferred to deciding which of those inputs need to be tested in order to see if discriminatory outputs are produced. However, this accountability approach transfers the responsibility of testing to the provider and puts the burden on them to make sure no discrimination or harmful outcomes are produced, or at least, the risk of them is minimal.

Still, full accountability for discriminatory outcomes is not achieved, as the EU and US Acts are tilted towards an obligation of means through thorough testing and safety requirements, but not an obligation of non-discriminatory results in the application. Both texts have loopholes regarding which is the appropriate response when discrimination has taken place despite having complied with the testing, robustness and cybersecurity analysis, so this remains a challenge for legislators.

5. Conclusions

Artificial Intelligence and algorithms have many applications that can improve human lives across the planet in unprecedented ways thanks to the increased efficiency in processes and productivity, the elimination of some arduous tasks, the facilitation of knowledge-sharing, its usage in medical procedures, and some other applications that have not even been discovered yet. However, despite its benefits, AI and algorithms also present some risks to human rights and social interaction, many of which cannot be covered by traditional legislation and many others that cannot even be predicted. The same way AI benefits can reach more people, its risks also present a larger potential impact. The more powerful and far-reaching a technology is, the more devastating its damaging effects can be, so it is paramount to establish mechanisms and legislation to prevent those risks and set appropriate compensations in case of malfunctioning or undesired outcomes.

We have studied how discrimination in AI and algorithms differs from traditional discrimination forms and how discrimination appears by the mere use of AI, both in theory and in practice, through studying Uber's case, as it is essential to understand the problem before trying to provide solutions to it or analyze the solutions that have been provided.

Discrimination in algorithms can come from the biases in the coders, that are translated into the algorithm; from the input parameters that the apparently neutral algorithm receives to make decisions, which are characteristics tied to protected groups that lead to discriminatory outcomes, or from the data the algorithm learns from, which has biased and discriminatory patterns because it is data that comes from society's decision patterns, which are discriminatory. Therefore, structural and systemic race and gender discriminations are perpetuated and accentuated through technology, as we have seen in Uber's case study. The company's algorithm presents at least the second and third type of discrimination. We cannot fully affirm that the discrimination comes from bias in coders, because the algorithm is not public. Nonetheless, we want to point out that biases are inherent to human nature, but their presence in AI may have been corrected through conscious analysis and a diverse team of coders. Regardless, what we can affirm is that the algorithm receives biased or discriminatory data in the form of ratings, as users determine the score in a directly or indirectly discriminatory way: depending on the user, he or she will have given a lower score due to racist or sexist motives, so it will consist on direct discrimination, or due to other motives, which will amount to indirect discrimination as it will lead to a discriminatory outcome. Moreover, in the discrimination consisting in different ride prices depending on the race, we can see that apparently neutral input data, such as neighborhood, is factored in to determine the price and it leads to discriminatory outcomes. Albeit Uber's justification for this difference is based on supply and demand, so it does not amount to direct discrimination, it does result in indirect discrimination based on race.

What has been done by Uber is making the underlying transaction with protected groups easier by decreasing the salience of race and gender identifying features, such as the name, as the Uber app now shows the name of the passenger after the ride has been accepted. However, this fix has not solved the problem, as now discrimination is deviated to post-acceptance cancellations once the name is shown.

These company-specific solutions are not sufficient, and discrimination forms are not exclusive to Uber, so scholars and legislators have tried to provide different mechanisms to respond to these new and complex technologies and its harmful outcomes. Scholars, AI experts and human rights advocates have called for algorithmic transparency. Legislators in the EU have established some transparency requirements for high-risk systems and non-high-risk systems when they interact with natural persons, but in the US transparency has been deemed voluntary, which is consistent with the higher protection of business freedom there is in that country.

However, transparency may not always be the best policy, as many authors are starting to point out, due to the large amounts of data that would need to be analyzed, its complexity, the generation of free riders that take advantage of knowing the functioning of algorithms and the fact that transparency, in and of itself, does not imply finding the root of the discrimination and eliminating it. Instead of transparency, or a complement to it, the mandatory audits and analysis of discrimination as a prerequisite and as a continuous requirement have been established by the EU Commission and the US in order to reduce the risk of AI bias and discrimination. This is in line with the call for accountability rather than transparency, but full accountability is not achieved either with these legislations.

When companies fulfill the requirements of the legislation because they have neutral setups but end up having discriminatory effects or because they input data with embedded discrimination, they are not held accountable for their results. Platform economies are especially sensitive to this, as some of the discrimination derives from the inner workings of the algorithms but another important part comes from the discrimination by humans, which is then magnified by the platform itself, generating important network effects that severely hinder human rights.

Therefore, to sum up and give an appropriate answer to the research question, platform economies and its use of AI and algorithms have an impact on discrimination in that they do not increase natural discriminatory tendencies of human behavior, but they generate worse effects to the discriminatory pattern of individuals. The same discriminatory conduct (for example, treating a woman poorly because she does not engage in conversation or not picking up a black person leading to longer waiting times) is channeled through network effects and is aggregated in the platform, which increases the discriminatory results (as the translation of discriminatory behaviors in lower scores creates an additional discriminatory effect of longer waiting times or worse benefits). This is because the discriminatory conduct of multiple individuals with respect to one individual due to that individual's race or gender is added up (and reflected in overall lower scores). Moreover, the use of algorithms can lead to discrimination forms tied exclusively to that system, as traditional functioning did not lead to expulsion of the market (or at least of the company's platform, which is an essential element) or differences in pricing (for example, taxis have a standard fare for the same city, but the price prediction done by Uber and other ride sharing platforms leads to different pricing discriminating by race). Expulsion of the market or differences in pricing may exist also in traditional methods, as somebody may not be accepted in a market because of gender or race or may be given a different price (as happens with salaries), yet the novelty is that the discrimination does not come from a specific individual's decision but an apparent aggregation of discrimination that deems it more difficult to determine responsibility. However, in the end, it is the company that enables this aggregation, so the liability should be imposed on them.

The attempts of companies to self-regulate in order to solve these problems have proven not to be enough and, while current legal solutions give a response to some of the issues AI and algorithms present and are good tools for risk prevention, they lack an adequate solution for actual discriminatory outcomes adapted to the actual dimension of those due to the network effects. Although steps in the right direction have taken place, the challenge AI poses has not been met with satisfactory solutions as the accountability demanded to companies does not provide strong enough incentives to eliminate discrimination overall.

6. References

Algorithmic Accountability Act of 2019, H.R. 2231, 116th Congress (2019-2020)

Altman, A. (2020, April 20). Discrimination. *Stanford Encyclopedia of Philosophy*, *Winter 2020*. https://plato.stanford.edu/archives/win2020/entries/discrimination/

Angwin, J., Larson, J., Mattu, S. M., & Kirchner, L. (2016, May 23). *Machine Bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Ayres, I., Banaji, M., & Jolls, C. (2015). Race effects on eBay. *The RAND Journal of Economics*, *46*(4), 891–917. https://doi.org/10.1111/1756-2171.12115

Barnett, J., Soares Koshiyama, A., & Treleaven, P. (2017, August 22). *Algorithms and the law*. Legal Futures. https://www.legalfutures.co.uk/blog/algorithms-and-the-law

Boix, A. (2020). Los algoritmos son reglamentos. *Revista de Derecho Público: Teoría y Método*, *1*, 223–270. https://doi.org/10.37417/rpd/vol_1_2020_33

Brown, A. E. (2019). Prevalence and Mechanisms of Discrimination: Evidence from the Ride-Hail and Taxi Industries. *Journal of Planning Education and Research*, 0739456X1987168. https://doi.org/10.1177/0739456x19871687

Ciechalski, S. (2017, June 28). *A civil rights organization is suing Uber for not having wheelchair-accessible cars*. Mashable. https://mashable.com/2017/06/28/uber-lawsuit-wheelchair-accessibility/?europe=true

Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., & Chaintreau, A. (2020). Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics. *Proceedings of the 21st ACM Conference on Economics and Computation (Pp. 679–681)*. Published. https://doi.org/10.2139/ssrn.3615404

Dastin, J. (2018, October 11). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Datta, A. (2021, February 24). *3 kinds of bias in AI models — and how we can address them*. InfoWorld. https://www.infoworld.com/article/3607748/3-kinds-of-bias-in-ai-models-and-how-we-can-address-them.html

Daws, R. (2020, July 20). *Uber hit with lawsuit to reveal how its algorithm works*. Internet of Things News. https://iottechnews.com/news/2020/jul/20/uber-hit-lawsuit-reveal-how-algorithm-works/

Desai, D. R., & Kroll, J. A. (2017). Trust but verify: a guide to algorithms and the law. *Harvard Journal of Law & Technology*, *31*. https://ssrn.com/abstract=2959472

Doleac, J. L., & Stein, L. C. (2010). The Visible Hand: Race and Online Market Outcomes. *The Economic Journal*. Published. https://doi.org/10.2139/ssrn.1615149

Edelman, B., & Luca, M. (2014). Digital Discrimination: The Case of Airbnb.com. *Harvard Business School Working Paper*, *14–054*. https://doi.org/10.2139/ssrn.2377353

European Commission - Joint Research Centre. (2019, January 2). *The Future of Cities*. European Commission. https://urban.jrc.ec.europa.eu/thefutureofcities/social-segregation#the-chapter

European Commission. (2021a, April 21). *Communication on Fostering a European approach to Artificial Intelligence* [Press release]. https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence

European Commission. (2021b, April 26). *A European approach to Artificial intelligence*. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

European Commission. Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts. COM/2021/206, 21st April 2021.

European Court of Justice. Judgement of the Court (Grand Chamber) of 20 December 2017, Asociación Profesional Elite Taxi v Uber Systems Spain, SL, C-434/15, ECLI:EU:C:2017:981

European Union. Agency for Fundamental Rights, Council of Europe, & European Court of Human Rights. (2018). *Handbook on European Non-discrimination Law*. European Agency for Fundamental Rights.

European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). OJ L 119, 4th May 2016, p. 1–88

Ge, Y., Knittel, C. R., MacKenzie, D., & Zoepf, S. (2020). Racial discrimination in transportation network companies. *Journal of Public Economics*, *190*, 104205. https://doi.org/10.1016/j.jpubeco.2020.104205

Gordon Clausen, K. (2020, November). *How to tackle bias in AI*. 2021.AI. https://2021.ai/how-to-tackle-bias-in-ai/

Hargrave, M. (2021). *Deep learning*. Investopedia. https://www.investopedia.com/terms/d/deep-learning.asp

Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The Evolution of Cognitive Bias. *The Handbook of Evolutionary Psychology*, 1–20. https://doi.org/10.1002/9781119125563.evpsych241

Hildebrandt, M. (2020). The Artificial Intelligence of European Union Law. *German Law Journal*, *21*(1), 74–79. https://doi.org/10.1017/glj.2019.99

Hosanagar, K., & Jair, V. (2018, July 23). *We Need Transparency in Algorithms, But Too Much Can Backfire*. Harvard Business Review. https://hbr.org/2018/07/we-need-transparency-in-algorithms-but-too-much-can-backfire

Iqbal, M. (2021, May 16). *Uber Revenue and Usage Statistics (2021)*. Business of Apps. https://www.businessofapps.com/data/uber-statistics/

Kantarci, A. (2021, April 17). *Bias in AI: What it is, Types & Examples of Bias & Tools to fix it*. AIMultiple. https://research.aimultiple.com/ai-bias/

Kossow, N., Windwehr, S., & Jenkins, M. (2021, February). *Algorithmic transparency and accountability*. Transparency International. https://knowledgehub.transparency.org/assets/uploads/kproducts/Algorithmic-Transparency_2021.pdf

Lee, D. (2019, January 29). *"Thrown to the wolves" - the women who drive for Uber and Lyft*. BBC News. https://www.bbc.com/news/technology-46990533

Manyika, J., Silberg, J., & Presten, B. (2019, October 25). *What Do We Do About the Biases in AI?* Harvard Business Review. https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai

Mullainathan, S. (2019, March 22). *Discrimination by Algorithm and People* [Research presentation]. Legal Challenges of the Data Economy conference, Paris, France. https://www.law.uchicago.edu/recordings/sendhil-mullainathan-discrimination-algorithm-and-people

New, J. (2019, September 23). *How to Fix the Algorithmic Accountability Act*. Center for Data Innovation. https://datainnovation.org/2019/09/how-to-fix-the-algorithmic-accountability-act/

New, J., & Castro, D. (2018, May). *How Policymakers Can Foster Algorithmic Accountability*. Center For Data Innovation. https://www2.datainnovation.org/2018-algorithmic-accountability.pdf

O'Neil, C. (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Reprint ed.). Crown Publishing Group (NY).

Pandey, A., & Caliskan, A. (2021). *Disparate Impact of Artificial Intelligence Bias in Ridehailing Economy's Price Discrimination Algorithms*. AIES Conference, Virtual. https://arxiv.org/pdf/2006.04599.pdf

Pope, D. G., & Sydnor, J. R. (2011). What's in a Picture?: Evidence of Discrimination from Prosper.com. *Journal of Human Resources*, *46*(1), 53–92. https://doi.org/10.1353/jhr.2011.0025

Rosenblat, A., Levy, K. E., Barocas, S., & Hwang, T. (2017). Discriminating Tastes: Uber's Customer Ratings as Vehicles for Workplace Discrimination. *Policy & Internet*, *9*(3), 256–279. https://doi.org/10.1002/poi3.153

Schoenbaum, N. (2016). Gender and the Sharing Economy. *Fordham Urban Law Journal*, *43*, 1023–1070. https://ir.lawnet.fordham.edu/ulj/vol43/iss4/4

Sweeney, L. (2013). Discrimination in Online Ad Delivery. *SSRN Electronic Journal*. Published. https://doi.org/10.2139/ssrn.2208240

Uber. (n.d.). *Uber Cities - Rides Around the World*. Retrieved April 2021, from https://www.uber.com/global/en/cities/

Waters, C. (2018, August 10). *'Nice girls' and Uber: Why the rating system is a gender trap*. The Sydney Morning Herald. https://www.smh.com.au/business/small-business/nice-girls-and-uber-why-the-rating-system-is-a-gender-trap-20180810-p4zwrs.html

Annex 1: Algorithmic Accountability Act of 2019 – Section 2

SEC. 2. DEFINITIONS.

(7) HIGH-RISK AUTOMATED DECISION SYSTEM.—The term "high-risk automated decision system" means an automated decision system that—

(A) taking into account the novelty of the technology used and the nature, scope, context, and purpose of the automated decision system, poses a significant risk—

(i) to the privacy or security of personal information of consumers; or

(ii) of resulting in or contributing to inaccurate, unfair, biased, or discriminatory decisions impacting consumers;

(B) makes decisions, or facilitates human decision making, based on systematic and extensive evaluations of consumers, including attempts to analyze or predict sensitive aspects of their lives, such as their work performance, economic situation, health, personal preferences, interests, behavior, location, or movements, that—

(i) alter legal rights of consumers; or

(ii) otherwise significantly impact consumers;

(C) involves the personal information of a significant number of consumers regarding race, color, national origin, political opinions, religion, trade union membership, genetic data, biometric data, health, gender, gender identity, sexuality, sexual orientation, criminal convictions, or arrests;

(D) systematically monitors a large, publicly accessible physical place; or

(E) meets any other criteria established by the Commission in regulations issued under section 3(b)(1).

Annex 2: Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) – Annex III

ANNEX III. HIGH-RISK AI SYSTEMS REFERRED TO IN ARTICLE 6(2)

High-risk AI systems pursuant to Article 6(2) are the AI systems listed in any of the following areas:

1. Biometric identification and categorisation of natural persons:

(a) AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons;

2. Management and operation of critical infrastructure:

(a) AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity.

3. Education and vocational training:

(a) AI systems intended to be used for the purpose of determining access or assigning natural persons to educational and vocational training institutions;

(b) AI systems intended to be used for the purpose of assessing students in educational and vocational training institutions and for assessing participants in tests commonly required for admission to educational institutions.

4. Employment, workers management and access to self-employment:

(a) AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests;

(b) AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behavior of persons in such relationships.

5. Access to and enjoyment of essential private services and public services and benefits:

(a) AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services;

(b) AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by small scale providers for their own use;

(c) AI systems intended to be used to dispatch, or to establish priority in the dispatching of emergency first response services, including by firefighters and medical aid.

6. Law enforcement:

(a) AI systems intended to be used by law enforcement authorities for making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending or the risk for potential victims of criminal offences;

(b) AI systems intended to be used by law enforcement authorities as polygraphs and similar tools or to detect the emotional state of a natural person;

(c) AI systems intended to be used by law enforcement authorities to detect deep fakes as referred to in article 52(3);

(d) AI systems intended to be used by law enforcement authorities for evaluation of the reliability of evidence in the course of investigation or prosecution of criminal offences;

(e) AI systems intended to be used by law enforcement authorities for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups;

(f) AI systems intended to be used by law enforcement authorities for profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of detection, investigation or prosecution of criminal offences;

(g) AI systems intended to be used for crime analytics regarding natural persons, allowing law enforcement authorities to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data.

7. Migration, asylum and border control management:

(a) AI systems intended to be used by competent public authorities as polygraphs and similar tools or to detect the emotional state of a natural person;

(b) AI systems intended to be used by competent public authorities to assess a risk, including a security risk, a risk of irregular immigration, or a health risk, posed by a natural person who intends to enter or has entered into the territory of a Member State;

(c) AI systems intended to be used by competent public authorities for the verification of the authenticity of travel documents and supporting documentation of natural persons and detect non-authentic documents by checking their security features;

(d) AI systems intended to assist competent public authorities for the examination of applications for asylum, visa and residence permits and associated complaints with regard to the eligibility of the natural persons applying for a status.

8. Administration of justice and democratic processes:

(a) AI systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts.