HiLL PUBLISHING

Journal of Humanities, Arts and Social Science, 2023, 7(1), 128-154
https://www.hillpublisher.com/journals/jhass/
ISSN Online: 2576-0548
ISSN Print: 2576-0556

Data Governance and Human Rights: An Algorithm Discrimination Literature Review and Bibliometric Analysis

Yi Wu

Goldman School of Public Policy, University of California, Berkeley, 94720, California, USA.

How to cite this paper: Yi Wu. (2023) Data Governance and Human Rights: An Algorithm Discrimination Literature Review and Bibliometric Analysis. *Journal of Humanities, Arts and Social Science*, 7(1), 128-154.
DOI: 10.26855/jhass.2023.01.018

Received: December 18, 2022
Accepted: January 12, 2023
Published: February 9, 2023

***Corresponding author:** Yi Wu, Goldman School of Public Policy, University of California, Berkeley, 94720, California, USA.
Email: yi_wu2022@berkeley.edu

Abstract

Algorithm discrimination becomes a serious problem. Its far-reaching implications have attracted the interest of many studies in a variety of disciplines, including computer science, law, sociology and economics, and hence the need for further review and categorization of the literature through systematic means. The research in this paper therefore systematizes 159 algorithms of discrimination papers and books. Primarily descriptive analysis is used to highlight different forms of discrimination in ten industries. The machine learning and deep learning processes played a leading role in the systematic review, involving both ethical topics and human rights themes. Accordingly, the discussion of implications and limitations on data governance is based on input-modeling-output pipeline. Ethical early intervention, algorithm's self-fulfilling loop, dynamic managerial capability, transparency and traceability of algorithms, and tradeoff between innovation and regulation are the main challenges to current data governance.

Keywords

Algorithm discrimination, data, fairness, equality, governance

1. Introduction

Algorithms now have penetrated into human society, which make is possible to communicate remotely with our friends and family, to discover new residence, trendy music, current news and desired information (Gangadhara, 2014). However, the public perception of this is not proportional. One survey of Facebook users revealed that most people were unaware that an algorithm were once used by to filter what they were feed daily (Sandvig, Karahalios & Langbort, 2014). This is what philosopher Charles Taylor describes as the new "social imaginary", a shared public understanding of how the world looks like and what can be called "normal" (Charles, 2009). Therefore, the normal world is being controlled by algorithms, even if most people don't notice it.

Since the 1960s and 1970s, i.e., after the Civil Rights movement and other upheavals in society, significant political and social transformation reduced discrimination effectively. However, with the rise of the Internet age and the explosive development of digital, discrimination lurks in social life in a very hidden way, disguised as innovation. Discrimination is a flagrant violation of human rights which we need to fight it.

The first step is to rearticulate human rights in this new era context. Article 1 of the Universal Declaration of Human Rights states that all human beings are born free and equal in dignity and rights. Article 2 emphasizes that everyone is entitled to all the rights and freedoms set out in this Declaration, without distinction of any kind, such as race, color, sex, language, religion, political or other opinion, national or social origin, property, birth or other status. The Australian Human Rights Commission's definition is even more straightforward: human rights are a set

of fundamental principles about equality and fairness. We have the freedom and power to choose our own lives, to develop our own inner potential and to live a life free from discrimination, harassment or fear. The consensus of the world is that human rights are the basic rights of human being to be equal and fair.

Algorithmic discrimination poses a threat to human rights, as there are algorithm-driven forces eager to categories, exclude, blow up and even treat people differently in a variety of ways. This mysterious force is shaping our society, exacerbating social and economic inequality and inequity. The 2014 White House report issued a warning that such algorithmic discrimination may be a side effect of Big Data Technologies (Executive Office of the President, 2014).

It not only exists our familiar contexts such as racial discrimination and gender discrimination. For example, the screening system for human resources automatically discriminates. But also, some hidden discrimination exists in our daily life. Like price discrimination, quality discrimination, opportunity discrimination and so on, all kinds of discrimination in the hotbed of data breeding.

This study aims to present a grander picture of how the algorithm construct decision trees in different industries and its link to social injustice and inequity. After that, the paper will discuss data governance, human rights, innovation and current regulation. Finally, this study reveals the existing limitations and possible directions of data governance on algorithmic discrimination.

2. Algorithm discrimination

The US anti-discrimination law aims to solve discrimination issues in other areas of social life such as employment, housing, education and public accommodation. Although the application of the algorithm is not specified, its definition of discrimination is worth referring to. Discrimination is reflected in the differential treatment of persons, i.e., the "discriminatory intent or formal application of different rules to different groups of persons", and/or the differential impact, i.e., the result of "different groups of persons" (Barocas & Selbst, 2014).

Price discrimination and quality discrimination, as the most prevalent algorithmic discrimination in the business world, have existed for nearly 60 years. American Airlines, the first to use it in 1960s, calls it a new competitive strategy which has another name “screen science” (Petzinger Jr., 1996).

The US Civil Aeronautics Administration and the Department of Justice immediately launched an antitrust investigation after major travel agencies and other national airlines responded that US flights were often the first flights back - not only much more expensive but also much longer than other flights (Sandvig, Hamilton, Karahalios, & Langbort, 2014).

Some scholars pointed that algorithmic discrimination is widely accepted as the production of discrimination (Noble, 2018; Eubanks, 2018) in the way that members of a protected group or class are considered or treated; it is largely traced to the existence of 'automation bias’s and 'bias by proxy' in algorithmic models. “Automation bias” and “bias by proxy” in algorithmic models. Automated biases are social and cultural biases that are propagated on a large scale via the machine learning process and are ingrained in the historical training data for driving algorithm. Proxy bias arises where unintended proxies for protected variables (gender, race, etc.) can still be adopted to reconstruct and infer biases by proxy that are very difficult to detect and eliminate. In order to ensure fairness in the algorithmic decision making and machine learning process, these methods of treating fairness as non-discrimination through unbiased machine learning models are highly operational and include “discrimination prevention analytics and strategies” (Romei & Ruggieri, 2014) and “fairness-and discrimination-aware data mining techniques”. These approaches are largely informed by calls in the industry for fairness in machine learning and are largely based on their integration with machine learning classifiers, the technical engineering of anti-discrimination standards and the control of distortion in the data for algorithms training (Ochigame, 2019).

Other researchers are trying to dissipate this doubt in different ways. Specifically, the different levels of data in the machine learning and deep learning process can be grouped and thus analyzed for differentiation. The three levels - input level, model level or output level - are the start of their entry intervention.

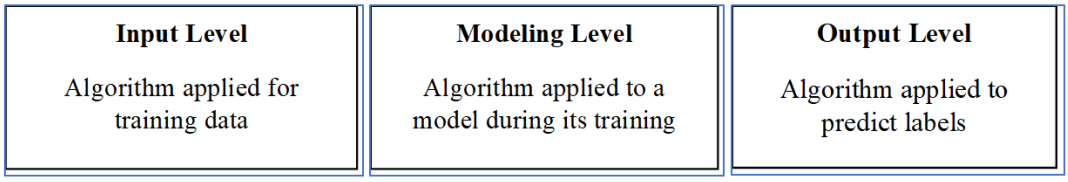

Figure 1. Input-modeling-Output of Algorithm.

At the input level, the simplest would be to drop the socio-demographic input variables, but this approach has not been successful because it reduces the visibility of the problem, ignores possible proxy variables for socio-demographic characteristics and opportunities to implement alternative solutions, thus inadvertently distracting from equity—For example, in one scholar's analysis of university enrolment in the US, he supported positive data-driven interventions at the output level (Kleinberg, Ludwig, Mullainathan, & Rambachan, 2018).

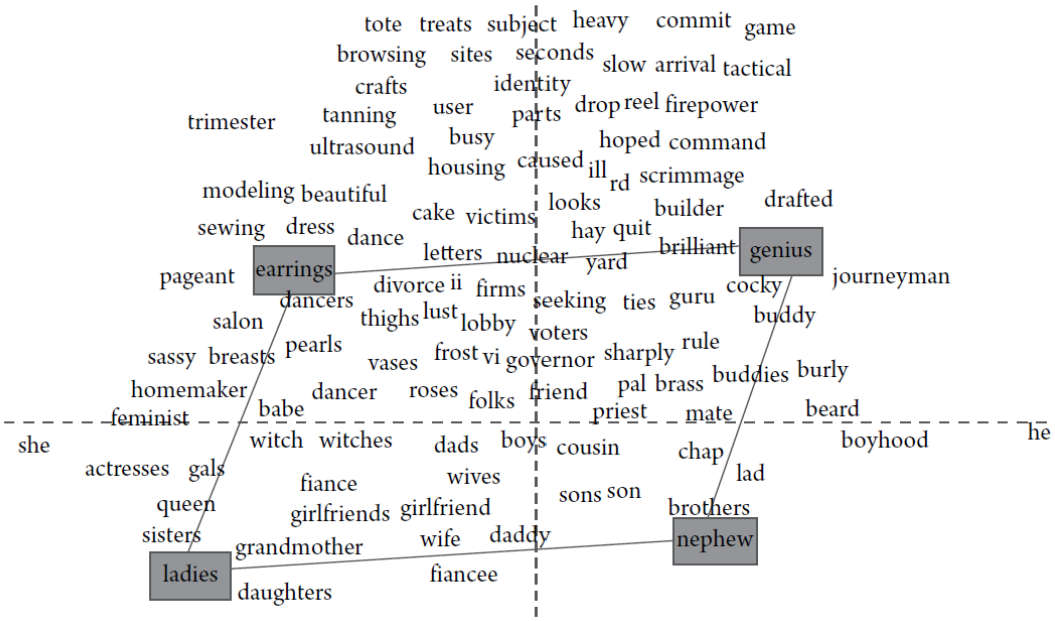

Figure 2. A word embedding exhibiting gender bias.

The problem is that the training data for machine learning applications often contains a variety of hidden (and not so hidden) biases, and in the complex models built from this data, these biases can be amplified and new biases can be introduced. Machine learning tends to be deterministic and bounded, and it does not give you things like gender neutrality 'for free' that you have not explicitly asked for. So even though there are few documents that can be created that exhibit a clear gender bias, when this data is compressed into a predictive model of word analogy, any language about this across the dataset converges into a collective force that creates a clear bias. And when such models are applied to a wide range of communication media, such as search engines, advertising placements, job boards, etc., these biases are continually propagated and amplified.

At the output level, Moerel deems **LinkedIn**, one of the world-famous recruitment tools, as a means of using algorithmically generated ranking for achieving a better predefined desirable ratio. The candidates are categorized by gender initially, each within each gender is ranked according to its algorithm and finally equal numbers of men and women to hiring are presented to manager for selection (Moerel, 2018). In 2018, Facebook disclosed an internal tool Fairness Flow and announced it will be tested for the identification of level discrimination, and in July 2020 Civil Rights Report the plan was discussed further for tackling intensified algorithmic discrimination (Murphy & Cacace, 2020). In 2018, **Google** launched the open-source What If Tool at Tensor Board to facilitate ML developers' work to analyze differences in the classification of key variables, check the boundaries of specific classifications, and document the impact of counterfactuals as a way to assess the likelihood of extracting inappropriate social bias from training data or otherwise mirroring it in their models (Wexler, 2018).

Some academics have also attempted to model specific loss functions subject to fairness constraints, a concept known in academic parlance as cost-sensitive classifiers. In simple terms, this is a professional empirical evaluation of the results on a range of datasets (Agarwal, Beygelzimer, Dudík, Langford, & Hanna, 2018). Others are working on a specific family of classifiers called plain Bayes, in which the concept of discriminative patterns is introduced and algorithms for mining that pattern. This is an iterative approach, with the overall aim of eliminating such patterns to obtain a fair model (Choi, Farnadi, Babaki, & Broeck, 2020). The following figure can help clarify what happened at the output level.

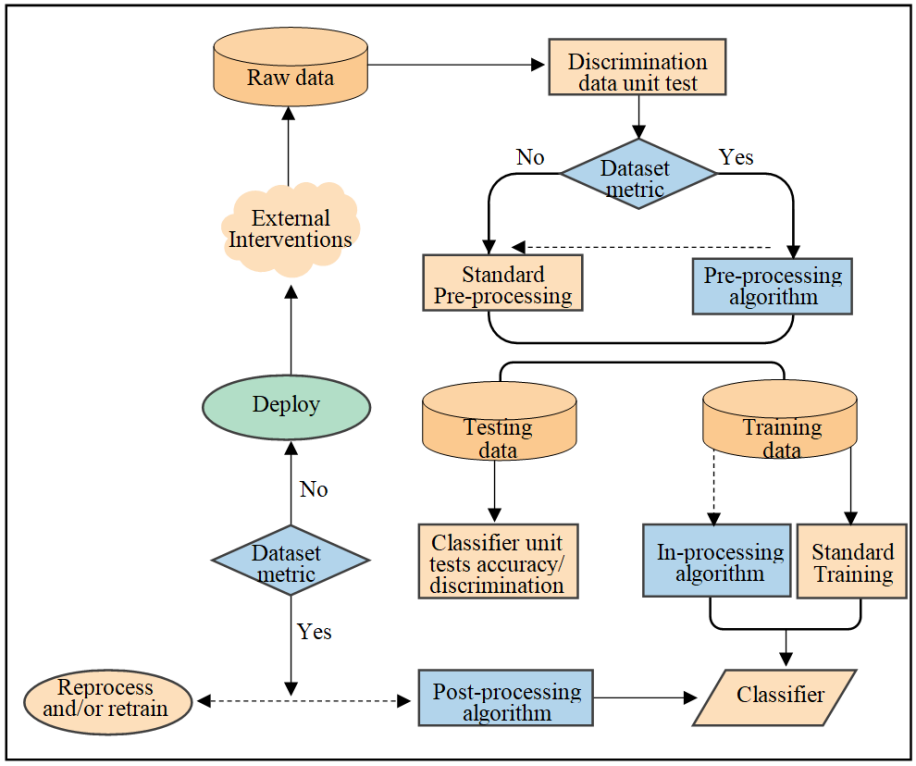

Figure 3. Machine Learning Framework.

Parameters or algorithms can be adjusted during the modelling process to reduce the model's dependence on some patterns in the input data. For example, the gender-debiasing techniques have been developed for word embedding solutions (Bolukbasi, Chang, Zou, Saligrama, & Kalai, 2016), where a mix of adjustment inputs and model-level adjustments are mentioned in the essay. The problem of biased training data is present in a variety of scenarios and decisions. Cameras used to recognize faces often fail to accurately identify specific faces: facial recognition software often has a higher recognition rate for white faces compared to non-white faces. The data scientist "eventually traced the error back to the source: In his original data set of about 5000 images, whites predominated" (Dwoskin, 2015). The data expert did not intend to write the algorithm to target only the white population, but he used mostly white faces. As Aylin Caliskan, a postdoctoral fellow at Princeton University, pointed out that "AI is biased because it reflects effects about culture and the world and language. Sowhenever you train a model on historical human data, you will end up inviting whatever that data carries, which might be biases or stereo- types as well" (Chen, 2017).

If we set a step back to the essence of algorithmic discrimination, which is the mechanism by which algorithms are generated, then understanding the relationship between Machine Learning (ML), Artificial Intelligence and Deep Learning (DL) and how they work can help us understand the underlying causes of algorithmic discrimination.

In the Internet era, people have invented smart devices that are connected to the Internet, which means that any device that can be connected to the Internet can be called a smart device. In this context, Artificial Intelligence (AI) came into being, coded to make devices even smarter—that is, with minimal or no human intervention. Two other algorithms, Machine Learning (ML) and Deep Learning (DL), are also prevalent today as new types of algorithms that can empower smart devices (Kumar, 2020).

AI is also a big concept that incorporates both Machine Learning (ML) and Deep Learning (DL), namely an umbrella term used for ML and DL. Deep learning, on the other hand, is a subset of machine learning (see above), and specifically refers to artificial neural networks (ANN) that contain multi-layered, large data sets that serves as powerful computer hardware to make complicated training models possible. This set of powerful neural network learning techniques encompasses multiple methods and techniques that use increasingly rich artificial neural networks with multiple layers of functionality (Gupta, 2020).

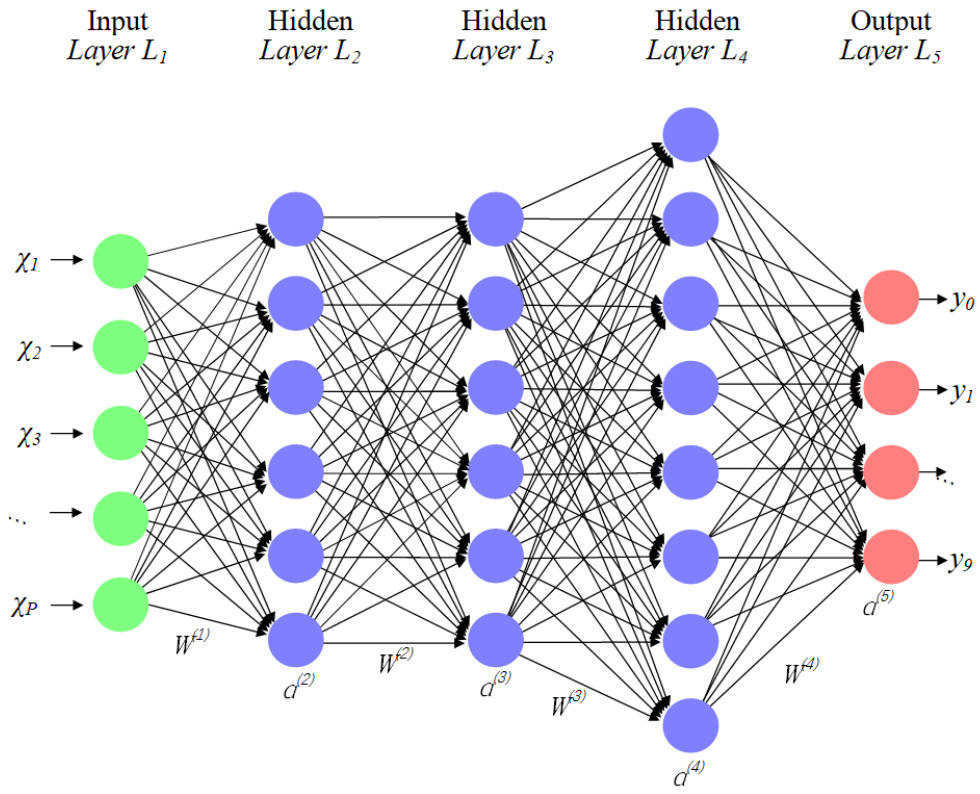

Figure 4. Machine Learning and Deep Learning Neuron Process.

3. Methods

The study was developed over a series of phases, in the same way as systematic literature reviews. See Figure 5.

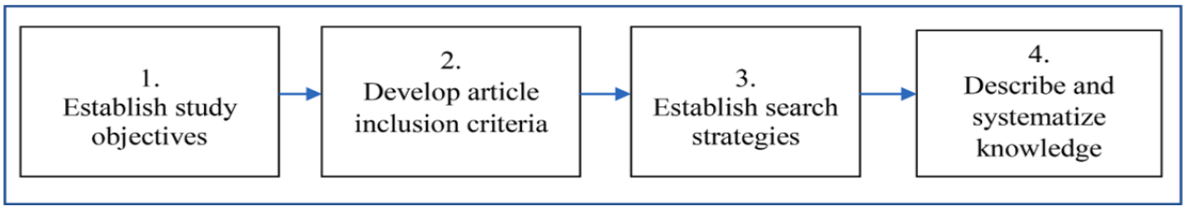

Figure 5. Stages of this study.

3.1 Study Objectives

For the first phase, the research focused on analyzing the extant academic literature on algorithmic discrimination, particularly in the area of ethics. The objectives of this phase are divided into three main parts:

- O1: To analyze the intrinsic links and characteristics of the relevant researches
- O2: To systematically classify and group the content and findings of the studies
- O3:To identify gaps and deficiencies in current knowledge in the field and thus identify new directions for future research

Accordingly, it needs to be recognized that algorithmic discrimination is a multidisciplinary nature of research that touches all aspects of life, such as online consumption, retail, public health, public finance, public education, human resources, criminal and judicial, ride-hailing services, automobiles and advertising. Machine learning and big data techniques are the main methodologies for algorithmic discrimination.

3.2 Criteria for selection of articles

The study sets a number of criteria for article-selection. Most importantly, the study must address data govern-

ance and human rights from the perspective of algorithmic discrimination only. Therefore, the two factors, data control and human rights, need to be distinguished from other areas under the algorithmic framework. Furthermore, the study has quality access criteria for the selection of manuscripts, i.e., it is limited to accepting only English-language manuscripts and does not restrict publishers or publication dates.

The articles that fail to meet the selection criteria are characterized by, firstly, their greater relevance to privacy and the right to information, but with no or only a hint of relevance to the issue of big data discrimination or inequality. Secondly, they relate to discrimination issues but are not relevant to the development of big data analytics. Thirdly, they focus on the widening gap between institutions that have used more power and resources to analyse and understand big data ("the Big Data rich") and those that do not ("the Big Data poor"), rather than on the concept of the digital divide, i.e., the difference in the conditions under which individuals have access to data on the internet. Fourthly they assess the differences that affect social media engagement. In a subsequent phase of the literature review study, this paper will analyse the 159 articles available.

3.3 Searching strategies

In this regard, I developed the appropriate search strategy and followed the selection criteria. Firstly, three search terms were extracted simultaneously: data control, human rights, and algorithmic discrimination. Secondly, the search was filtered by title, abstract and article keywords.

During the search process, other search terms were added, one of which was highly cited in the relevant literature. Therefore, for the keyword "algorithmic discrimination", the alternative word "digital bias," "algorithm bias" could be used. Similarly, regarding human rights, "fairness," "equality," and "ethics" were also the alternative words.

In the search process, the first round of filtering is done by reading the title and abstract of the article before moving on to the second round of purification. One of the biggest challenges encountered was that although some of the articles revolved around discrimination and contained words such as human rights or near-synonyms in their titles, abstracts, and introductions, they were not clearly presented. From the beginning of the review to the end of 23 October 2022, a total of 159 relevant articles met the criteria and were selected as reference.

3.4 The methods for description and systematization of literature

Various bibliometric and systematization techniques were used to characterize and systematize the content of the 159 articles. Firstly, bibliometric techniques were used to assess the descriptive data of relevant content in the literature. Secondly, a visual analysis of co-occurrence was conducted. The study systematizes the literature by analyzing the knowledge and findings in the literature in order to build a scientifically sound knowledge architecture (Kaur, Dhir, Tandon, Alzeibyg, & Abohassan, 2021).

The study focused on the process of identifying a universal algorithmic discrimination generation and therefore setting the study selection process as input variables - modelling - output variables.

4. Descriptive and bibliometric analysis

4.1 Descriptive analysis of related articles

Of the final 159 selected papers and books, the main industries covered are healthcare, criminal activity prediction (US criminal justice system), police, credit, insurance, recruiting and employment, university admissions, housing, lending, criminal justice, sentencing, business promotion, advertising and marketing, search results, prioritization of news, health data forecasting, facial recognition, risk assessment, forecasting recidivism (criminal), mortgage, government services. The platforms involved are Google, Facebook, Wikipedia, Uber, Lyft.

Relevant laws and regulations include, Universal Declaration of Human Rights, Anti-Discrimination Law, International Covenant on Civil and Political Rights, USCredit Reporting Act, EU Employment Equality Directive (2000/78/EC), EU General Data Protection Regulation (GDPR), the European Convention for the Protection of Human Rights, the US Civil Rights Act, the UK Equality Act 2010.

4.2 Analysis of the co-occurrence

This study adopted a co-occurrence analysis as a way to assess the relationship of each topic in articles related to algorithmic discrimination. Reliable results were obtained in two specific ways: on the basis of keywords and on the basis of titles and abstracts.

Figure 6. Co-occurrence network of articles on the basis of keywords.

Cluster 1 and cluster 2 cross at "moral and ethical aspects". Cluster 1 (right) gathers terms such as "technological innovations", "economic aspects" and "information technology". Cluster 2 (left) gathers terms such "cloud computing", "decision making" and "algorithm". Remarkably, the way they cluster is unrelated to each other. For example, algorithms are related to decision making, and technological innovation is related to economic impact.

Further, to allow for greater context richness, the second analysis was based on the titles and abstracts of selected articles. The data was cleaned by elimination or grouping. For example, generic or irrelevant words, such as "item," "author," and "study," were eliminated. However, similar terms were grouped together, as in the case of "decision making," "transparency," and "accountability." Figure 7 shows the results of the co-occurrence analysis based on titles and abstracts.

To add credibility to the study, the second analysis was conducted from the titles and abstracts of the selected articles. These data have been censored or restructured. For example, on the one hand, generic words or irrelevant words such as "item," "author," and "study," were removed. On the other hand, similar words were selected in the same group, like "decision making,""transparency," and "accountability". As shown in Figure 7.

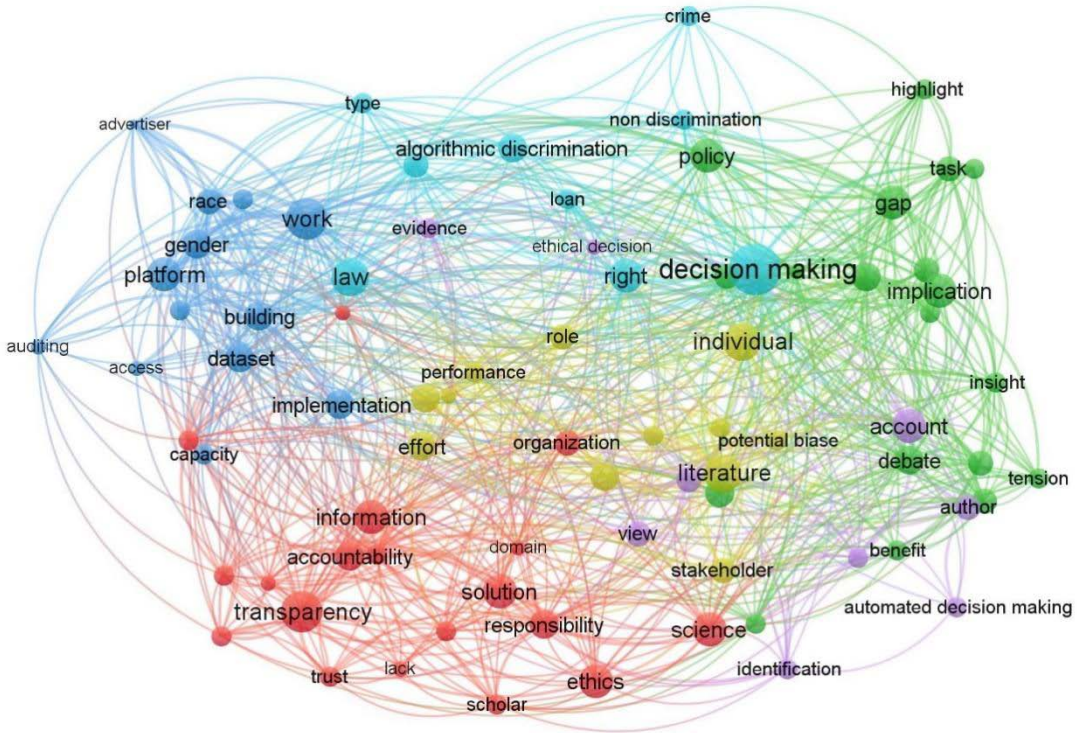

Figure 7. Co-occurrence network of articles on the basis of titles and abstracts.

Four clusters were derived from the co-occurrence analysis and are marked by four colors. In Cluster 1, (blue-green) “decision making” performed as a prominent node and included terms like “right”, “individual”, “policy” “gap” and “implication”. Thus, this cluster is related to decision making related policy implication and human rights. In Cluster 2 (red), many words became prominent nodes, including “transparency”, “accountability”, “trust”, “responsibility”, “solution,” “ethics,” and “scholar.” Hence, it regards data governance and human rights. In Cluster 3 (blue), no particular prominent nodes showed up and it included “law,” “platform,” “gender,” “race,” “dataset,” “building,” “work,” “auditing,” and “access,”, reflecting that this cluster represents the macro, micro and internal impact of the algorithm.

The results of the keyword co-occurrence analysis were similar to those of the title and abstract, which strengthens the credibility of the results. Meanwhile, there are more specific words in the abstract need our attention. This reveals the complexity presented in the area of algorithmic discrimination such as transparency, responsibility, accountability, identification, automated decision making. It also presents the suggestions that scholars have formed in the existing literature research to data governance.

5. Systematization of the related literature

5.1 Online Consumption

Initially one was able to buy the same goods at the same price when shopping online. The region-free nature of digital online shopping can eliminate the differences that previously arose in terms of geographical location. With the development of information technology, the growth of e-commerce has largely expanded and homogenized regional markets. Even if you live in a region where physical shops are very expensive, you can enjoy the same low prices for online shopping as consumers in other regions online. On a price level, you are just one of millions of consumers, anonymous (Mcsweeny & O’DEA, 2017).

In 2000, Amazon classified users based on their geographical location, spending history and online shopping behaviors (BBC News, 2000). The news soon came to light and, in the face of consumer outrage, Amazon quickly held a press conference claiming that the company was simply conducting a special experiment with a random number of accounts and compensating consumers who paid more than the average price for their online purchases (Amazon Press Center, 2000). In 2013, antivirus software developer McAfee adopted a categorical pricing strategy. Older users paid $79.99 for the same software when they upgraded, while new users paid only $69.90 (Caillaud & Nijs, 2013).

This type of price discrimination is one of algorithmic discrimination and is achieved mainly by companies offering different prices for goods to different classes of customers with different characteristics (Siegert & Ulbrichtb, 2020). In contrast to discounts or promotions, if a regular customer has to buy the same item at a higher price than the new customer would have to pay, the regular customer will have a sense of betrayal of the shopping experience and will therefore lose trust in the seller.

The collected literature suggests that there are three levels of price discrimination (Wu, Yang, Zhao, &Wu, 2022). **First-degree** discrimination is concerned with the customer's willingness to pay and is used by most retail industries to maximize revenue.For example, the use of a name-your-own-price (NYOP) strategies not only increases revenue and volume, such adaptive threshold prices can increase revenue by up to 20% without compromising consumer satisfaction (Hinz, Hann, & Martin Spann, 2011). **Second-degree** price discrimination, where consumers can identify or choose their own prices, is also an effective means of increasing revenue. For example, consumers may pay a higher price for a benefit or to improve their status (e.g., a high price for a first-class seat or a package ticket at the cinema with a massage) (Phillip Leslie, 2004).

Third-degree price discrimination refers to the establishment of different price systems for consumer groups with different characteristics for a particular category of goods. For example, the online travel agencies Hotel Tonight and Orbitz.com set prices according to the location and the computer system used (Mac or PC users) (Howe, 2017). If properly regulated, a third level of discrimination can be a win-win situation for both the company and the consumer. For example, a customer who thinks the product is of high quality is more likely to pay a higher price than a customer who thinks it is of low quality (Armstrong & Vickers, 2001). In this way, the company can generate higher revenue. For example, a customer who goes to a restaurant is willing to pay more for a specific table with a great view or better service (Borgesius & Poort, 2017). Some academics have specifically studied the revenue impact of two price discrimination mechanisms on retailers, namely second-degree price discrimination based on consumers' own volition pricing and third-degree price discrimination with store-level pricing. The results show

that pricing in combination with the second and third levels of price discrimination can allow companies to maximize their revenue (Khan & Jain, 2005).

5.2 Retailing

Recommendation algorithms help connect customers to products they need or want to buy. They also increase visibility to promotions and, in some cases, make shopping more fun (Ikeda, 2021). Even more and more people realize online consumption is the home of consumer data generation, however, online sellers are not the only area affected by algorithm-driven price discrimination. The offline scene of retail stores is still manipulated by algorithms, which is hard for consumers to detect.

For example, in brick-and-mortar retailers, merchants can use mobile phone Wi-Fi signals to track consumers in the shop to obtain their mobile phone model. In addition, brick-and-mortar retailers set prices for shelf items in advance and then issue coupons to individual consumers via mobile phones or through shop loyalty programs, offering them different discounts to increase revenue, a process also known as price discrimination.

In addition, retailers are encouraging their consumers to use mobile devices in their 'showrooms'. For example, to check prices in physical shops, or to buy products at lower prices from their devices rather than from physical shops. Consumers feel that it brings them convenience and physical store executives believe this encourages in-store loyalty. The multiple data of a customer brought by loyalty will form a complete portrait and score of a customer, and merchants can use the score to conduct differentiated marketing on this customer. This is the root of retailing data discrimination.

In Amazon's 'Go' shops or China's Bingo Box, in-store systems are in place to capture customers' movements and facial expressions and personalize products or services (Soo, 2017). Similar tracking systems are used by retailers in the UK and Switzerland, where customers are tracked in other forms such as test beacons, and the data collected informs personalized pricing in online shops. This tracking technology has generated higher profits for most retail shops (Retail News Insider, 2014). As a result, algorithmic pricing strategies are spreading rapidly across online and offline channels across the sales industry, against a backdrop of constantly updated information technology and an explosion in data collection.

It should be pointed out that due to the particularity brought by the immediacy of the sales nature of retail stores, algorithmic discrimination has its positive effect, that is, to avoid waste through dynamic pricing. Some pricing efforts also exist to reduce food waste. A startup called Wasteless Algorithms designed a dynamic pricing algorithm for supermarkets, whereby pricing is periodically adjusted based on product expiry dates. A Spanish retailer used the algorithm and the results showed that this not only effectively reduced food waste by a third, but also increased revenue by 6.3% (Rochelle, M, 2019). Wolak has proposed a non-linear pricing scheme based on water supply facilities set up in relation to changes in water demand and demographic household characteristics, which not only helps water companies to realize revenues and meet water saving targets, but is also a boon to society and the environment (Wolak, 2016).

5.3 Public health

Algorithms are infiltrating the healthcare. Primarily began in the 1970s, from consultation programming for glaucoma to automated intake processes in primary care to scoring systems that evaluate newborns' health conditions, patients regularly encounter these technologies and algorithms whether they know it or not (Christensen, Manley & Resendez, 2021). Personal patient data is the main training basis for these algorithms, which help the healthcare providers involved to make health decisions. The growing deployment of and dependence on algorithms has made healthcare beneficial for operational efficiency, waste reduction, health research and healthcare resource management. Insurance companies also use algorithms to determine risk and adjust healthcare costs.

Meanwhile, violation of fairness looms large. One such program for docking specialized resources for care is being used in thousands of hospitals in the United States. This program has been found to be biased and to further widen social disparities. When ranking patients in need in order of priority, white patients tend to be treated ahead of black patients with more serious conditions due to the erroneous use of past medical expenditure databases (where black patients have historically had lower levels of medical expenditure) as supporting data (Obermeyer, Powers, Vogeli, & Mullainathan, 2019).

Some scholars pointed out that applying the health care algorithm used in the study hospital to the rural clinic under the assumption that the rural clinic would have access to the same level of resources as the study hospital, the

algorithm generated health care resource allocation decisions that were not accurate and had many gaps (Danks & London, 2017).

Currently, the algorithms used in the clinic to predict future healthcare spending are color-biased, with white patients likely to enjoy higher healthcare spending than black patients, which would lead to disparities in healthcare delivery (Obermeyer, Powers, Vogeli, & Mullainathan, 2019).

Several other potentially racially biased models of ML are still in pre-clinical development. Several studies have clearly demonstrated that Black patients suffer from a strong racial bias in medical treatment. Specifically, black patients show higher rates of false negatives when using the opioid abuse classifier (Thompson, Sharma, Bhalla, Boley, McCluskey, Dligach, Churpek, Karnik, & Afshar, 2021).

Differences in mortality prediction and x-ray diagnosis among other races and ethnicities; and differences in burn recognition and diabetic retinopathy recognition among dark- and light-skinned patients. While no conclusions can be drawn regarding the prevalence of racial bias in published ML clinical studies, the breadth of ML clinical models susceptible to racial bias in this review exposes that a significant element of racial bias is drawn into the coding process in ML models and is likely to have a strong negative impact on patients in all aspects of healthcare.

As presented by Panch, Mattie and Rifat Atun, the application of algorithms not only has a macroeconomic and social impact, but also exacerbates inequalities in religion, socioeconomic status, gender, race, ethnic background, disability or sexual orientation to the detriment of equitable sharing of resources in the health system (Panch, Mattie, & Atun, 2019).

Discrimination against disadvantaged groups who are less vocal needs time to be rectified by giving them the necessary help in the meantime, and by trying to 'protect' and care for them so that they can access the same services and rights as the majority in society. This can be achieved through an algorithm that incorporates an artificial criterion that overemphasizes these groups over others (Igoe, 2021). This project is a pioneering area of research that is currently speculative and unproven, and still requires considerable effort to break through the technical barriers.

5.4 Public Finance

The existing legal system and regulatory bodies in the US are not yet sufficient to regulate and control algorithmic discrimination and are not sufficiently effective to prevent discrimination and enforce fair lending. The only existing laws are the legal systems developed in the 1960s and 1970s, such as the Fair Housing Act of 1968, the Truth in Lending Act of 1968, the Truth in Lending Act of 1968, and the Equal Credit Opportunity Act of 1974. At the time, we faced an almost exact opposite dilemma: there were not enough standardized data sources to make decisions and far too little credit data available (Klein, 2020).

At the end of 2019, an investigation of Apple cards by the New York Department of Financial Services revealed that Apple applied relevant algorithms to offer different credit limits to male and female users with apparently similar financial situations. This case is a typical application of algorithmic discrimination in the financial sector.

In the non-algorithmic world, credit is allocated based on the risk of the borrower, hence the term 'risk-based pricing'. The lender simply assesses the real risk of the borrower being able to pay back the money on time and then charges the borrower a fee for borrowing. The reference factors for determining risk, many of which are almost always linked to one or more of the protected categories at the social level. It is clear that determining the ability of a repayer to repay a loan is a legitimate commercial influence. As such, financial institutions are entitled to, and do, use factors such as debt, income and credit history to determine whether to offer credit and the interest rate at which to offer it, even if this relates to factors such as gender, race, etc.

In the algorithmic world, artificial intelligence combines machine learning and big data to incorporate different types of data into credit calculations. For example, your web browsing history, the type of phone you use, the movie tickets you buy, the shoes you wear and the places you shop. All traces of data you leave on the web, including online purchases and video browsing, have the potential to be integrated into credit models. But all that exists here is a statistical relationship; this data is still unpredictable, much less likely to be legally allowed to be incorporated into credit decisions.

Specifically, they surveyed Wayfair, a European online shopping company similar to Amazon but on a much larger scale, using the application of credit to complete online purchases as a special survey target. These five digital footprint variables are not only simple, but easy to access at no cost to the lender. Specifically, there is the type of computer the borrower has (Mac or PC); the type of device (phone, tablet, PC); the time of day the borrower ap-

plies for credit (if someone borrows at 3 am, it must not be a good sign); the borrower's email domain (Gmail is riskier than Hotmail); and is the borrower's name part of their email (if it is of the borrower's name, it indicates a high level of trust). Unlike improving credit scores, which is a traditional method used to determine who will receive a loan and to predict the interest rate on the loan (Berg, Burg, Gombovic, & Puri, 2018).

Algorithmic discrimination can transform consumer lending. On the positive side, the introduction of algorithmic discrimination could help identify millions of good credit risks who are currently being denied access to credit and allow them to re-qualify for credit, thereby escaping the criticism associated with meta-buying empiricism and allowing financial institutions to judge users more efficiently and comprehensively, resulting in a win-win situation for all. On the negative side, however, AI could bring about a wave of implicit discrimination, driven easily by historical discrimination factors ingrained in society, which could lead to the denial of credit or the use of a range of variables to increase interest rates (Klein, 2019).

5.5 Public Education

The three factors of race/ethnicity, nationality and gender have been the focus of most research on educational algorithms.

Several scholars have utilized five different algorithms to identify differential performance between students of different racial/ethnic groups in a six-year college graduation model. In the algorithm results it was found that white students had tended to have higher rates of false positives, while Latino students had higher rates of false negatives (Anderson, Boodhwani, & Baker, 2019).

Yu and colleagues made research on the potential rate of college dropout, discovering that non-white or Asian students were scored with worse true negative rates and easier recall of their identities in the algorithm models, and if the student is studying in person the results will get worse in accuracy (Yu, Lee, & Kizilcec, 2021). When the essay was scored automatically, it was shown that the E-Rater system has distinct preference for 11th grade Hispanic and Asian-American students, who were received significantly higher scores compared with human essay raters, while White and African American students' showing more accurate (Bridgeman, Trapani, & Attali, 2009). This effect did not appear again in subsequent studies of GRE students using the new version of E-Rater; instead, E-Rater gave much lower scores to African American students than manual graders for certain types of essays (Bridgeman, Trapani, & Attali, 2012).

Differences in the reliability of models of learners' native languages have been investigated in the context of educational applications of natural language processing, with automatic essay scoring as the main object of study. Naismith and others have found that word lists are one of the common measures of vocabulary complexity and that native speakers are better suited to use this learning aid as opposed to second language learners; while the method is effective in distinguishing learners with different proficiency ratings, there are also systematic differences in ratings for learners from different countries (simply put, in the case of two people from different countries with equal language Arabic-speaking learners tend to be rated lower than Chinese-speaking learners when their proficiency levels are comparable) (Naismith, Han, Juffs, Hill, & Zheng, 2018). In addition to this, they found evidence of significant differences between the corpora used by the different testing agencies.

Kai and other academics have examined the performance of male and female students and found that both groups performed very well. In particular, the JRip decision tree model presented fairer data than the J48 decision tree model, but the JRip model showed that girls still received slightly better grades than boys (Kai, Andres, Paquette, Baker, Molnar, Watkins, & Moore, 2017). Hu and Ranwala conducted a study on a range of algorithms that predict whether students are at risk of failing and found that the data measured by these models indicated that male students performed worse than female students, but this result could not match the true performance grades in university courses (Hu & Rangwala, 2020). In addition, Gardner and others looked at dropout prediction systems for MOOC universities and found that several algorithms presented data indicating that female students performed worse than male students, but this was attenuated in courses with 50-80% male student numbers. But again, surprisingly, when in courses with less than 45% male students, female students perform much worse (Gardner, Brooks, Andres, & Baker, 2018).

Western scholars Ryan S. Baker and Aaron Hawn point out that there are currently two challenges that prevent deeper research into groups known to be at risk of algorithmic bias as a way of detecting just how much bias exists. One of these is the economic challenge, where the power of capital is so great that algorithm developers encounter layers of obstacles to uncovering algorithmic biases in commercial systems, and where simply collecting the data

needed to investigate algorithmic biases is inherently risky and can easily run afoul of legal boundaries or core corporate interests (Baker & Hawn, 2021). Another challenge arises in the educational field, where algorithmic bias in the education system is intersectional, for example, where the specific identities of members of multiple groups can have a specific impact and hold each other back (Crenshaw, 1991).

5.6 Human Resources

Even before algorithms became commonplace, discrimination existed in human resources (HR), and Frijters notes that in HR recruitment and HR development, algorithms would recommend not hiring or supporting someone in their current position if their productivity-irrelevant characteristics were unpopular or not compatible to the value of the company (Frijters, 1996). As technology has evolved, algorithms have widely been applied to human resource management. This has triggered a new type of algorithm-based discrimination in the field of human resources.

Similar algorithmic discrimination existed in Amazon's original hiring algorithm when the option of male was the preferred criterion for career "fit", which resulted in female applicants not having priority for resume screening. These facts not only existed in the past, but are still very common in the development and implementation of existing algorithms, and this phenomenon is further exacerbated by the lack of historical data diversity in the field of computer and data science (Lee, 2019).

The original goal of algorithm-based HRM systems was to improve efficiency, particularly in recruitment applications. As early as 1996, it was suggested that implementing algorithmic systems to help HR make decisions could reduce the pressure on HR departments and increase efficiency by screening high-match talent for companies, reducing exceptionalism and providing companies with a continuous supply of a large workforce (Porter, 1996). For example, in industries such as hotel chains and retail, where staffing fluctuates, companies have to go through a large number of CVs and conduct a large number of interviews each year to fill vacancies and keep the company running (Leicht-Deobald, Busch, Schank, Weibel, Schafheitle, Wildhaber, & Kasper, 2019).

Providers of HR tool services are seeing new opportunities, touting their ability to provide staffing solutions to companies and citing them as the obvious choice to help them win the war for talent. This is because algorithmic results have been proven to be fairer and more reliable than manual screening.

Therefore, the main benefits of algorithm-based HR systems for companies are 1) efficiency 2) reduced bias from staff and 3) based on past hiring evidence, decisions are superior to human intuition. However, in recent years, many scholars have questioned the objectivity and fairness claimed by algorithm-based HR systems.

O'Neill, a mathematician with previous experience working in finance, points out that two factors, race and gender bias, still influence the algorithms of HR processes, specifically in the areas of staff recruitment, performance evaluation, etc (O'Neil, 2016). For example, after being trained by historical employment data, the hiring algorithm will position men as being more suitable for managerial positions, leading to the preconceived assumption that women are less interested in managerial positions, ultimately leading to the social outcome that managerial positions are mostly male. As a result, the recruitment algorithm is much less likely to push management job advertisements to women on social media, so that women do not have access to such advertisements and therefore do not have the opportunity to apply for such positions. In this case, the recruitment algorithm further 'optimizes' the algorithm by assuming that women are not interested in such positions and then actively generates a specific raw gender bias, which further exacerbates the situation (Devlin, 2017). In this context, Buolamwini and Gebru examined how facial recognition algorithms could create a clear racial bias, as research has shown that some algorithms determine that African-American women are less capable of working than Caucasian women (Buolamwini & Gebru, 2018). The discrimination found in the above two studies is generated by the process of using data to train algorithms by machine-learning developers.

Another observation about the reason of racial discrimination and racial discrimination is about the input of algorithm. Crawford speculates that since the majority of algorithm developers are white males, there is inevitably a "white" gene etched into them, and it is often difficult to detect. It is also difficult to verify because the algorithms used to make HR decisions are often stored in black boxes of proprietary code, and technology companies are reluctant and unwilling to disclose this core code to the public (Pasquale, 2015).

5.7 Criminal (Justice)

This prediction system refers to historical crime data and tends to contribute to the risk of recurrence of improper

enforcement (Koepke & Robinson, 2016). Vendors shield the technology in secrecy, and informed public debate is rare. Some algorithmic processes within the justice system are suspected of discriminating against African-Americans, sometimes judging them to have a higher likelihood of committing dangerous crimes, but the opposite may be true (Angwin, Kirchner, Larson, & Mattu, 2016).

Correctional Offender Management Profiling for Alternative Sanctions (The COMPAS) algorithm is a means used by judges to predict whether defendants should be detained or released on bail pending trial. A report from ProPublica shows that it was found to be biased against African-Americans (Brennan, Dieterich, & Ehret, 2009). The algorithm determines the defendant's risk level for future offending and assigns a corresponding risk score based on a large amount of data such as the defendant's household registration, arrest record, family situation and other variables. Among people who are equally likely to re-offend, African-Americans have a higher probability of receiving a higher risk score than whites, thereby prolonging the time that such people are detained pending trial (Corbett-Davies, Pierson, Feller, Goel, & Huq, 2017). Finally, the facts show that among African Americans, only 20 per cent of those predicted to commit a violent crime actually do so.

The presence or absence of a criminal history among a defendant's friends is one of the assessment criteria used by the COMPAS risk assessment algorithm to predict whether a defendant has committed a crime. Thus, if a defendant has friends with a criminal history, he will be caught in a vicious circle where he is judged by the system to have a higher likelihood of committing a crime, and the likelihood of being sentenced to prison will increase. Thus, the mere fact of relevance increases the number of people in a given population who are likely to commit a crime (Grgić-Hlača, Redmiles, Gummadi, & Weller, 2018).

Angwin, along with other scholars, has demonstrated the inaccuracy and discrimination of the crime prediction algorithms predominantly employed in the US criminal justice system. For example, experiments have demonstrated that both black and white subjects are at risk of being misassessed, with blacks having a much higher risk of reoffending than whites. Although race itself was not one of the reference criteria used to generate this risk score, the other reference criteria and characteristics used, such as information about interests, family beliefs, and family members, were all highly race-related factors, which explains the differences in accuracy of the test results between races. In addition, the risk of reoffending is often underestimated in the social life of white defendants, which is not the case, as is rarely the case with black defendants. Their findings are summarized in Table 1 (Angwin, Kirchner, Larson, & Mattu, 2016).

Table 1. The examples of Prediction fail for white and black defendants

The percentage of prediction fails for white and black defendants		
	White Defends (%)	Black Defends (%)
Being labeled higher risk, but didn't re-crime	23.5	44.9
Being labeled higher risk, yet did re-crime	47.7	28

5.8 Ride Hailing

The prices of taxi apps are automatically generated in accordance with the length of the journey demanded by the user and the supply and demand for taxi services in the area. There is a great deal of concern about the AI algorithmic bias embedded within taxi-hailing programs, as it could have a disparate impact not only on all types of consumers, but also on the future of intelligent automation, resource allocation, intelligent urban automation and resource allocation in cities around the world.

Uber can use artificial intelligence predictive models to determine high-demand locations for dispatching rides by using historical customer demand information collected to predict future demand for rides (Hermann, 2019). Uber's own algorithms calculate fares based on the length of the ride, the distance travelled and the time spent using the program. In addition, the system automatically increases the price if, for example, demand for a trip is greater than the attack, based on local supply and demand, a concept known as "surge ridership". A similar concept is used by the famous taxi company Lyft, which internally refers to it as "prime time pricing" (Banerjee, Johari, & Riquelme, 2016). "Surge ridership varies by separate "surge zones" within a city, with each area having a separate "surge multiplier" (Chen, Mislove, & Wilson, 2015).

Researchers Akshat Pandey and Aylin Caliskan conducted an in-depth study of a sample of 100 million ridesharing rides in the city of Chicago and found that the dynamic pricing algorithms of most ridesharing software con-

tinuously learn AI preferences based on the diversity of the population. The algorithm for measuring bias can be used not only to measure AI bias in carpooling, but is also applicable to examine bias in a dataset of location-based predictive AI models that have continuous outcomes such as pricing. During the period from November 2018 to September 2019, the use of the ride hailing app would have faced higher fares for non-white residents, younger, poorer, and more educated populations in the city of Chicago. In addition, demand and pricing interact, and future demand is also influenced by current and past pricing, which in turn can affect pricing (Pandey & Caliskan, 2021).

5.9 Autonomous Vehicles (AVs)

Many studies have warned that the completely different results of the data mining process can furniture exacerbate algorithmic discrimination in Avs (Barocas & Selbst, 2016) and that algorithmic bias may lead AVs to prioritize the safety of specific groups of road users when making decisions (Selbst, 2017). Meanwhile, there is a warning voice that he manufacturers and algorithm designers prioritized the safety of auto vehicle passengers to maximize profits, by introducing unintentionally or intentionally discrimination against drivers, which is lack of legal frameworks (Liu, 2018).

Benjamin Wilson, Judy Hoffman and Jamie Morgenstern found that the latest object detection system used in self-driving cars had a higher recognition rate for lighter-skinned pedestrians, and conversely a lower recognition rate for darker-skinned pedestrians. Specifically, the technology was five percentage points less accurate in detecting darker-skinned people when factors such as intervals or obstructed vision were controlled (Wilson, Hoffman, & Morgenstern, 2019).

Kevin Todd raised some serious moral and legal questions which cannot be clearly answered till now. For example, if a car is not inherently a safety hazard, but is more likely to run over a black or brown person than a white person, should that car be allowed on the road? What is the benchmark for the safety of such a vehicle? Should the criterion be, "Should an AV be equally likely (hopefully not very likely) to hit any given pedestrian?" Or, "Should an AV be less likely to hit any given pedestrian than a human-driven vehicle?" Given what we know about algorithmic bias, should an automaker be liable for more damages if a car hits a black or brown pedestrian than if it hits a white pedestrian? Do tort law claims, such as design defects or negligence, provide sufficient incentive for automakers to address algorithmic biases in their systems? Or should the government create a uniform regulatory and testing system for algorithmic bias detection in self-driving cars and other advanced and potentially dangerous technologies? (Todd, 2019).

As auto vehicles get smarter, they could pose all sorts of unforeseen problems. Tesla launched version 2021 of its self-driving car software after it led to more than a dozen emergency vehicle collisions, an issue that caught the attention of a federal agency investigation and was cited as a major source of cases to advance the legislative process to regulate self-driving technology. While there are many reasons why these collisions occur, one of the main factors is that the artificial intelligence driving the car can be disturbed by flashing lights and vehicles pulling over, leading the underlying algorithms to react in unpredictable and catastrophic ways (Kim, 2021).

There is another example. Suppose that, in search of convenience, the AI system initially chooses a route that avoids a certain part of town, and thereafter chooses this same route every time. Gradually, machine learning or deep learning may stagnate computationally and always take the same route. This would mean that driverless cars would never appear in other parts of the town, and this is where the lack of flexibility and shortcomings of artificial intelligence come into play (Eliot, 2020).

In order to keep liability claims at a constant level, manufacturers also program the driving behavior of AVs according to the average income of a given area. In simple terms, this means that self-driving cars will be "more cautious" in "affluent" areas than in "economically deprived areas" (Himmelreich, 2018). This could be seen as discrimination based on income levels, as it effectively shifts the safety risk from areas with higher income levels to areas with lower income levels.

In the literature, Hazel Lim and Araz Taeihagh describe Singapore and Japan's approach to mitigating algorithmic discrimination in autonomous driving technology. The governments of Singapore and Japan have issued a series of guidelines that emphasize the explanatory power and transparency of algorithmic programming, but not specifically in the field of AI for AVs, while the EU GDPR explicitly prohibits the use of sensitive personal data for algorithmic decision-making. In the context of AVs, further measures are needed to analyse bias and its impact on security and discrimination (Lim & Taeihagh, 2019).

Identifying discrimination and bias in algorithms is a long way off, as complex machine learning algorithms are

not opaque, but transparency can still be increased by both ensuring traceability of output to input and increasing the interpretability of algorithms, so it is clear that there is still a huge challenge ahead.

5.10 Advertisements

Scholar Latanya Sweeney investigated the role that the element of race plays in Google ads. First, she searched Google pages for common African-American names and noted down the ads that appeared with the results. She then followed the same steps to search for names that were more common among whites. The results showed that ads were more likely to be generated in the results of searches for names that looked like African-American names. In addition to racism, the Google search engine also suffers from gender bias. anja Lambrecht and Catherine Tucker looked at the Google website and found that 20% more men than women had seen the ads. In particular, women aged 25-34 were 40% more likely to have seen STEM ads than women of the same age (Lambrecht & Tucker, 2021).

One of the reasons for this may be that the algorithm has learned the appropriate behaviour from women - women click on ads less often than men. In addition to this, the algorithm may face some sort of sample data shortage, as fewer women have access to advertisements or, alternatively, to the web. Thirdly, the algorithm reflects cultural traditions and underlying mechanisms of discrimination against women in some countries.

A study finds that in 2021 advertisers will use Facebook's advertising platform to collect private data to target cultural interest groups, or groups that use the Lookalike and Special Ad Audience tools, to discriminate on the basis of race and ethnicity. These advertisers also provide evidence that this approach works - demonstrating that the idea of preventing discrimination "through unconscious fairness", i.e., by eliminating the use of protected class variables or proximity agents from the model, does not reduce the likelihood of algorithmic bias (Zang, 2021).

Datta and his colleagues found evidence that when men see relevant ads on Google's platform and try to get jobs that pay more than women, such as coaching services, this not only increases the gender pay gap, but the status of women is further jeopardized (Datta, Tschantz, & Datta, 2015).

Due to the opaque nature of the ad recommendation system, the authors were unable to determine the reasons for this effect; the policy of the advertiser's algorithm may be to tailor the ads displayed according to gender, which is not in itself illegal or discriminatory. The prevailing argument is that men are more likely than women to click on job ads in the coaching services category, so it is the difference in online behavior between men and women that drives the algorithm to show these coaching services jobs to men.

Sweeney, on the other hand, demonstrated that, regardless of whether the name searched for had an arrest record, ads suggesting an arrest record appeared more frequently when searching for black names than when searching for white names. It is unclear why this discrimination occurs, and here are a few reasons: First, it may be that advertisers are defining specific search terms as targets for their ads. Targeting is one of the key tools for accurate advertising and this is not suspected of being illegal. Secondly, it could be because user behavior, such as clicking behavior, has driven machine learning algorithms to make these decisions based on specific names (Sweeney, 2013).

6. Discussion, implication and limitations

6.1 Data governance and algorithm discrimination

Some Scholars suggests audit study: using fictitious correspondence to detect algorithm discrimination detect. This is the most recognized social science method for detecting racial discrimination in employment and housing (Pager, 2007). Specifically, an external auditor would submit a fictitious CV or housing application in the same way as an employed person, and prepare two or more equally valid documents reflecting the same background (including education level and experience) to test the employer or landlord shi'fou, but these documents would vary only by race. For example, researchers can manipulate the fictitious applicant's race between the conditions "Emily" and "Lakisha" in order to issue prospective employers with "Caucasian" and "African-American". Thus, the different reactions of employers to two identical CVs reflect racial discrimination. This method can detect obvious discrimination (such as race and gender), but it is difficult to detect discrimination caused by the company's personalized settings, of which the content is only known to company insiders.

If we explore the source of algorithmic discrimination around the whole process of algorithm and data generation, as Figure 8 shows, we will have more detailed findings. We begin with processing. Pre-processing the data used for training algorithms can be effective in reducing discrimination. Typically, discriminatory datasets are biased in some way, so when they are used to train machine learning algorithms, they will inevitably result in

certain users or groups of users being unwelcome or discriminated against. Once the vulnerabilities within the dataset have been patched as a way to reduce the probability of discrimination, the data can be reused to train any other algorithm type.

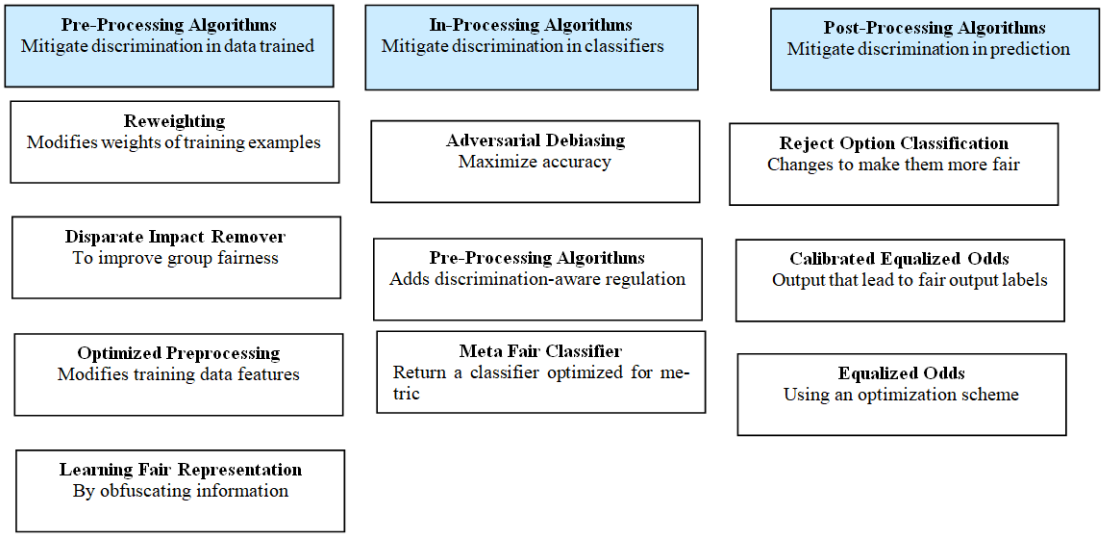

Figure 8. Data Governance Process.

Take an unbalanced dataset, which is a dataset that does not contain a true data distribution and contains more instances of a particular type of user than other datasets. The problem of unbalanced datasets is not limited to numerical discrimination and different techniques have been proposed to deal with unbalanced datasets in the more traditional machine learning and data mining domains that exist, such as: oversampling, which replicates under-represented data elements; under sampling, which removes data elements from over-represented classes; and resampling, which involves swapping labels on data elements. Within the field of digital identification, the proposed approach includes re-tagging data elements to ensure fairness, which has similarities to the aforementioned resampling approach.

Other scholars have proposed eliminating the effects of direct or indirect discrimination directly from the dataset (Feldman, Friedler, Moeller, Scheidegger, & Venkatasubramanian, 2015). Among them, the authors propose the principle of the "80% rule" test to detect the effects of differential discrimination, a principle advocated by the US Equal Employment Opportunity Commission.

Structured data is a type of data that is highly specific and stored in a predefined format. Unstructured data is an aggregation of many different types of data stored in its original format. Structured data is typically stored in data warehouses and utilizes a write model, while unstructured data is stored in data pools and utilizes a read model. Both can use cloud space, but structured data requires less cloud storage space, whereas unstructured data requires more. For the average business user, structured data is the easiest to get started with, as the use of unstructured data requires data science expertise from which accurate business intelligence can be obtained (Talend, 2022). Therefore, for most people, unstructured data is the most challenging data element to solve algorithmic discrimination.

6.2 Human rights and algorithm discrimination

Algorithmic discrimination is not only a violation of human rights, it is also a violation of the right to fairness and equality as required by fundamental human rights.

Take the search engine business as an example. Search algorithms and search engines are not set up to treat all information the same, as creating differentiation is one of the main sources of funding for search engine companies. While processes used to select and index information may be applied consistently, the search results will typically be ranked according to perceived relevance. Accordingly, different items of information will receive different degrees of visibility depending on which factors are taken into account by the ranking algorithm (ARTICLE19, 2016).

While the process of selecting and indexing information is the same, the system tends to rank search results based on perceived relevance. The factors taken into account by the algorithms used for ranking therefore dictate

that different items of information will receive different levels of visibility.

Search engines and search algorithms do not treat all users equally. The different results produced by the system are presented to users in order according to their behavior or other characteristics, including personal risk characteristics. Individual risk profiles can be used to assess a user's insurance or credit score, or applied to more general differential pricing, where different prices are set for the same goods or services based on different consumer characteristics.

If we extend the context of human rights violations leading to inequalities from individuals to companies, the inequalities caused by data discrimination are even more pronounced. Influenced by data aggregation and data analytics, ads from smaller companies registered in less affluent communities are ranked lower than larger companies by search algorithms and search engines, putting these smaller companies at a commercial disadvantage by reducing their exposure.

Some scholars point out that human rights violations can be judged by whether people intentionally violate human rights. This requires us to focus on the process of machine learning to measure at which critical steps proactive intervention is needed to detect human rights violations.

It is also concerned with the parties' perceptions of rights and obligations under different systems of legal language - whether subjective intent constitutes a violation of human rights, or whether facts cause a violation of human rights. In the case of human rights law, it is based on facts. Whether or not a violation is caused is the key to determining whether a human rights violation has occurred. If, instead, the focus is on subjective intent in the framework of algorithmic discrimination, it makes subjective intent difficult to prove and can create a conflict with human rights law. For example, in the non-discrimination law introduced in the EU it is clear that intent to infringe is not a requirement.

6.3 Innovation and Algorithm Discrimination

From a positive perspective, algorithmic discrimination is an innovation in precision marketing, media and retailing. It can help individual consumers find the goods or services they want more quickly, and choose their needs efficiently among the limited goods. It can also improve the efficiency of the enterprise in marketing, promoting the user's second or multiple consumption without additional cost. Diminishing marginal costs will help companies to earn excess returns. There are many examples of innovation changing the consumer sector and even the whole society.

Table 2. Comparison of Structured Data and Unstructured Data

	Structured Data	Unstructured Data
Characteristics	• The model for pre-defined data • Text only (majority of time) • Easy to search	• The model for no pre-defined data • Video, Image, Sound, Text
stores in	• Relational databases • Data warehouses	• Applications • Data Warehouses • NoSQL databases • Data lakes
Created by	Humans or machines	Humans or machines
Typical applications	• Inventory control center • The systems for airline reservation • ERP systems • CRM systems	• Word processing • Presentation software • Tools for viewing or editing media • Email clients
Examples	• Phone numbers • Product names and numbers • Dates • Social security numbers • Credit card numbers • Customer names • Addresses • Transaction information	• Tools for viewing or editing media • Text files • Audio files • Images • Video files • Surveillance imagery • Email messages

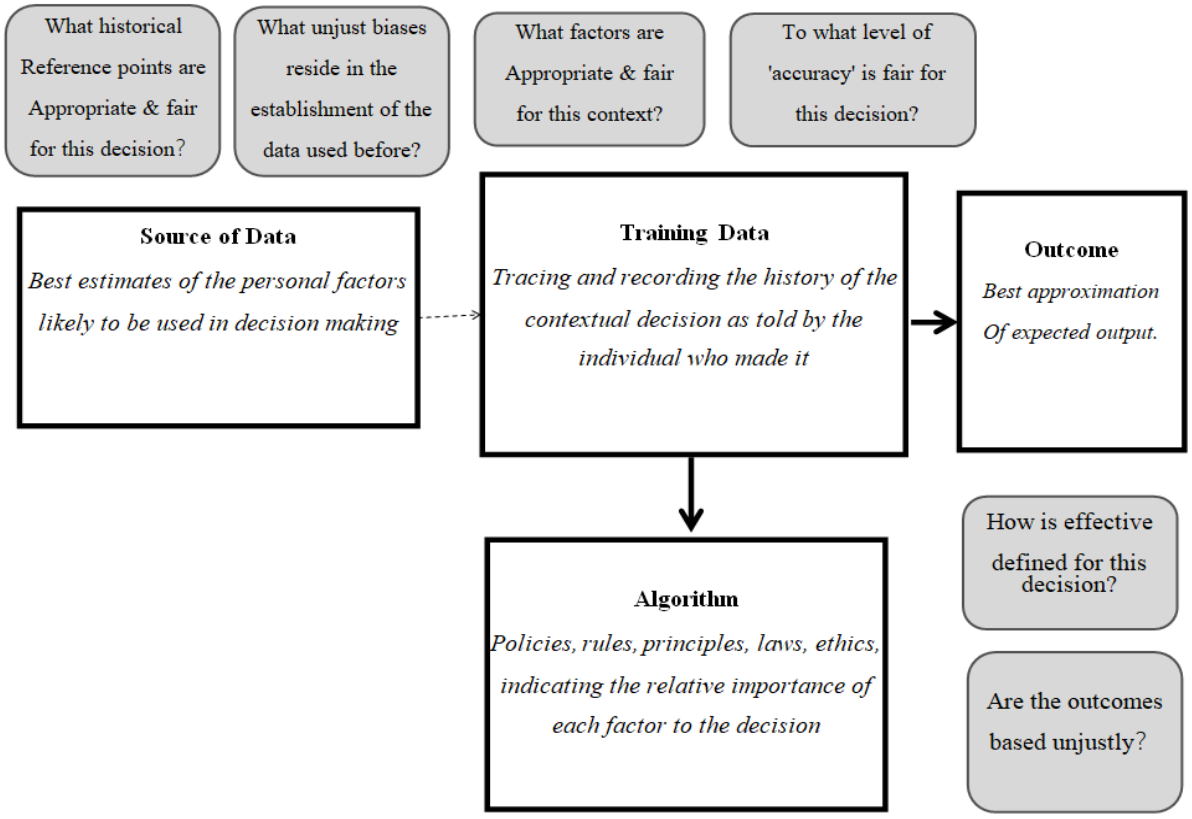

Figure 9. Human rights violation detection path.

Customer experience is enhanced by innovation. Famous technology companies such as Facebook, Google, Facebook, Yahoo and LinkedIn build data products “People You May Know,” “Groups You May Like,” or “Jobs You May Be Interested In” by collecting information about users' education, employment and mutual friends. "Jobs you might be interested in". Bridgestone USA uses a combination of internal data, software provider data and car manufacturer data to advise customers on the choice of repair shops.

Innovation leads to more efficient business management, which promotes innovation in society as a whole. Netflix introduced filtering algorithms to predict customers' movie ratings. Amazon's powerful recommendation engine enables predictive modelling in data products. Trifacta has developed Cloud Dataprep to rapidly capture, process and model data as a way of providing data preparation services to users. Google Data Studio has developed a visual analytics-based data product that analyses embedded visual insights to unlock potential value for improved decision-making. Google delivers targeted advertising for advertisers to reduce the cost of exposure to non-effective users, thereby reducing their unnecessary marketing costs.

Innovation in the field of algorithms, whether for personal consumption or enterprise management, must be highly dependent on the rich soil of customers. Data is generated from increasing customer interactions. Consumers are constantly adapting to the changing digital ecosystem with companies or platforms and their products, and the continuity, interactivity and parallelism of algorithms relies heavily on this consumer interaction. We encourage innovation because of the social efficiency it brings. However, we are concerned about the abuse of innovation, that is, substance discrimination will arise after excessive use.

Some companies see their algorithms as a valuable form of intellectual property, but also to hide the truth from outsiders under the guise of trade secrets and to keep their core interests out of reach (Pasquale, 2010). In response to this proposition, one scholar has further proposed a possible solution: disclose the algorithm to expert third parties who would hold it in safe custody; allow it to be reviewed for public interest reasons, but not made public. The dubious aspect of this solution is that the algorithm is not straightforward and cannot be interpreted to determine whether discrimination exists. Even with audit staff, a large number of algorithm engineers working together to write complex code packages would not allow for effective review.

6.4 Current Regulation and Algorithm Distribution

The US and the EU are leading the way globally in terms of legislative protection against algorithmic discrimination.

In the EU, non-discrimination law and data protection law are the two most relevant legal instruments for protection against algorithmic discrimination. The European Convention on Human Rights states in Article 14. The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, color, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.

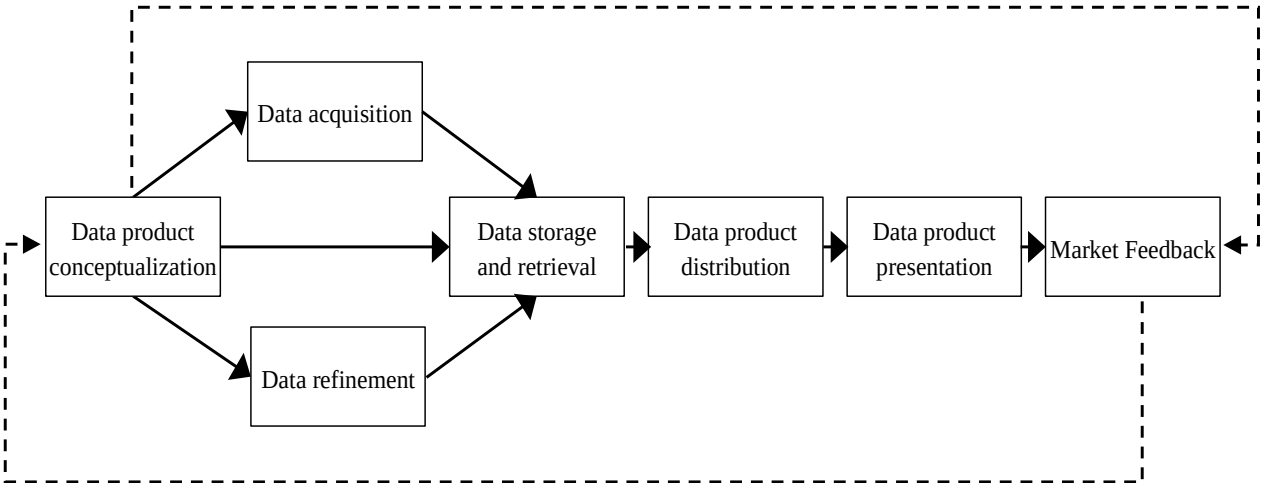

Figure 10. Algorithms and Innovation Loop.

Case law states that the European Convention on Human Rights explicitly prohibits any form of direct or indirect discrimination. **Direct discrimination** is discrimination on the basis of a protected ground (e.g., ethnic origin). The European Court of Human Rights has described direct discrimination as follows: "there must be a difference in the treatment of people, which is based on "identifiable characteristics"."

A similar definition can be found in EU non-discrimination law. Indirect discrimination refers to practices that appear neutral on the surface, but in fact end up discriminating against people of a particular ethnic origin or other protected group. It is worth noting that in the United States indirect discrimination is referred to as "disparate impact"; direct discrimination is referred to as "disparate treatment". The European Court of Human Rights has described indirect discrimination as follows: “a difference in treatment may take the form of disproportionately prejudicial effects of a general policy or measure which, though couched in neutral terms, discriminates against a group. Such a situation may amount to ‘indirect discrimination’, which does not necessarily require a discriminatory intent.”

In practice, the above non-discrimination law has the dilemma that they cannot be implemented. This is partly because it is difficult to prove, or because it gives the algorithm enough room for the discriminator to claim a sufficient objective reason to discriminate. As the European Court of Human Rights has said, the grounds for algorithmic discrimination must be reasonable and objective.

There is a risk that a policy or measure with a disproportionate prejudicial effect on a particular group may be considered discriminatory, even if it is not specifically designed for that group and does not contain a discriminatory intent. However, such a situation only arises when there is no 'objective and reasonable' justification for such a policy or measure. The government is deemed not to be pursuing a legitimate aim if there is 'no objective and reasonable justification' for such an approach, in other words, if there is an unreasonable proportionality between the aim sought and the means employed.

More detailed references can be made to the European Commission's Data Protection Convention 108+ and the EU's General Data Protection Regulation (GDPR). The main rules in both instruments have a high degree of similarity. Among them, the eight core principles are as follows. (a) Personal data may only be processed in a lawful,

fair and transparent manner. (b) Such data may only be collected for a pre-specified purpose and only for a purpose consistent with the original purpose. (c) Organizations should collect or use data within a limited amount. (d) Organizations must ensure that data are accurate and up-to-date. (e) Organizations should not store data for excessive periods of time. (f) Organizations must ensure data security. (g) The organizational decision-maker ("controller") of the purpose and means of processing is the first responsible person to comply with the control provisions.

Similar to our reference to anti-discrimination law, the GPDR also has a power of interpretation. Since there is no legal practice in this part, it is limited to intense interdisciplinary discussions among scholars. On the one hand, scholars such as Edwards and Veale and Wachter doubt the validity of this right and point out that many algorithmic decisions are not subject to the rules of the GDPR (Wachter, Mittelstadt, & Floridi, 2017). For example, the rigor of the extent to which Article 22 of the GDPR applies to algorithmic decisions that are primarily, rather than "exclusively", based on automated processing needs further discussion. For example, if, guided by the algorithmic system's recommendations, a bank employee refuses to extend credit to a customer, then Article 22 would not apply (Kay, Matuszek, & Munson, 2015).

On the other hand, scholars such as Malgieri and Comandé say that the GDPR does provide a right of interpretation (or 'readability') of automatic decisions. As Veale and Edwards point out, the rules in Convention 108 are not as strict as they should be with respect to binding individuals to automatic decision making. Under Convention No. 108, a person has the right "to obtain knowledge of the reasons for the processing of data upon request, where the results of the processing apply to him or her" (Veale & Edwards, 2018). Convention No. 108 therefore does not limit this right to decisions with legal or significant implications. Similarly, we will have to wait and see what effect this different wording will have.

Unlike direct and indirect discrimination in European Law and Regulation Framework, antidiscrimination law in the US is divided into intentional and unintentional discrimination (Selmi, 2011). Intentional discrimination is further divided into two classes, claims where intent is proved through traditional means and claims involving systemic discrimination that generally involve the use of statistics to prove intent. Many of the issues relating to the discriminatory potential of algorithms are not unique to algorithms but are problems that have been a staple of antidiscrimination law. In the cases that have occurred so far, both plaintiff and defendant have relied on anti-discrimination laws to defend themselves, and some decisions have had far-reaching implications for algorithmic discrimination. For example, *Ricci v. DeStefano* involved a promotion test conducted by the City of New Haven for various supervisory positions in its fire service. As a result of the disparate impact, the city effectively discarded the test results by not certifying them, and they were then sued by a group of white and Latino firefighters who, based on their test scores, would likely have been promoted during the two-year life of the list. In the writing for the court in *United States v. Brennan*, issued shortly after Ricci and curiously ignored by the algorithm literature, Judge Calabresi noted that a strong evidentiary basis requests more “than speculation” and “more than a mere fear of litigation, but less than the preponderance of the evidence that would be necessary for actual liability.”

In addition to the difference in definition between the US and EU, there are more specific operational protections in practice in the US. For example, there is legal support for outsiders who want to initiate algorithmic discrimination investigations. Thanks to the protection of the Computer Fraud and Abuse Act (CFAA), the government can prosecute anyone who violates the terms of service of a website. Any researcher investigating how proprietary information systems operate, even if it is to look for discrimination, is likely to violate the TOS.

6.5 Limitations

One thing we must now realise is that we have little control over how our information is collected or used, and very often have to let it happen. While we may be able to intervene before and after an algorithm is created, this does not complete the task of eliminating the occurrence of algorithmic discrimination. In the age of the internet, despite our desire for fairness, the nature of the internet itself dictates that different groups of people will receive different influences and treatment.

Ethical management should be ex ante intervene. We need to set new ethical standards for machine learning and data mining. To put it another way, we may design algorithms in an immoral way. For example, Greyball, a software tool of Uber, aims to predict which taxi drivers may be undercover law enforcement officers, thus enabling the company not to violate local regulations. When Volkswagen accepts the vehicle test, its algorithm allows the vehicle to pass the emission test by reducing nitrogen oxide emissions (Char, Shah, & Magnus, 2018). Protection is not enough if there is only the accountability model after harm.

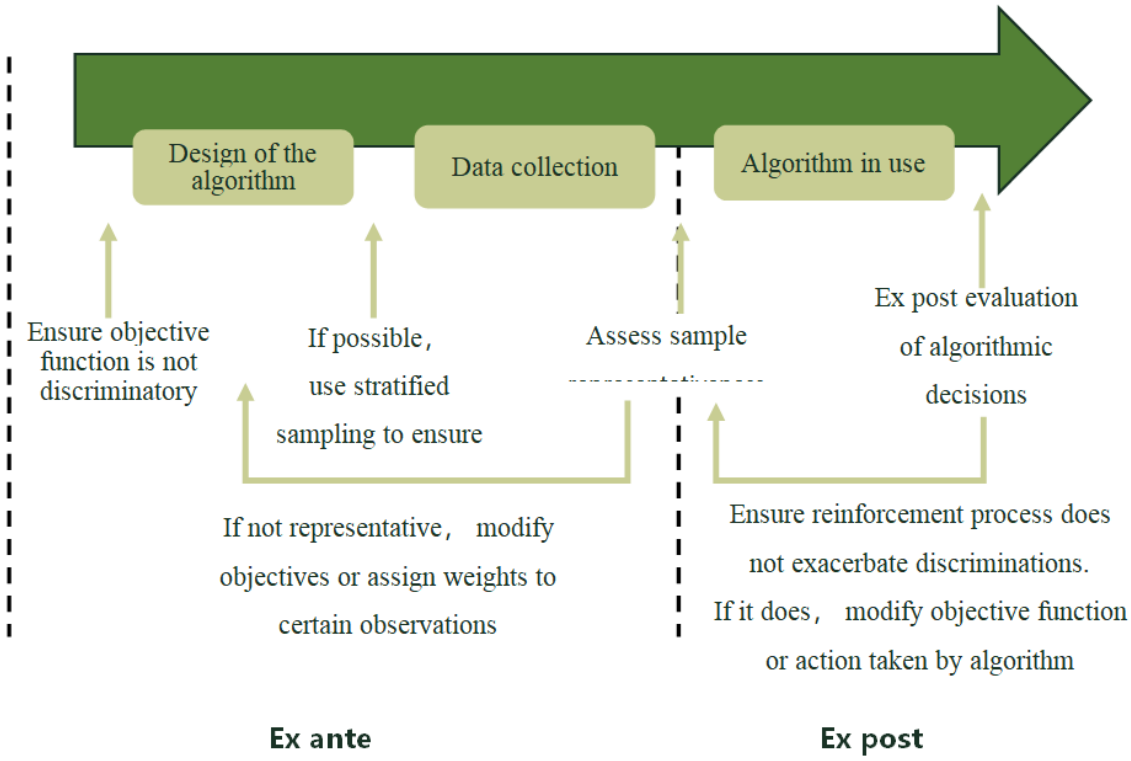

Figure 11. Ex ante VS Ex post Algorithm.

The learning ability of the algorithm is too fast, and it can start a causal self-realization cycle. Scholars have discussed the possible reasons of feedback loop and system loop as unfair prediction (Brayne, 2017). This involves creating a negative vicious circle, that is, some inputs in the data set will lead to statistical deviation, which will be learned by the algorithm and made permanent in the self-realization cycle of cause and effect. An example may help to clarify this mechanism: in some urban areas, the police's crime notification will increase police patrol activities, because the crime notification is considered as an omen of the increase of criminal activities. However, intensive parole will lead to a higher and higher reporting rate of criminal activities in this area, regardless of the real crime rate in this area relative to other areas (d'Alessandro, O'Neil, & LaGatta, 2017). In addition, as a computer language, the algorithm is not intentionally hidden but difficult to detect. Even the experts who have established the algorithm system may not know how the system will behave in practice and have been provided with some data. This aggravates the negative self-circulation effect.

The dynamic managerial capability calls for an innovative and more high-level entrepreneurship and leadership. Dynamic management ability includes management cognitive ability, social ability and human ability, which play a key role in the dynamic ability construction of perception, acquisition and reconfiguration at the organizational level (Helfat & Peteraf, 2003). Managerial cognitive ability refers to the cognitive ability of managers to accomplish tasks that require a lot of cognitive participation, such as the ability to perceive details, solve problems and reason (Helfat & Peteraf, 2015). Managerial social capability refers to the ability to develop contacts and contacts through the social networks of organizations and individuals, allowing effective contact with key information channels, important resources and opportunities, thus bringing competitive advantages to enterprises (Adler & Kwon, 2002). Managerial human capability is the ability to apply skills, knowledge and innovation, which are developed through past experience and educational background (Castanias & Helfat, 2001). A complete dynamic capability must exist in an organization. By integrating new technologies and successful innovations, internal resources and capabilities can be effectively transformed according to the changes of external environment. This is also a management challenge brought by technological innovation.

The effectiveness of self-disclosure by large companies is insufficient. And the third-party auditing is feasible in form but impossible in substance. Google's artificial intelligence has seven principles, the second of which is to avoid creating or reinforcing unfair prejudice. However, only after the algorithm makes a decision will it realize

that there is prejudice and discrimination (Dwork, Hardt, Pitassi, Reingold, & Zemel, 2012). More often, discrimination occurs and is not known until it is investigated. Even if a company allows third-party audits, the auditor has the handicap of not knowing the essence of the algorithm. As for the algorithm itself, its complexity determines that even the algorithm developers cannot clearly judge whether it is subjectively malicious discrimination or objectively specific discrimination. If data discrimination is judged and held accountable only by results, it will do great harm to the overall efficiency of society.

Lacking of Transparency and Traceability leads to challenges. Ex ante inventions for transparency are hard. There are three main reasons. First of all, algorithms are usually complex, because their learning structure is often irregular, and it is difficult to trace and understand. Second, algorithms are being trained to huge data sets. Third, the source code of the algorithm is rarely available. In practice, several active attempts are to make algorithms transparent happen. For example, a decision algorithm in Pennsylvania is being developed by a public institution, which is open to the public for analysis (Smith, 2016). Similarly, a company named CivicScape published its algorithm and data online, so that experts can check the deviation of the algorithm and provide it (Wexler, 2017). In addition, due to the structure and operation of the data agency market, once the data is introduced into the market, it is impossible to "trace any given data to its original source" in many cases. Including the protection of trade secrets; A complex market that "separates" the data collection process from the sales and purchase process; And a large amount of information generated by calculation is mixed with "there is no' real' experience source" and real data (Crain, 2016).

If the data governance against algorithmic discrimination is excessive, it will bring obstacles to commercial civilization and entrepreneurship. Differentiated marking or operation is the product of promoting economic efficiency for both consumers and producers. Algorithmic discrimination is a by-product of the brutal growth of data in this evolution. It seems full of evidence to discuss human rights violations only based on the results of algorithmic discrimination. However, if we step back and look at the whole story of algorithmic discrimination from a global perspective, the logic behind it is the differentiation between business operation and social development. And the essence of differentiation is efficiency. So discussions about algorithmic discrimination, data governance, and human rights are essentially discussions about fairness and efficiency. Equity at the expense of efficiency can lead to setbacks, and efficiency at the expense of equity can lead to serious social problems. Finding the critical point of fairness and efficiency, and continuously optimizing this critical point is the next step in the field of data governance regarding algorithmic discrimination to be solved.

References

Aaron Klein. (2020). Reducing bias in AI-based financial services. Brookings. July 10, 2020.

Aaron Klein. Credit denial in the age of AI. Brookings. April 11, 2019.

Akshat Pandey and Aylin Caliskan. Disparate Impact of Artificial Intelligence Bias in Ridehailing Economy's Price Discrimination Algorithms. 2021. ACM ISBN 978-1-4503-8473-5/21/05.

Alekh Agarwal, Alina Beygelzimer, Miroslav Dudík, John Langford, & Hanna Wallach. A Reductions Approach to Fair Classification, 35th International Conference on Machine Learning, ICML 2018, Stockholm, Sweden, July 2018.

Algorithms and automated decision-making in the context of crime prevention. ARTICLE19.com. December 02, 2016.

Amazon Press Center. (2000). Amazon.com Issues Statement Regarding Random Price Testing. Sep. 2000.

Amit Datta, Michael Carl Tschantz, and Anupam Datta. (2015). Automated experiments on ad privacy settings. Proceedings on Privacy Enhancing Technologies, 2015(1), 92–112.

Andrea Romei and Salvatore Ruggieri. A multidisciplinary survey on discrimination analysis. 2014. The Knowledge Engineering Review, 29(5):582–638.

Andrew D. Selbst. (2017). Disparate Impact in Big Data Policing. Ga. L. Rev., 52, 109.

Angela Chen. (2017). AI picks up racial and gender biases when learning from what humans write. The Verge. April. 2017.

Anja Lambrecht and Catherine Tucker. Algorithm-Based Advertising: Unintended Effects and the Tricky Business of Mitigating Adverse Outcomes. NIM Marketing Intelligence Review, vol.13, no.1, 2021, pp.24-29.

Ankit Gupta. What is Deep Learning and Neural Network. The windows club. November 8, 2020.

Article 5(1) (a)–5(1)(f) GDPR; article 5, 7, and 10 COE Data Protection Convention 2018.Article 5(2) of the GDPR; article 10(1) COE Data Protection Convention 2018.

Arun Kumar. What are Machine Learning and Deep Learning in Artificial Intelligence. The windows club. February 3, 2020.

BBC News. (2000). Amazon's Old Customers 'Pay More'. Sep. 2000.

Ben Naismith, Na-Rae Han, Alan Juffs, Brianna Hill, and Daniel Zheng. (2018). Accurate Measurement of Lexical Sophistication with Reference to ESL Learner Data. Proceedings of 11th International Conference on Educational Data Mining, 259–265.

Benjamin Wilson, Judy Hoffman, and Jamie Morgenstern. (2019). Predictive Inequity in Object Detection. arXiv:1902. 11097.

Bernard Caillaud & Romain De Nijs. (2013). Strategic Loyalty Reward in Dynamic Price Discrimination. Oct. 2013.

Brent Bridgeman, Catherine Trapani, and Yigal Attali. (2009, April 13-17). Considering fairness and validity in evaluating automated scoring. Annual Meeting of the National Council on Measurement in Education (NCME), San Diego, CA, United States.

Brent Bridgeman, Catherine Trapani, and Yigal Attali. (2012). Comparison of Human and Machine Scoring of Essays: Differences by Gender, Ethnicity, and Country. Applied Measurement in Education, 25(1), 27–40.

Brian d'Alessandro, Cathy O'Neil, and Tom LaGatta. (2017). Conscientious classification: a data scientist's guide to discrimination-aware classification. Big Data. 2017; 5(2):120–34.

Caspar Siegert & Robert Ulbrichtb. (2020). Dynamic oligopoly pricing: evidence from the airline industry. Int. J. Indus. Organ. 71, 1026–1039.

Christian Sandvig, Karrie Karahalios, and Cedric Langbort, Uncovering Algorithms: Looking Inside the Facebook News Feed, In the Berkman Center Seminar Series: Berkman Center for Internet & Society, Harvard University. 2014.

Christian Sandvig, Kevin Hamilton, Karrie Karahalios, & Cedric Langbort. Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. 64th Annual Meeting of the International Communication Association. 2014.

Cleber Ikeda. (2021). Do retailers have a recommendation bias problem? RetailWire. Dec 17, 2021.

Constance E Helfat, and Margaret A. Peteraf. (2015). Managerial cognitive capabilities and the microfoundations of dynamic capabilities. Strategic Management Journal, 36(6), 831–850.

Constance E. Helfat and Margaret A. Peteraf. (2003). The dynamic resource-based view: Capability lifecycles. Strategic Management Journal, 24(10), 997–1010.

Cynthia Dwork, Moritz Hardt, ToniannPitassi, Omer Reingold, and Rich Zemel. (2012). Fairness through awareness. arXiv:1104.3913.

Danton S. Char, Nigam H. Shah and David Magnus. (2018). Implementing Machine Learning in Health Care-Addressing Ethical Challenges. 2018 Mar 15: 378(11): 981-983.

David Danks and Alex John London. Algorithmic bias in autonomous systems. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 4691–97. 2017. Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization.

Devah Pager. (2007). The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future. The Annals of the American Academy of Political and Social Science 609, no. 1 (2007): 104- 33.

Donna M. Christensen, Jim Manley, and Jason Resendez. Medical Algorithms Are Failing Communities of Color. Health Affairs Forefront. September 9, 2021.

ECtHR, Biao v. Denmark (Grand Chamber), No. 38590/10, 24 May 2016, para. 89.

ECtHR, Biao v. Denmark (Grand Chamber), No. 38590/10, 24 May 2016, para. 103.

ECtHR, Biao v. Denmark (Grand Chamber), No. 38590/10, 24 May 2016, paras. 91 and 92. I deleted internal citations and numbering from the quotation.

ECtHR, Biao v. Denmark (Grand Chamber), No. 38590/10, 24 May 2016, para. 90.

Elizabeth Dwoskin. (2015). How social bias creeps into web technology. Wall Street Journal. August 2015.

Executive Office of the President. Big Data: Seizing opportunities, preserving values. Washington D.C.: White House. 2014.

Frank A. Wolak. (2016). Designing nonlinear price schedules for urban water utilities to balance revenue and conservation goals. National Bureau of Economic Research Working Paper 22503, 1–41.

Frank Pasquale. (2010). Beyond Innovation and Competition: The Need for Qualified Transparency in Internet Intermediaries. 104 Northwestern University Law Review 105. (2010).

Frank Pasquale. The black box society: The secret algorithms that control money and information. 2015. Cambridge, MA: Harvard University Press.

Frederik Zuiderveen Borgesius and Joost Poort. (2017). Online price discrimination and EU data privacy law. J. Cons. Policy 40, 347–366.

Frijters Paul. Discrimination and job-uncertainty. 1996. Journal of Economics Behavior & Organizations, 36(4):433-446.

Hale M Thompson, Brihat Sharma, Sameer Bhalla, Randy Boley, Connor McCluskey, Dmitriy Dligach, Matthew M Churpek, Niranjan S Karnik, and Majid Afshar. Bias and fairness assessment of a natural language processing opioid misuse classifier: detection and mitigation of electronic health record data disadvantages across racial subgroups. J Am Med Inform Assoc 2021 Oct 12; 28(11):2393-2403.

Hannah Devlin. AI programs exhibit racial and gender biases. The guardian.com. April 2017.

Hazel Lim and ArazTaeihagh. An examination of discrimination and safety and liability risks stemming from algorithmic decision-making in Avs.2019. The Fourth International Conference on Public Policy (ICPP4) June 26-28, 2019 – Montreal, Canada.

Henry Anderson, Afshan Boodhwani, and Ryan S. Baker. (2019). Assessing the Fairness of Graduation Predictions. Proceedings of the 12th International Conference on Educational Data Mining, 488–491.

Hin-Yan Liu. (2018). Three types of structural discrimination introduced by autonomous vehicles. Univ. Calif. Davis Law Rev. Online 2018, 51, 149–180.

J. Kleinberg, J. Ludwig, S. Mullainathan, Ashesh Rambachan. Advances in big data research in economics: Algorithmic fairness. AEA Papers and Proceedings 2018, 108: 22–27.

James Wexler. The What-If Tool: Code-Free Probing of Machine Learning Models. Sep 11, 2018.

Jeremy Hermann. (2019). Scaling Machine Learning at Uber with Michelangelo. Uber Blog. Nov 2, 2018.

Jinyan Zang. How Facebook's Advertising Algorithms Can Discriminate by Race and Ethnicity. Technology Science. October 19, 2021.

Johannes Himmelreich. (2018). Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations. Ethical Theory and Moral Practice. Vol. 21, No. 3, Special Section: BSET Papers (June 2018), pp. 669-684 (16 pages). Published By: Springer.

Josh Gardner, Christopher Brooks, Juan Miguel Andres, Ryan S. Baker. (2018). MORF: A Framework for Predictive Modeling and Replication at Scale with Privacy-Restricted MOOC Data. 2018 IEEE International Conference on Big Data (Big Data), 3235–3244.

Joy Buolamwini and Timnit Gebru. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency, 2018. pp. 77–91.

Julia Angwin, Lauran Kirchner, Jeff Larson, and Surya Mattu. (2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. May 23, 2016.

Julia Angwin, Lauran Kirchner, Jeff Larson, and Surya Mattu. (2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. May 23, 2016.

Katherine J. Igoe. Algorithmic Bias in Health Care Exacerbates Social Inequities — How to Prevent It. March 12, 2021.

Kevin Todd. The Problem of Algorithmic Bias in Autonomous Vehicles. 2019. University of Michigan Law School: Law and Mobility Program and the Journal of Law and Mobility. March 2019.

Kimberle Crenshaw. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. Stanford Law Review, 43(6), 1241–1300.

Lance Eliot. (2020). Overcoming Racial Bias in AI Systems and Startlingly Even in AI Self-Driving Cars. Forbes. Jan 4, 2020.

Latanya Sweeney. (2013). Discrimination in online ad delivery. arXiv:1301.6822.

Laura W. Murphy and Megan Cacace. Facebook's Civil Rights Audit – Final Report. July 8, 2020.

Le Chen, Alan Mislove, and Christo Wilson. (2015). Peeking Beneath the Hood of Uber. In Proceedings of the 2015 Internet Measurement Conference (Tokyo, Japan) (IMC '15). Association for Computing Machinery, New York, NY, USA, 495–508.

Logan Koepke and David Robinson. (2016). Stuck in a pattern: Early evidence on ''predictive policing'' and civil rights. Issue Lab Aug 01, 2016.

Lokke Moerel. Algorithms can reduce discrimination, but only with proper data.2108. Publ. 16 Nov 2018 by IAPP, 2018.

Mark Armstrong and John Vickers. (2001). Competitive price discrimination. RAND J. Econ. 32, 579–605.

Matthew Crain. (2016). The limits of transparency: data brokers and com- modification. New Media & Society, 20(1), 88–104.

Matthew Kay, Cynthia Matuszek, and Sean A. Munson. (2015). Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems April 2015 Pages 3819–3828.

Michael Feldman, Sorelle Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. (2015). Certifying and removing disparate impact. arXiv:1412.3756.

Michael Selmi. (2011). Theorizing Systemic Disparate Treatment Law: After Wal-Mart v. Dukes, 32 BERKELEY J. EMP. & LAB. L. 477, 478, 481–83 (2011).

Michael Veale and Lilian Edwards. Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling. Computer Law & Security Review 34 (2018): 398.

Mitch Smith. (2016). In Wisconsin, a backlash against using data to fore- tell defendants' futures. The New York Times. June 22, 2016.

Neil Howe. (2017). A Special Price Just for You. Forbes. Nov. 2017.

Nicol Turner Lee. Inclusion in Tech: How Diversity Benefits All Americans. Subcommittee on Consumer Protection and Commerce, United States House Committee on Energy and Commerce (2019).

Nina Grgić-Hlača, Elissa M. Redmiles, Krishna P. Gummadi, and Adrian Weller. (2018). Human perceptions of fairness in algorithmic decision making: a case study of criminal risk prediction. ArXiv:1802.09548.

Oliver Hinz, II-Horn Hann, and Martin Spann. (2011). Price discrimination in E-commerce? An examination of dynamic pricing in name-your-own-price markets. MIS Q. 35, 81–98.

Paul S. Adler and Seok-Woo Kwon. (2002). Social capital: Prospects for a new concept. Academy of Management Review, 27(1), 17–40.

Phillip Leslie. (2004). Price discrimination in Broadway theater. RAND J. Econ. 35, 520–541.

Puneet Kaur, Amandeep Dhir, Anushree Tandon, Ebtesam A. Alzeibyg, & Abeer Ahmed Abohassan. (2021). A systematic literature review on cyberstalking: an analysis of past achievements and future promises. Technol. Forecast. Soc. Change. 163, 120426.

Qian Hu and H. Rangwala. (2020). Towards Fair Educational Data Mining: A Case Study on Detecting At-risk Students. Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020), 431–437.

Rebecca Wexler. (2017). Code of Silence. Washington Monthly. June 11, 2017.

Renzhe Yu, Hansol Lee, and René F. Kizilcec. (2021). Should College Dropout Prediction Models Include Protected Attributes? In Proceedings of the Eighth ACM Conference on Learning@ Scale (pp. 91-100).

Retail News Insider. (2014). In-Store Tracking: Personalization Innovation or Privacy Invasion? Retail News. May 2014.

Ricci v. DeStefano, 557 U.S. 557 (2009).

Richard P. Castanias and Constance E. Helfat. (2001). The managerial rents model: Theory and empirical analysis. Journal of Management, 27(6), 661–678.

Rochelle, M. (2019). Press & media. Wasteless.

Rodrigo Ochigame. The invention of “Ethical AI”. The Intercept.com. December 20, 2019.

Romana J. Khan and Dipak C. Jain. (2005). An empirical analysis of price discrimination mechanisms and retailer profitability. J. Market. Res. 42, 516–524.

Ryan S. Baker and Aaron Hawn. Algorithmic Bias in Education. International Journal of Artificial Intelligence in Education (2021).

Safiya Umoja Noble. Algorithms of oppression: how search engines rein- force racism. New York University Press, New York. 2018.

Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. Algorithmic Decision Making and the Cost of Fairness. ArXiv:1701.08230 [Cs, Stat], January 27, 2017.

Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. (2017). Why a Right to Explanation of Automated Decision-making Does Not Exist. International Data Privacy Law, Volume 7, Issue 2, May 2017, Pages 76–99.

Sarah Brayne. (2017). Big Data surveillance: the case of policing. American Sociological Review, 2017, Vol. 82(5) 977–1008.

Seeta Pena Gangadharan. Data and Discrimination: Collected Essays. New America’s Open Technology Institute. 2014.

Shimin Kai, Juan Miguel Andres, Luc Paquette, Ryan S. Baker, Kati Molnar, Harriet Watkins, and Michael Moore. (2017). Predicting Student Retention from Behavior in an Online Orientation Course. Proceedings of the 10th International Conference on Educational Data Mining, 250–255.

Siddhartha Banerjee, Ramesh Johari, and Carlos Riquelme. (2016). Dynamic Pricing in Ridesharing Platforms. SIGecom Exch. 15, 1 (Sept. 2016), 65–70.

Solon Barocas & Andrew D. Selbst. Big data’s disparate impact. Calif. Law Rev. 2014, 104, 671–732.

Solon Barocas and Andrew D. Selbst. (2016). Big data's disparate impact. Cal. L. Rev., 104, 671.

Structured vs. Unstructured Data: A Complete Guide. Talend.com.2022.

Taylor Charles. A Secular Age. Cambridge, MA: Harvard University Press, 2009.

Terrell Mcsweeny & Brian O’DEA. The Implications of Algorithmic Pricing for Coordinated Effects Analysis and Price Discrimination Markets in Antitrust Enforcement. Antitrust, Vol. 32, No. 1, Fall 2017.

The two leading Supreme Court cases on systemic discrimination are International Brotherhood of Teamsters v. United States, 431 U.S. 324 (1977), and Hazelwood School District v. United States, 433 U.S. 299 (1977).

Theodore Kim. Op-Ed: AI flaws could make your next car racist. Los Angeles Times. Oct 2021.

Theodore M. Porter (1996). Trust in numbers. The pursuit of objectivity in science and public life (p. 1996). Princeton, NJ: Princeton University Press.

Thomas Petzinger Jr. (1996). Hard landing: the epic contest for power and profits that plunged the airlines into chaos. New York: Rando, House.

Tim Brennan, William Dieterich, and Beate Ehret. (2009). Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System. Criminal Justice and Behavior, 36 (2009): 21–40.

Tobias Berg, Valentin Burg, Ana Gombovic, and Manju Puri. On the Rise of the FinTechs—Credit Scoring using Digital Footprints. FDIC CFR WP 2018-04.

Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. Available via arXiv:1607.06520v1 [cs.CL] 21 Jul 2016.

Trishan Panch, Heather Mattie, Rifat Atun. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health. 2019 Dec; 9(2): 020318.

Ulrich Leicht-Deobald, Thorsten Busch, Christoph Schank, Antoinette Weibel, Simon Schafheitle; Isabelle Wildhaber; Gabriel Kasper. The Challenges of Algorithm-Based HR Decision-Making for Personal Integrity. Journal of Business Ethics, (2019) 160:377–392.

United States v. Brennan, 650 F.3d 65, 109–14 (2d Cir. 2011); Briscoe v. City of New Haven, 654 F.3d 200, 205–09 (2d Cir. 2011); Maraschiello v. City of Buffalo Police Dep’t, 709 F.3d 87, 95 (2d Cir. 2013).

Virginia Eubanks. Automating inequality. How high-tech tools profile, police, and punish the poor. St Martin’s Publish-

ing, New York. 2018.

YooJung Choi, Golnoosh Farnadi, Behrouz Babaki, and Guy Van den Broeck. Learning Fair Naive Bayes Classifiers by Discovering and Eliminating Discrimination Patterns. In Proc. AAAI Conference on Artificial Intelligence, AAAI 2020, New York, NY, February 2020.

Zen Soo. (2017). Bingo Box to expand its unstaffed store concept beyond mainland China. South China Morning Post. Nov. 2017.

Zhiyan Wu, Yuan Yang, Jiahui Zhao, and Youqing Wu. (2022). The Impact of Algorithmic Price Discrimination on Consumers' Perceived Betrayal. 2022. Front. Psychol. 13:825420.

Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. 2019. Science 366:447–453.

Zied Obermeyer, Brian Powers, Cristine Vogeli, Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25; 366(6464):447-453.