Facultad de Ciencias Económicas y Empresariales
ICADE

TO WHAT EXTENT DO LARGE CORPORATIONS TAKE ACCOUNTABILITY WHEN RACE AND GENDER BIAS IS FOUND IN THEIR DATA SYSTEMS AND WHERE DO DIGITAL HUMANITIES COME INTO PLAY?
TRABAJO FIN DE GRADO

Autor: Alicia Senna-Prime
Director: María Eugenia Ramos Fernández

MADRID | Junio y 2024

Abstract

The purpose of this study is to discover whether large corporations take accountability when bias against ethnic minorities and women are found in their data systems, specifically in AI, algorithms, and Large Language Learning Models. In this case, the focus is on 3 large corporations: Google, Airbnb and Amazon, all leaders in their respective sectors. The literature review gives context as to why this bias in Big Data is an issue that needs to be addressed and how digital humanities can help in future. The research section is comprised of 3 case studies that were analysed to conclude on whether companies take accountability for bias in their systems. First, there is an in-depth description of each case and then an analysis of the actions taken by the company to counter the bias. This focuses on the company culture, their response to public outcry and the measures taken to fix their algorithms. The overall conclusion is that in general companies are slow to take accountability for bias found in their systems and only take responsibility when they are faced with bad publicity or a negative public reaction. The conclusion states that companies need to be more proactive testing their systems for bias in order to eradicate it overall.

Key Words: Big Data, Digital Humanities, Artificial Intelligence (AI), Algorithm, Bias, Black and Minority Ethnicities (BAME), Critical Thinking, Analyse

Resumen

El objetivo de este estudio es descubrir si las grandes empresas asumen su responsabilidad cuando se detectan prejuicios contra las minorías étnicas y las mujeres en sus sistemas de datos, concretamente en IA, algoritmos y grandes modelos de aprendizaje de idiomas. En este caso, la atención se centra en 3 grandes corporaciones: Google, Airbnb y Amazon, todas ellas líderes en sus respectivos sectores. La revisión bibliográfica contextualiza por qué este sesgo en Big Data es un problema que hay que abordar y cómo las humanidades digitales pueden ayudar en el futuro. La sección de investigación se compone de 3 estudios de casos que se analizaron para concluir si las empresas asumen la responsabilidad del sesgo en sus sistemas. En primer lugar, se describe en profundidad cada caso y, a continuación, se analizan las medidas adoptadas por la empresa para contrarrestar el sesgo. El análisis se centra en la cultura de la empresa, su respuesta a las protestas públicas y las medidas adoptadas para corregir sus algoritmos. La conclusión general es que, en general, las empresas tardan en responsabilizarse de los sesgos detectados en sus sistemas y sólo asumen su responsabilidad cuando se enfrentan a una mala publicidad o a una reacción pública negativa. La conclusión es que las empresas deben ser más proactivas a la hora de poner a prueba sus sistemas en busca de sesgos para erradicarlos en general.

Palabras Clave: Big Data, Humanidades Digitales, Inteligencia Artificial (IA), Algoritmo, Sesgo, Etnias negras y minoritarias, Pensamiento Crítico, Analizar

Table of Contents

1.0 Introduction

Data has always been around. However, it was previously in the form of words in a book, stored in libraries. The first mass produced computer was made available in 1957 (Van Es & Masson, 2016); with the emergence of technology, more specifically computers, our ability to store information has improved drastically.

What actually is Big Data? Bean (2021) considers that the term ‘Big Data’ could be applied to all forms of data, no matter the structure or volume. However, Ferreira et al. (2021) states that for something to be considered Big Data it must have one of the 5 following components: a large volume of data, high velocity of data - meaning it must be produced at a certain speed, high veracity - it must be good quality data, a variety of data - this refers to the different formats data can come in social media pages, news articles, online messages etc, finally the data must have value or add value in some form (Ferreira et al., 2021).

As said by Randy Bean the term Big Data only became a common phrase in 2011 even though it had been around for decades; the progression of modern technology at such a rapid pace has made information more accessible and launched us into the era of Big Data that we are now in (Bean, 2021). More data is being recorded every 5 seconds than in the last 20 years and in 2015, 5 billion gigabytes of data was recorded every 10 seconds (Zwitter, 2014). But why is big data even important? Big Data can detect patterns in mass amounts of data where humans cannot (Bean, 2021). What’s more, companies, governments, and other institutions can leverage the technologies and information provided by Big Data, making research easier and faster for much cheaper than previously capable (Bean, 2021).

When we first think of Big Data we may think of technology companies. Big Data is everywhere, social media companies collect your data to show you personalised ads and searches (Gregersen, 2024). Hospitals use Big Data to collect information about patients' medical history and also when transferring data to other medical centres. Companies use Big Data to look at market trends, collect market research and afterwards analyse the patterns in the data. Governments use big data to find out what the public desires and how

the economy can be improved. Educational institutions use Big Data to find better tools to improve student results and satisfaction (Gregersen, 2024).

Digital Humanities is the application of technology or computational tools in the humanities subjects (Burrows & Falk, 2021); for example, the digitisation and visualisation of historical texts (Noguera, 2021). The implementation of technologies in the humanities is a highly relevant issue in order to help advance the fields that they are in and is argued as needing to be taught in universities and schools, however, it is often an afterthought as its importance is often disputed. Regardless, Digital Humanities has been progressing at an alarming rate as it is increasingly being used in all lines of work and research such as history, religion and most importantly literature and languages (Burrows & Falk, 2021). Digital Humanities is now being used in languages and literature to analyse texts of information to find key themes and even to understand poetic expression within complicated pieces of literature and poetry (Jeffrin, 2023).

It is undisputed that Big Data and Digital Humanities are intertwined as they become more necessary and relevant to mundane topics and daily life (Burrows & Falk, 2021). Technology is present in both and as stated by Stephen Ramsay - a member of the: Center for Digital Research in the Humanities (Ramsay, 2024) - you need to know how to code in order to be a Digital Humanist (Burrows & Falk, 2021). Digital Humanities can be the key to analysing patterns within big data. Therefore, it is important that people are taught how to properly use the necessary tools so that they can extract and look at the correct data (Burrows & Falk, 2021), which should start at a school level.

While being taught to examine data is important, being taught to examine the bias in that data should be a priority. Humans are not born with bias, rather it is a trait that children are taught from a young age (Perez, 2020); however, throughout history, humans have repeatedly shown and acted upon bias, which is why it is unsurprising that this is shown in data. If not used with caution, data can magnify cultural and economic bias, which is problematic as people often look at numbers and view them as absolute fact (Z. Chen, 2023) without critically thinking about certain points such as: Who conducted the research? Where did they get their facts from? Did they interview a diverse number of

people? Did their study involve an equal number of men and women? Did they account for their own potential bias? (Z. Chen, 2023). Critical thinking is the way in which a person analyses and interprets a situation or a piece of text (Gosner, 2024) and it is extremely important that humans use their critical thinking and rationale to interpret data, however, first they must be taught the different kinds of bias that can be present in Big Data.

The following data biases are the most common. They include selection bias or bias of omission which refers to only a certain type of people being chosen to take part in a study (Ntoutsi et al., 2020) (for example upper class, middle aged white males) or when the data being analysed simply is not representative of the population as a whole (Perez, 2020). Researchers may choose information or data that is easier to select and more accessible to them meaning that minority groups and women are often left out of the equation (Z. Chen, 2023). Sparse data bias occurs when there is not enough differentiation in the data input meaning the results can be misinterpreted or blatantly incorrect (Greenland et al., 2016). Statistical bias is very common in data groups and is defined by Z Chen in his 2023 paper as "prejudice from assessment criteria that generalise group characteristics to individuals" - see annex 2 (Z. Chen, 2023).

Bias can have extremely negative impacts on the people it affects as it can hinder their capability of engaging in society and the economy (Manyika et al., 2019), which is a disadvantage to society as a whole. Furthermore, it is our responsibility to eradicate bias in Big Data and training models to promote diversity, equality and to improve decision making (Manyika et al., 2019) whether that be in the workplace and in our day to day lives. This paper will discuss the bias that occurs from Big Data, Machine Learning and AI and the impacts that this can have on society. The literature review will focus on Big Data bias in society, highlighting the problems that ethnic minorities and women have when it comes to these biases and the effects this has on daily lives, law enforcement and job opportunities. The research section will discuss 3 case studies; focusing on 3 different corporations and their response to Big Data bias found in their algorithms which had a negative impact on ethnic minorities and women. The literature review will cover different aspects in which bias has affected women and ethnic minorities in everyday life.

2.0 Literature Review

2.1 Bias Against Ethnic Minorities and Women in Big Data

Unfortunately, bias exists in all forms of data, and it is everywhere. One notable example is ShotSpotter, a sensor that was used in Chicago to detect when gunshots occur (Prescott, 2023). The police then leveraged another technology called Hunchlab that used the information detected by ShotSpotter to predict locations of potential gun crime, meaning that the police were able to focus more resources on these areas; these also happened to be in the poorer neighbourhoods which were mainly inhabited by ethnic minorities; however, the police stated that they were able to reduce crime by 24% (Prescott, 2023). And this is not the only way that predictive policing is used in the law system. A computer software programme - COMPAS - was being used to predict whether a convicted criminal was more likely to reoffend; shockingly, the algorithm misclassified black offenders as twice as likely to reoffend than white offenders (Gupta et al., 2021) and were much more likely to be classified as high risk (Gillborn et al., 2017). Ironically, the reason the computer algorithm was put in place was to ensure the exclusion of human judgement and bias (Gillborn et al., 2017). Additionally, online searches for black sounding names were much more likely to provide search results of criminal records even if there was no record associated with the name, the same search with white sounding names did not have any such results (Z. Chen, 2023).

In 1988 St George's Hospital Medical School was proven to be discriminating against student applications if they were a woman or if they did not have a European sounding name (Gillborn et al., 2017), this resulted in women and ethnic minorities reaching the interview stage of the application process significantly less (Gillborn et al., 2017). You would think that making it into the 21st Century means that we would have progressed significantly in fixing gender bias, alas it has not, meaning it creeps into data; and it still occurs even in big companies. When Amazon decided to use AI to help sort through job applications for various positions the algorithm had taught itself that men were the preferred choice over women; this is due to the computing industry being significantly male dominated meaning that if a CV contained the word 'women' it was seen as less favourable by the algorithm (Prescott, 2023). However, there isn't just outright bias to

account for but also hidden bias that is much more difficult to spot. For example, Gild - an online tech hiring platform - allows employers to search through a candidate's online data that is easily traceable (Perez, 2020). Gild's data showed that visiting a Japanese Manga site gave an indication that the candidate would be strong at coding, giving them the edge in the candidate selection process; of course, women tend to steer clear of sites if they include sexist or derogatory imagery; thus, unintentionally discriminating against women in the application process. This is increasingly worrying considering that in the US, 72% of CVs are screened by computers and never make it to the second stage (Perez, 2020).

Biases aren't just affecting the criminal justice system or potential job opportunities, but they are also affecting the health system as well. The NHS uses a deterministic algorithm called HESID in order to match names with patient history (Prescott, 2023). Due to Black and Minority groups having names with more variable spelling a probabilistic algorithm would much more accurately link names with background data. Unfortunately, HESID was unable to match 4.1% of names with details of BAME groups and 0.2% had false matches (Prescott, 2023). Biases are also affecting patient diagnosis. Convolutional neural networks - otherwise known as CNNs - have been trained to offer a potential diagnosis of melanoma in patients with skin lesions (Norori et al., 2021). The issue is this programme is trained with images of predominantly white people and between 5% to 10% of images are people with a darker skin tone in which skin lesions may appear different. Unsurprisingly, the algorithm is much less accurate in diagnosing black patients, which could ultimately prove fatal, and it has as the mortality rate among Black patients with melanoma is the highest out of all groups (Norori et al., 2021).

There are also smaller instances of bias in data, for example Google's targeted ads advertised jobs to women that were much lower paid than the ones they showed men (Ntoutsi et al., 2020). In a study conducted by Northeastern University and USC, Facebook ads were found to have unwanted bias with 85% of people being targeted for cashier jobs being women and 75% of people targeted for taxi jobs, were black (Bogen, 2019). Many recruiter algorithms are taught recruiters' preferences by using previous applicants in which they have interacted with to suggest possible applicants; this can lead

to accidental bias, for example, if recruiters have previously preferred applicants that coincidentally have the same name, the algorithm will teach itself that this name is a preference in applicants. Other algorithms may be used to anticipate which applicants are best suited to roles by searching for other criteria such as online presence, lack of disciplinary action, job performance and awards such as tenure (Bogen, 2019), which of course can be another source of accidental bias if tenure roles are traditionally held by white men from highly educated backgrounds. If computers are used to select potential candidates the AI is making decisions based on what the programmer of the algorithm desires and thinks (Z. Chen, 2023).

As well as algorithms, physical technology is less efficient for people of colour. For example, facial recognition works best on white male subjects and is less accurate on black women; when 5 different facial recognition software were tested, they were all significantly less accurate on darker skinned females with 3 of the technologies having around a 34% error rate for this demographic - see annex 1 (Najibi, 2020). This is concerning, as facial recognition is used in law enforcement where Black people are more likely to be arrested for minor indecencies, subsequently there is much more mugshot data for Black people, which is used in predictive policing. It is also used in border control of many countries which has the potential to discriminate against other minority groups including the Muslim community (Najibi, 2020).

As well as facial recognition, voice recognition also does not work well for BAME accents or dialects (Prescott, 2023). For example, voice recognition tools do not understand African American Vernacular English, a dialect commonly used by Black Americans. They struggle with the ‘phonological, phonetic or prosodic characteristics’ of their speech. Not only this but when transcriptions of the African American dialect were fed into an AI machine to give job recommendations, the AI machine said that the candidate should not be hired, characterising the language used as ‘lazy and aggressive’ (Prescott, 2023). If speech recognition tools are used in situations that require transcripts, such as court hearings, this can have negative outcomes for the defendant (Prescott, 2023). Smart home devices also struggle to comprehend Indian, Spanish and Chinese accents, as well as strong Welsh or Scottish accents with just above 45% of people saying

that their devices could not understand them (Prescott, 2023). The voice recognition software in many cars also has trouble understanding higher pitched voices which of course affects women; upon further research Perez found multiple accounts of women trying and failing to set up the voice activation features and was told to get a man to set up the feature as it would not work for her (Perez, 2020). The voice recognition feature in her mother's car also had difficulty understanding her, but upon lowering the pitch of her voice it worked perfectly the first time (Perez, 2020).

There are a multitude of examples of researchers proving bias in big data and AI tools by using word association. A study conducted by Capitolina Diaz Martinez et al assessed the hidden gender bias in algorithms by feeding words into the system, the most notable and well-known outcome being 'Man is to Woman, Computer Programmer is to Homemaker' (Diaz et al., 2024). Martinez stated that the computer had never been fed information about gender or word association however the algorithm still produced extremely sexist results including: 'Man is to experts and woman is to know-it-all', 'Man is to fidelity as woman is to obedience' and 'Man is to work as woman is to mother' (Diaz et al., 2024). It is shocking that technology can produce such sexist results without being taught anything to do with traditional gender roles and stereotypes.

Additionally, in a paper by Thomas J Misa (2022), bias in various algorithms is uncovered. An algorithm called Namsor was fed articles on computing science where it then predicted whether the author was male or female. The algorithm mistakenly identified 17.6% of authors as male for female whereas only misidentified male authors as female 0.3% of the time - due to there being a lack of female researchers in computer science only 10% of articles included in the study were written by females, meaning they were significantly underrepresented (Misa, 2022).

Finally, AI images are becoming increasingly used and are a very relevant topic. Unfortunately, and unsurprisingly, images reproduced by AI when prompted can include stereotypes or even racist and sexist imagery. A study done by the Washington Post using Stable Diffusion XL - an AI image tool - highlighted a clear bias. When the imaging tool was prompted to show 'A portrait of a person at social services' they produced images

solely of non-white people, with the majority having darker skin. In contrast when asked to show 'A portrait of a productive person' it generated predominantly white males in suits shown to be working a desk job (Chen, 2023). The algorithm also had extreme bias towards women. When prompted to show 'A portrait of a person cleaning' it reproduced pictures only of women completing household chores. Even more disturbing, when asked to generate 'A photo of a Latina' Diffusion XL created images of women in barely any clothes, posing suggestively (Chen, 2023).

2.2 Studies in which Bias in Data is Eradicated

There are a multitude of studies in which researchers try to account for bias in big data. These include analysing data and using critical thinking to determine how biassed the data is. Critical thinkers must look at variables such as where the data came from: is there a wide variety of sources that include ethnic minorities, women and are they correctly represented in the data? (Basili et al., 2017). It is important to look at articles or news sources and think, what is the educational background of the author? Are they accounting for their own intrinsic bias, what sources are they getting their data from and what are their political views? All these questions can help an analyst to pinpoint racial or gender bias in data. However, if people are less used to facing discrimination or bias, then they are much less likely to spot or prioritise accounting for bias in the data (Gupta et al., 2021). For example, in a study conducted by Manjal Gupta et al (2021), women were shown to have a much higher AI questionability than men which is due to the bias that they have been exposed to. We've already seen the problems AI can produce to uphold sexist and racist stereotypes; therefore, it is up to researchers to hold these tools accountable (Gupta et al., 2021). This is common knowledge among researchers and as stated by Carla Basili et al in their research paper - Digital Humanities and Society: an impact requiring 'intermediation':

"A crucial question is how theoretical models and socio-economic pressures affect research evaluation and, consequently, the funding criteria of research projects."- page 49 (Basili et al., 2017).

Computers will always find patterns in data that humans miss, however, there is the worry that if the data set is large enough, patterns will always be found, as a large dataset

increases the chances of there being coincidental similarities in the data. Nevertheless, just because patterns in data can be found does not mean that an AI model can produce recommendations that should be followed; computers do not have intuition and cannot explain why patterns appear, as they have a linear way of thinking. Humans can interpret patterns and see why they occur in the data (Viola, 2023).

Critiquing big data tools may be the simplest way to account for bias, however, in the long run it is not actually eradicating these biases (Viola, 2023). Algorithms need to be changed and correctly trained in order for their responses to not contain bias towards certain groups of people. However, the bias present in our algorithms is due to the 'historical, systemic and societal issues' that have been present in human society long before Big Data came to be (Gupta et al., 2021). Algorithms are what the engineers behind their construction make them, meaning that if the individual who made the algorithm has any bias, then that will appear in the decisions made by the model (Z. Chen, 2023). As stated in the previous section, sometimes companies need to consider what kind of algorithms they are using more carefully; with HESID, in the NHS's case, this was a deterministic algorithm meaning it needs exact data - Date of Birth, NHS number etc - in order to make matches (Gupta et al., 2021). If a probabilistic algorithm was used it would be able to make matches even if there is data missing, and there is proof: for black patients, the probabilistic algorithm had a missed match rate of 2.3% whereas the deterministic algorithm had a missed match rate of 7%. Furthermore, patients from disadvantaged economic groups also benefited from the probabilistic algorithm with a mismatch rate of 2.2% rather than 6.8% from the deterministic. This clearly highlights the difference testing a range of methods and algorithms can do to ensure greater accuracy and reduce the risk of error or bias in algorithms (Gupta et al., 2021). Companies need to test their algorithms thoroughly and push for the betterment of race and gender disparity within the technology industry (Z. Chen, 2023). And there are many other studies that show how algorithms can be altered to provide a more just outcome. A study done from MIT showed how DB-VEA (an AI system) was able to minimise bias by resampling the data meaning that the machine was able to learn physical characteristics, including skin colour and gender without putting them in categories such as race or gender that would encourage bias. Another algorithm named Blendoor allows for equal opportunity of

applicants by removing the names, dates, and pictures on applicant's CVs (Z. Chen, 2023), allowing for applications to be judged purely on work experience and performance.

In the previous section where I discussed the paper by Thomas J Misa, he states that often companies try to account for bias by just removing the problem altogether (Misa, 2022). Misa analysed a study conducted by Jevin D West et al on gender predictions of authors using algorithms. With names that could be used for both genders they decided to remove them completely:

"As with Leslie or Sidney we are unable to identify the gender and do not include that author in our analysis" - page 291 (Misa, 2022).

This just seems like another way to ignore the bias in algorithms as opposed to fixing them. Of course there are better ways to mitigate bias such as, clarifying the data before it is put through a learning model, modifying any decisions after the data has been processed or adding clear definitions to training models so they know how to make fair decisions (Manyika et al., 2019). Other platforms have also attempted to implement tools to detect bias such as Google's 'What-if' tool that helps programmers and engineers to seek out the reasons behind incorrect groupings (Z. Chen, 2023); 'Fairness Flow', created by Facebook, alerts designers if their algorithm has made poor decisions based on race or gender (Z. Chen, 2023).

There are also mathematical ways to account for bias, but this is a much more complicated approach for people who are not trained in statistical methods and reasoning. Jianguo Lü and Dingding Li (2013) conducted research into how a bias correction estimator can predict the bias in a data sample (Lü & Li, 2013). However, this requires knowledge and expertise that not all companies may have or will be willing to invest in.

2.3 Where Digital Humanities Can Help in Future

Discrimination not only has impacts on individuals, but also on society as a whole. For example, if there are individuals who are unable to find work due to hiring bias or educational bias, it can have negative economic implications (Z. Chen, 2023); the economy prospers if there are more people employed and able to take part in the job market. Luckily, it is much easier to fix the bias in machines and in data than to fix the intrinsic human bias that is unfortunately woven into every person and society (Manyika et al., 2019). However, as stated by James Manyika in his 2019 paper, addressing bias also involves 'defining what is fair', and everyone may have their own individual idea on what fair is. Defining what is fair involves discussing culture, language and ethics which is where Digital Humanities can be introduced (Manyika et al., 2019).

Digital Humanities deal with a range of subjects, whether that be culture, linguistics, history and media (Prescott, 2023) all of which are vital in understanding the context of race and gender bias that plagues society. Digital Humanities are of the utmost importance in all aspects of Big Data and if our goal is to fix these biases within algorithms and data collection, then it is key that we incorporate those that work in humanities in the making and mending of them. Language learning, linking records and image processing are all key aspects of the Digital Humanities (Prescott, 2023). For example, those who have expertise in gender studies will be needed to help train large language learning models in order to remove the gender bias that can be formed with word associations (Werthner et al., 2024). Furthermore, Digital Humanities and AI tools will become intrinsically linked; to form non-biassed algorithms and AI tools Digital Humanities expertise is needed and to examine historical studies and find links in the humanities, AI can be leveraged (Prescott, 2023), one cannot thrive without the other. As AI improves, so will Digital Humanities meaning they will be able to track AI performance; as AI tools become more extensive and used in more aspects of life such as legislation and law, it is critical that the current tools used, become less biassed and that their usage is fully understood by everyone so that race and gender bias can be accounted for and removed from our systems (Prescott, 2023). Those who work in humanities can also help devise guidelines for best uses of AI and how to use it responsibly in different aspects of society; Andrew Prescott in his 2023 paper 'Bias in Big Data, Machine Learning and AI: What Lessons for the

Digital Humanities?' numbered 10 ways in which Digital Humanities can help with the progress of AI, these include:

1. Demonstrating how Information Technology can be inclusive and diverse for all.
2. Creating guidelines for the use of AI in all aspects of society including workplace and daily life
3. Explaining the process structure of AI tools
4. Promoting awareness of bias in their own field as well as in Big Data and AI to eliminate bias in algorithms
5. Facial recognition tools and how they are tested
6. Linguistic processing in AI and voice recognition

(Prescott, 2023).

While fixing algorithms would be a step in the right direction, decisions made by AI are based on the information that it has been fed (Z. Chen, 2023); therefore, fixing systemic and cultural bias, whether it is to do with gender or racism, is critical if we are to develop tools that are free of human bias (Gillborn et al., 2017). This means we need to teach society how to be fairer, which starts as early as school; children are not born with bias, it is something they are taught through life. Before starting school, children were asked to draw an image of a person: the number of drawings of men and women were almost even; a year later - after starting school - the same group of children were asked to draw an image of a person again, this time the majority of students drew men (Perez, 2020). If children are taught unconscious bias at such a young age, it seems obvious that it will only get worse as they grow up. Children's textbooks need to include more female historical figures and not just in an 'influential women' section, to ensure that women aren't segregated to a single paragraph while the rest of the textbook only includes men (Perez, 2020), implying that men have accomplished far more than women throughout history. History is a key topic in the humanities; therefore, the digital technologies will be imperative in making sure women's history is included in children's textbooks and is taught in schools and to ensure women's efforts are not completely erased from history. Furthermore, it's important that everyone has access to the same level of education: often by offering top jobs to those that went to elite universities, we are leaving out a huge

portion of the population - often minority groups - that do not have access to the same level or cannot afford the same level of education as their white peers (Gillborn et al., 2017). In the UK, at a glance, it may seem as if minority ethnicity prospective university students are just as likely to attend elite universities as white students - see annex 3 (Gillborn et al., 2017). However, if we split the minority ethnicities into smaller groups, it is clear that this is not the case - see annex 4 (Gillborn et al., 2017).

But what does this have to do with Digital Humanities? Humanities and now digital humanities subjects are the basis of our culture, language and history; if these subjects are inhabited by mainly white or male individuals, only a small section of the population is being represented. Furthermore, different ideas and perspectives come from teams of people that are diverse and have been exposed to different upbringings and cultures (Perez, 2020); therefore, it is important to diversify the workspace of AI and other technologies (Manyika et al., 2019) if we are to combat the bias of these systems; and, as stated previously, people are more likely to spot bias in systems if they have been exposed to it themselves (Gupta et al., 2021), and they are more likely to want to rid algorithms and systems of bias if it directly affects them or people they care about.

3.0 Investigation

My literature review has set up the context to prove how dangerous AI and Big Data can be for marginalised communities, describing in detail the various ways in which it affects day to day life for BAME and women including: work life, law enforcement and health care. For my in-depth research section I have analysed three case studies, each tackling a different organisation as the study will analyse responses of big corporations to Big Data bias proven to be in their system and how they proceeded moving forward. Large corporations are ultimately the people who hold the power in the innovation of technologies therefore it is of paramount importance that they accept the risks Big Data poses and invest in having fairer AI and data; we must hold them accountable for their active bias or they must admit to the bias and aim to fix their system and company culture.

For each case study I have given an in-depth background of the issue that arose within the organisation and how it relates to Big Data and the harm that it had on certain groups of people. I then described the response from the company and whether it could be deemed as appropriate. Using the information from the case study I have then concluded whether they have taken accountability for their actions.

3.1 Case Study 1

Timnit Gebru: “SILENCED No More” on AI Bias and The Harms of Large Language Models by Tsedal Neeley and Stefani Ruper

3.1.1 Gebru’s History and Experience

The motivation behind writing the case study was the treatment of Timnit Gebru - a former employee of Google - who tried to call out the lack of ethical research and diversity within Google and tried to promote research into mitigating bias in large language models before her untimely departure from the company. Timnit Gebru is a renowned expert in AI and a computer scientist. She was the co-lead of Google’s Ethical AI Team before she was fired and is the founder of the organisation ‘Black in AI,’ which aims to increase the representation and inclusion of Black individuals in the sector of AI technologies. Gebru

has had significant achievements within her academic career, including her doctoral research where she developed an algorithm that identified cars from street view images. From this algorithm she was able to discern patterns that allowed her to correlate car ownership with various socioeconomic factors such as household income, crime rates, and voting patterns; she went on to present this research at Silicon Valley. It was during this research where she began to highlight the potential risks of AI technologies, leading her to advocate for ethical AI practices.

Gebru's recognition in the AI community is well-earned. Her innovative approach to data analysis through AI highlighted critical socio-economic patterns and raised awareness about the implications of AI bias. Her contributions extend beyond her technical expertise, as she has been a vocal advocate for ethical standards in AI development. A quote from her thesis emphasises the prevalence of algorithmic bias:

"One of the most important emergent issues plaguing our society today is that of algorithmic bias. Most works based on data mining, including my own works described in this thesis, suffer from this problem." - page 4 (Neeley & Ruper, 2022)

3.1.2 Google, Other Technologies and the Bias Present Within Them

Google has been at the forefront of AI innovation, developing advanced technologies such as the language model BERT, trained on 3.3 billion words. However, it has fierce competition in this area as OpenAI's GPT-3, was trained on half a trillion words. The vast amount of data used to train these models introduces significant challenges in identifying and mitigating biases. As mentioned earlier in this paper, IBM and Microsoft were shown to have significant biases in their facial recognition technology, this was first highlighted by Gebru, along with Joy Buolamwini. They demonstrated that these technologies were less accurate in identifying individuals with darker skin tones, thus perpetuating racial biases.

The biases embedded in AI models are a significant concern because these technologies are increasingly integrated into various aspects of daily life. Gebru raised concerns about how these biases could manifest in AI outputs, potentially leading to discriminatory

practices and reinforcing existing inequalities. Despite the evidence, Google's response to these concerns was inadequate and Gebru was told by Google to focus on the positive aspects of these technologies, often downplaying or ignoring the potential harms. In May of 2023 former employee Geoffrey Hinton stepped back from the company so that he would have the freedom to speak out against AI and the risks it poses. He himself was responsible for many of the technological developments of Google which he now regrets stating that AI has the power to stamp out certain jobs and forge a world in which people will be unable to tell what is factual (Korn, 2023).

Gebru's critical stance on large language models emphasised that the sheer volume of data used to train these models makes it challenging to detect and eliminate bias. She argued for the necessity of ethical scrutiny and rigorous testing to ensure that AI systems do not amplify harmful stereotypes or unfair treatment of marginalised communities. Her work highlighted the importance of being transparent and taking accountability when developing and implementation of AI technologies.

3.1.3 Timnit Gebru’s Experience at Google

During her tenure at Google, Gebru faced significant challenges related to the company's culture and lack of diversity. Google has a history of underrepresentation of BAME groups and women, which can lead to issues such as racism, misogyny, and a certain disregard for bias or equality within different race groups. Gebru's efforts to promote ethical AI and diversity were often met with hostility and dismissiveness. When she and Emily Bender, a linguistics professor at the University of Washington, wrote a paper on the ethical implications of large language models, Google demanded the paper be retracted, citing that it was too pessimistic. Gebru refused to retract the paper without clarification or reasoning, instead offering to make adjustments, or remove her name from the paper; she was ultimately dismissed. This incident highlighted the company's reluctance to acknowledge and address biases in its AI technologies and its broader cultural issues.

Google’s internal culture often marginalised voices that raised critical issues. Gebru's experience is a testament to the systemic problems within the company. When she pointed

out how GPT-3 could reproduce sexist and racist responses, her concerns were largely ignored, and a colleague even made a comment that suggested he understood why she faced online harassment. Such attitudes reflect a broader issue of how diversity and inclusion are managed within the tech company. Paul Lambert, a product manager at Google attempted to fix bias in Google's Smart Compose as it only used male pronouns when referring to investors or engineers; when no bias free solution was found, they decided to ban gender pronouns.

3.1.4 Google's Response

The dismissal of Timnit Gebru led to significant backlash, prompting Google's CEO, Sundar Pichai, to issue a public statement. He acknowledged the need to assess the circumstances leading to Gebru's departure and committed to reviewing the processes and strategies within the company. Pichai emphasised the importance of learning from this incident and improving how the company handles such situations. He also acknowledged that Gebru's departure would have a large impact on the company's least represented communities and the field of AI ethics, underscoring the need for Google to continue making progress in this critical area:

"We need to assess the circumstances that led up to Dr. Gebru's departure, examining where we could have improved and led a more respectful process. We will begin a review of what happened to identify all the points where we can learn - considering everything from de-escalation strategies to new processes we can put in place. Jeff and I have spoken and are fully committed to doing this '…' we need to accept responsibility for the fact that a prominent Black, female leader with immense talent left Google unhappily. This loss has had a ripple effect through some of our least represented communities, who saw themselves and some of their experiences reflected in Dr. Gebru's. It was also keenly felt because Dr. Gebru is an expert in an important area of AI Ethics that we must continue to make progress on - progress that depends on our ability to ask ourselves challenging questions". - page 10 (Neeley & Ruper, 2022)

Pichai's statement was a rare public acknowledgment of the internal issues within Google. However, it also highlighted the gap between Google's public commitments to diversity

and the realities experienced by its employees. While Pichai's commitment to a review was a positive step, Timnit perceived it as disingenuous and insufficient given the gravity of the situation.

3.1.5 Analysis

It is very clear, firstly, from Google's company culture and lack of diversity that it is not a nurturing environment committed to promoting inclusivity and diversity for the BAME community and for women. Not only did white male colleagues make misogynistic comments towards Gebru but also reinforced and endorsed the harassment that she faced online. When researching a topic in which she is clearly an expert and is very relevant, in which she critiques the necessity of large language models and the bias that they contain - including those created by Google - she received no support from management but was told that framing technology in such a negative way is prohibited.

This is clear refusal from Google's management to accept that AI can harm marginalised communities and further refusal to take accountability that there are any issues within their own AI tools in the company. Additionally, Pichai's public statement falls short of addressing the root causes of the issues Gebru highlighted. The company's initial resistance to her research and the lack of support she received reflect a broader reluctance to confront biases within its AI systems and its workplace culture.

True accountability requires concrete actions to foster an inclusive environment and rigorous efforts to eliminate biases in AI technologies. Google's handling of Gebru's dismissal suggests a lack of commitment to ethical AI and its internal practices. To truly take accountability, Google must support and protect employees who raise critical issues, implement comprehensive strategies to address biases, and ensure that ethical considerations are integral to its technological advancements. They must make substantial changes that go beyond surface-level responses. This includes creating a safe and supportive environment for employees to voice concerns, investing in diversity and inclusion initiatives, and committing to transparency in how they address and mitigate biases in their AI systems. Only through sustained and meaningful actions can Google bridge the gap between its stated values and the experiences of its employees and users.

By acknowledging the systemic issues and taking decisive steps to address them, Google can set a precedent for the tech industry, showing that it is possible to innovate responsibly and ethically. The journey towards eliminating bias in AI is ongoing, and companies like Google must lead by example to ensure that technology serves all members of society fairly and equitably.

One critique of this case study is that it was written by an external source and is very clear that it takes a lot of personal anecdotes about one person's experience within the company, of course this means that the results are likely to be biassed and considering Timnit Gebru had a very negative experience of her time at the company it will of course have tainted her perception on how the company acts in certain situations. We must also look at other people's experiences and specific situations in order to get a more accurate view on how they deal with data bias within their internal systems.

3.2 Case Study 2

Measuring discrepancies in Airbnb guest acceptance rates using anonymised demographic data by The Airbnb anti-discrimination team:
Sid Basu, Ruthie Berman, Adam Bloomston, John Campbell, Anne Diaz, Nanako Era, Benjamin Evans, Sukhada Palkar and Skylar Wharton.

3.2.1 Background of the Company (Airbnb)

Airbnb, founded in 2008, is an online platform that allows people to rent out their homes or spare rooms to travellers. It quickly grew into a global phenomenon, offering millions of listings worldwide. The company aimed to create a sense of belonging and to make it possible for anyone to travel and stay anywhere. However, as the platform expanded, issues related to discrimination and bias emerged, leading to significant challenges in maintaining its ethos of inclusivity.

The case study, titled "Project Lighthouse," is written by Airbnb's anti-discrimination team, individuals who are dedicated to identifying and combating discrimination within

Airbnb's ecosystem. Their work is supported by several notable reviewers, including experts from Harvard's Data Privacy Lab, O'Neil Risk Consulting & Algorithmic Auditing, and the Electronic Frontier Foundation, among others. Importantly, the study took place within the US, meaning it only used those who holiday in the US and those who own property on Airbnb's platform in the US.

3.2.2 Bias and the Impact on Guests

Airbnb faced significant scrutiny due to reports of racial discrimination of its algorithm against guests, particularly those of with darker skin tones. This issue was highlighted by the hashtag #AirbnbWhileBlack, where a significant number of guests shared their experiences of having their booking requests rejected based on the colour of their skin. It was found that black guests had a lower acceptance rate in comparison to white guests, creating an unequal and unfair experience for users of the platform.

The impact of this bias was profound, as it not only hindered the travel experiences of many potential guests but also tarnished Airbnb's reputation. Individuals began to mistrust the company and the blatant discrimination which contradicted the company's mission of fostering a sense of belonging. To respond to these claims, Airbnb acknowledged the existence of bias and committed to addressing it through a series of initiatives, including revising their non-discrimination policy and formed an anti-discrimination team. This case study was also a response to the racial discrimination; a way of analysing the situation and coming up with a possible solution.

3.2.3 The Mathematical Approach

To address and measure the bias in its booking process, Airbnb implemented a comprehensive system designed to anonymise data while preserving its utility for analysis. There were 5 steps in their approach which focused on the probability of whether the booking of a guest would be accepted if their personal information, specifically their profile photo, was anonymised or not.

Firstly, the process began with collecting booking data, and seeing whether reservations were accepted or rejected. The data was then anonymised using k-anonymity and p-sensitive k-anonymity methods to ensure that individuals would not be re-identified.

Secondly, an external research partner was tasked with assigning perceived race to guests based on their first names and profile photos. This step was crucial as it allowed the measurement of acceptance rates by perceived race without revealing personal identities to those taking part in the study within Airbnb. Furthermore, names are also a huge indicator of perceived race in the US, so by matching names to faces, it created an idea of how people viewed each individual participant. To further protect privacy, asymmetric encryption was used to ensure that data could flow securely without the risk of re-identification of individuals.

The anonymised data was then used to identify whether anonymising guests' personal information closed the acceptance rate gap between different racial groups. This gap indicated the extent of bias in the booking process.

Lastly, they analysed the impact of anonymising the data to ensure that no important information was lost in the process so that it remained useful. There was a certain level of questioning on whether anonymising the data would allow for retained quality of the data and while it is certain that some data would be lost when anonymised it was important that the data retained its utility in order to be accurate.

3.2.4 The Effect of Anonymising Data on Racial Discrimination

The immediate effect of anonymising data was the ability to identify and quantify the extent of discrimination more accurately. By comparing acceptance rates between perceived racial groups, Airbnb could pinpoint where the discrepancies lay and take targeted actions to address them. This approach also reduced the potential for conscious or unconscious bias by hosts, as the anonymised data did not reveal specific identifying characteristics of the guests.

Furthermore, the system was designed to ensure that the data remained useful for analysis even after anonymisation. The simulation-based framework used by Airbnb demonstrated that although there was some reduction in data utility, the anonymised data still provided enough accuracy to measure and reduce the acceptance rate gap. Meaning that it was possible to balance privacy concerns while addressing discrimination effectively.

3.2.5 Analysis

Airbnb's response to the identified bias can be deemed both proactive and comprehensive. The company not only acknowledged the issue but also took physical measures in addressing it by improving its non-discrimination policies and forming a specialised anti-discrimination team. The implementation of Project Lighthouse, in anonymising and improving its data analysis processes, demonstrated a strong commitment to tackling discrimination on the platform.

However, whether Airbnb's actions constitute full accountability for the racial bias in their system can be debated. While the technical measures and policy changes are significant, the effectiveness of these interventions in the long term needs to be analysed themselves and these measures have not been in place long enough to do so. Additionally, the loss of potential data utility by anonymising guests' personal information, means important information could be dismissed. The success of these efforts must be met with continuous monitoring, transparency in reporting results, and a willingness to adapt strategies based on findings. Additionally, ensuring that hosts comply with non-discrimination policies and that guests feel genuinely included are ongoing challenges that will require ongoing effort and continuous implementation of resources.

Regardless of these limitations, it can be seen that Airbnb has in fact taken great steps in order to ensure equality for their guests and have also taken accountability for their bias in their Big Data and algorithms.

One limitation of this case study is that it was written by an internal source meaning it is likely to be biassed in favour of the company and paint the situation in a less negative light or downplay the severity of the situation. They may also exaggerate to what extent

they actually acted upon the bias and make it seem like they did an awful lot more than what they actually did. In order to overcome this, an external perspective should be introduced, in which they look at the results of implementing anonymised guest booking and enforcing the new anti-discrimination policies.

3.3 Case Study 3

Gender Bias in Hiring: An Analysis of the Impact of Amazon's Recruiting Algorithm by Xinyu Chang

3.3.1 Background of the Company (Amazon)

Amazon, founded by Jeff Bezos in 1994, is one of the most if not the most well-known global tech giant in e-commerce, cloud computing, and AI technology. By 2022 Amazon had more than 310 million users and 2.2 billion visits globally. The company has grown exponentially, employing hundreds of thousands of people globally. In 2017 Amazon's employees were 60% males and 40% females. Amazon has faced scrutiny over gender disparities of their employees in more technical and high paying roles, and they do not publish gender breakdown of those in higher up positions.

Amazon's history with gender bias is reflective of the tech industry as a whole, and not specifically on Amazon, where women are often underrepresented, particularly in technical roles. Although Amazon publicly committed to diversity and inclusion in the workplace, they still have had issues and challenges with data bias within the company, the most notable being their adoption of an AI-powered hiring tool, which inadvertently perpetuated gender biases based on the data it was trained on. This of course raises the question of algorithmic bias, a key topic that I covered in my literature review.

3.3.2 Algorithmic Bias in their Hiring System

The AI-powered recruitment tool was introduced in 2014 and it was eventually scrapped in 2018 due to the significant gender bias shown in the system. The tool was used for automating the first stage of the candidate selection process; it could filter through numerous applicant's CVs and choose the most desirable applicants based on previous

candidate selections made by the company. It was shown to favour CVs that included certain words and phrases that were more commonly used by male candidates, leading to a disproportionately higher selection rate for male applicants.

This had a large impact on external applicants applying for roles within Amazon as only 20% of applicants hired were female in comparison to the 40% of male applicants who were hired. This reinforced the gender disparity of more technical roles within Amazon but also meant that the barriers women face when attempting to enter the tech industry remained firmly in place, as well as not allowing them to progress within the technology field. As if the algorithmic bias wasn't enough, female candidates also experienced biassed treatment during interviews, including questions that were irrelevant to the job role and being offered less salary in comparison to their male counterparts for similar job roles.

The issue drew widespread attention from the public and those affluent within the tech industry experts, who perceived it as a clear example of how AI can amplify existing biases. The revelation also prompted broader discussions about the ethical implications of using AI in recruitment and the need for greater transparency and accountability in algorithmic decision-making.

3.3.3 Amazon's response

In response to the bias identified in its hiring tool, Amazon took action to redeem the situation. The company acknowledged the issue and immediately scrapped the biassed AI tool in 2018. They committed to improving the transparency and fairness of its recruitment processes, including further testing of the algorithms to eliminate gender biases, and increase the accountability of the recruitment team, and setting clearer goals intended to encourage a more equal and diverse workforce. Mentors and training programmes were put in place to encourage female employees and raise awareness about unconscious biases among recruiters and hiring managers. This was in response to the hiring tool which had disadvantaged women within Amazon and the tech industry for 4 years but also in order to foster a more inclusive environment.

The public reaction to Amazon's response was mixed. While some applauded the company for taking quick action once the bias was uncovered, others criticised the company for allowing such biases to exist in the first place and allowing it to go unseen for 4 years. This is a clear example as to why continuous supervision and analysis of data and AI tools are needed in order to ensure it is not disadvantageous to certain communities or groups of people.

3.3.4 Analysis

In this instance, Amazon took a proactive and swift approach to correcting the gender bias in its recruitment algorithm. By acknowledging the problem and taking steps to dismantle the biassed tool, Amazon demonstrated a commitment to addressing the issue head-on. The company's efforts to revise its algorithms, increase transparency, and implement mentorship and training programmes were commendable and necessary steps toward fostering a more inclusive work environment.

While these measures marked significant progress, they also underscored the complexity of eliminating algorithmic bias. Biases ingrained in historical data and human behaviour are challenging to eliminate completely. However, the company's decision to set diversity goals and hold recruiters accountable for meeting these targets was a positive move, emphasising the need for a cultural shift towards greater inclusivity.

It is crucial for Amazon to maintain a long-term commitment to addressing gender bias, including ongoing research, collaboration with diverse stakeholders, and adaptation to emerging best practices in AI and recruitment. Furthermore, transparency with the public about the steps being taken and the progress made is vital for building trust and demonstrating accountability.

One limitation of the case study is that it was written by an external source so they may not have the full intel of how the recruiting process works within Amazon nor have access to statistics on the difference between male and female employment rates in the company.

However, as stated in the text, Amazon does not publish this information therefore it would be impossible for external sources to comment and only people from Amazon can tell us this information, but then there is no way of telling whether the information is correct or not.

4.0 Discussion

These 3 separate Case Studies all show important issues within the tech industry but also highlight the approach that different companies take in dealing with bias in their data and tools. These 3 companies, especially Google and Amazon, are leaders in their sectors and it is extremely important that they take accountability for their actions. Not only this, but as leaders in technology they need to be the ones promoting ethical AI and actively implementing tools and doing research that eradicate bias in Big Data and promoting an equal environment for all their workers. From the case studies, it is clear to see that companies only respond to the bias in their systems and algorithms once forced to do so. Their responses are often prompted by negative public reaction and done as a measure to staunch the outcry that they have caused.

The 3 case studies are very different situations but all critique Large Language Models and algorithms and the background of why the bias occurred is the same; they were not tested sufficiently beforehand and the teams creating these algorithms were not diverse enough. Meaning that bias was often overlooked or spotted at the testing stage, or even thought of as unimportant as it did not affect the creators.

Digital humanities ties into every case study as Google would’ve needed linguists to fix the large language learning models and to help with testing. Airbnb would’ve needed image analysts to help with the testing stage of anonymising data. Amazon would also have needed linguists in order to ensure that the language potential candidates used did not end up being biassed against women.

5.0 Conclusion

The exploration of Big Data, its biases, and its interconnection with Digital Humanities highlights the imperative need to address social and systemic inequalities embedded in technology. The Big Data era has revolutionised the ways in which corporations work and the ease in which we are able to gather data as well as how we analyse it. However, this evolution has also uncovered the biases that technology can amplify if left untested, specifically the cultural biases that have been rampant throughout human history. Biases that affect people of colour and women is an ongoing issue that needs to be addressed as the bias becomes even more prevalent and inexcusable in Big Data.

Governments, corporations, educational institutions, and healthcare systems all leverage Big Data to enhance efficiency and decision-making. Yet, as highlighted by the cases of ShotSpotter, COMPAS, and Amazon's AI recruitment tool, these technologies can inadvertently reinforce biases against marginalised communities, particularly ethnic minorities and women. The biases in predictive policing, hiring practices, and healthcare algorithms exemplify how Big Data can exacerbate existing disparities rather than mitigate them and it is an issue that needs to be addressed by leaders within the technological field as they are the ones with resources to make a difference in how we manage data. These companies need to involve the humanities in their Big Data decisions as they have the knowledge of linguistics, history, culture and media; the main components that affect the bias in Big Data. Digital Humanities can help identify and rectify biases in Big Data, ensuring a more equitable and inclusive technological landscape.

One of the critical steps in mitigating biases is acknowledging their existence and impact. As seen in the case studies of Google, Airbnb, and Amazon, organisations must take quick and proactive measures to identify, address, and prevent biases in their algorithms or data practices. Google's dismissal of Timnit Gebru, Airbnb's Project Lighthouse, and Amazon's scrapping of its biassed recruitment tool illustrate varying levels of accountability and response to identified biases. These cases highlight the necessity for continuous monitoring, transparency, and a commitment to ethical AI development.

Moreover, the role of education in combating bias cannot be overstated. Teaching critical thinking and bias recognition from a young age is essential in fostering a generation that is aware of and equipped to challenge systemic inequalities; and the humanities is a major part in this. By incorporating diverse perspectives and histories into educational curricula, we can begin to dismantle the deeply ingrained biases that shape our society and, consequently, our technologies.

The potential of Digital Humanities in this endeavour is vast. By leveraging computational tools to analyse historical texts, social media, and other data sources, Digital Humanities scholars can uncover patterns of bias and develop strategies to counteract them. Additionally, interdisciplinary collaboration between technologists and humanities scholars can lead to the creation of more inclusive and fair algorithms.

In conclusion, the journey toward unbiassed Big Data and AI is ongoing and multifaceted. It requires a concerted effort from technologists, humanities scholars, educators, and policymakers. By recognising the inherent biases in our data and taking deliberate steps to address them, we can harness the power of Big Data to create a more just and equitable society. Digital Humanities plays a pivotal role in this process, offering the critical lens and analytical tools necessary to navigate the complexities of bias in technology. As we progress, it is imperative that we continue to prioritise diversity, equity, and inclusion in all aspects of technological development and data analysis.

6.0 Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Alicia Senna-Prime, estudiante de ICADE de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado “To What Extent Do Large Corporations Take Accountability When Race And Gender Bias Is Found In Their Data Systems And Where Do Digital Humanities Come Into Play?", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación el resto.:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Constructor de plantillas:** Para diseñar formatos específicos para secciones del trabajo.
3. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 05/06/2024

Firma: _____

7.0 Bibliography

Basili, C., Biorci, G., & Emina, A. (2017). Digital Humanities and Society: an impact requiring 'intermediation.' DOAJ (DOAJ: Directory of Open Access Journals). https://doi.org/10.6092/issn.2532-8816/7196

Basu, S., Berman, Bloomston, A., Campbell, J., Diaz, A., Era, N., Evans, B., Palkar, S., & Wharton, S. (2019). Measuring discrepancies in Airbnb guest acceptance rates using anonymized demographic data. In Airbnb. https://news.airbnb.com/wp-content/uploads/sites/4/2020/06/Project-Lighthouse-Airbnb-2020-06-12.pdf

Bean, R. (2021). Fail fast, learn faster: Lessons in Data-Driven Leadership in an Age of Disruption, Big Data, and AI (T. H. D. Davenport, Ed.). John Wiley & Sons.

Bogen, M. (2019, May 6). All the Ways Hiring Algorithms Can Introduce Bias. Harvard Business Review. https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias?registration=success

Burrows, S., & Falk, M. (2021). Digital humanities. Oxford Research Encyclopedia of Literature. https://doi.org/10.1093/acrefore/9780190201098.013.971

Chang, X. (2023). Gender Bias in hiring: An analysis of the impact of Amazon's recruiting algorithm. Advances in Economics, Management and Political Sciences, 23(1), 134–140. https://doi.org/10.54254/2754-1169/23/20230367

Chen, N. T. K. S. S. Y. (2023, November 1). AI generated images are biased, showing the world through stereotypes. *Washington Post*. https://www.washingtonpost.com/technology/interactive/2023/ai-generated-images-bias-racism-sexism-stereotypes/

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. Humanities & Social Sciences Communications, 10(1). https://doi.org/10.1057/s41599-023-02079-x

Ferreira, C., Merendino, A., & Meadows, M. (2021). Disruption and Legitimacy: big data in society. Information Systems Frontiers, 25(3), 1081–1100. https://doi.org/10.1007/s10796-021-10155-3

Gillborn, D., Warmington, P., & Demack, S. (2017). QuantCrit: education, policy, 'Big Data' and principles for a critical race theory of statistics. Race, Ethnicity and Education, 21(2), 158–179. https://doi.org/10.1080/13613324.2017.1377417

Gosner, W. (2024, April 29). Critical thinking | Definition, History, Criticism, & Skills. Encyclopedia Britannica. https://www.britannica.com/topic/critical-thinking

Greenland, S., Mansournia, M. A., & Altman, D. G. (2016). Sparse data bias: a problem hiding in plain sight. BMJ, i1981. https://doi.org/10.1136/bmj.i1981

Gregersen, E. & The Editor of Encyclopedia Birtannica. (2024). Big Data, Computer Science. Encyclopedia Britannica. https://www.britannica.com/technology/database

Gupta, M., Parra, C. M., & Dennehy, D. (2021). Questioning racial and gender bias in AI-based recommendations: Do espoused national cultural values matter? Information Systems Frontiers, 24(5), 1465–1481. https://doi.org/10.1007/s10796-021-10156-2

Jeffrin, N. J. (2023). Recent Trends in Digital Humanities: a focus on language and literature. *Shanlax International Journal of English*, *12*(S1-Dec), 197–198. https://doi.org/10.34293/rtdh.v12is1-dec.125

Korn, J. (2023, May 3). *AI pioneer quits Google to warn about the technology's 'dangers.'* CNN. https://edition.cnn.com/2023/05/01/tech/geoffrey-hinton-leaves-google-ai-fears/index.html

Lü, J., & Li, D. (2013). Bias Correction in a Small Sample from Big Data. IEEE Transactions on Knowledge and Data Engineering, 25(11), 2658–2663. https://doi.org/10.1109/tkde.2012.220

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

Martínez, C. D., Bou-Belda, E., & Sustaeta, P. N. (2024). Sesgos de género ocultos en los macrodatos y revelados mediante redes neurales: ¿hombre es a mujer como trabajo es a madre? / Hidden Gender Bias in Big Data as Revealed Through Neural Networks: Man is to Woman as Work is to Mother? Revista Española De Investigaciones Sociológicas, 172, 41–60. https://doi.org/10.5477/cis/reis.172.41

Misa, T. J. (2022). Gender bias in big data analysis. Information & Culture, 57(3), 283–306. https://doi.org/10.7560/ic57303

Najibi, A. N. (2020, October 26). Racial discrimination in face recognition technology. Science in the News. https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/

Neeley, T., & Ruper, S. (2022). Timnit Gebru: "SILENCED No More" on AI Bias and The Harms of Large Language Models. In Harvard Business School.

Noguera, J. C. R. (2021, November 8). *Some applications of digital humanities: data visualization, digitization, spatial analysis and representation, social network analysis, and text analysis.* Cultural Heritage Informatics Initiative. https://chi.anthropology.msu.edu/2021/11/some-applications-of-digital-humanities-data-visualization-digitization-spatial-analysis-and-representation-social-network-analysis-and-text-analysis/

Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. Patterns, 2(10), 100347. https://doi.org/10.1016/j.patter.2021.100347

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K.,

Ramsay, S. (2024, January). Stephen Ramsay - about. Retrieved June 3, 2024, from https://stephenramsay.net/about/

Wagner, C., Karimi, F., Fernández, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., . . . Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery/Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery, 10(3). https://doi.org/10.1002/widm.1356

Perez, C. C. (2020). Invisible Women : Exposing data bias in a world designed for men. https://lib.ugent.be/en/catalog/rug01:002787216

Prescott, A. P. (2023). Bias in Big Data, Machine Learning and AI: What Lessons for the Digital Humanities? Scholarly, 17(2). https://www.proquest.com/scholarly-journals/bias-big-data-machine-learning-ai-what-lessons/docview/28429%2008427/se-2?accountid=11979

Van Es, K., & Masson, E. (2016). Big Data Histories: An Introduction. Journal for Media History.

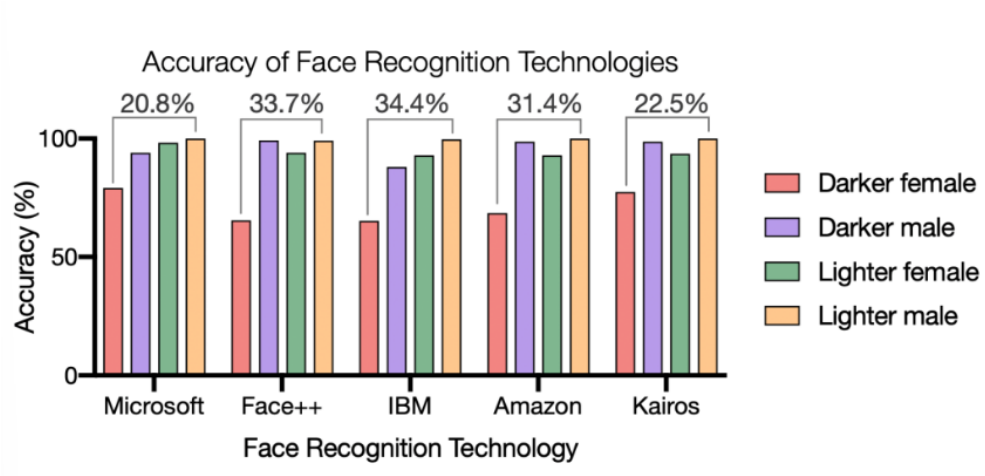
Viola, L. (2023). The Humanities in the Digital: Beyond critical digital humanities. In Springer eBooks. https://doi.org/10.1007/978-3-031-16950-2

Werthner, H., Ghezzi, C., Kramer, J., Nida-Rümelin, J., Nuseibeh, B., Prem, E., & Stanger, A. (2024). *Introduction to digital humanism: A Textbook*. Springer Nature.

8.0 Annex

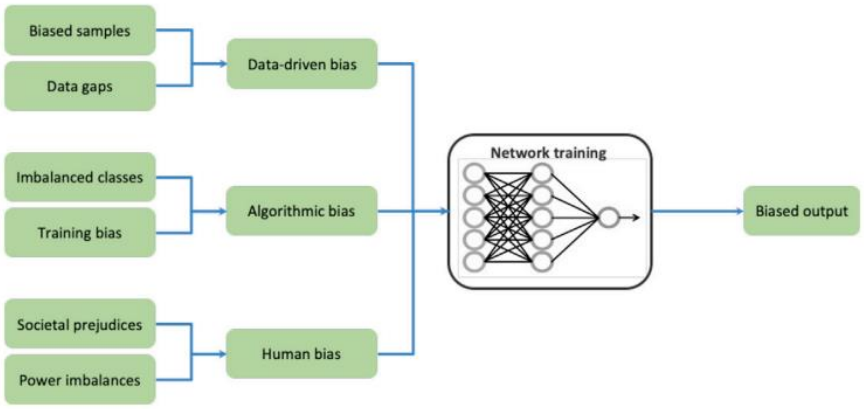
Annex 1

(Najibi, 2020)



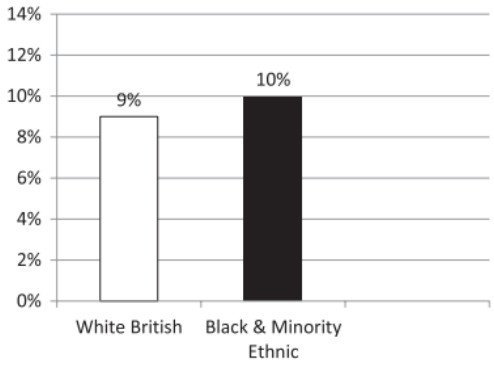
Annex 2

Page 2 (Norori et al., 2021)



Annex 3

Page 164 (Gillborn et al., 2017)



Annex 4

Page 164 (Gillborn et al., 2017)