PHRG
PEACE HUMAN
RIGHTS GOVERNANCE

Volume 6, Issue 2, December 2022

New perspectives on the mitigation of gender bias in AI by EU regulations

Monique Munarini

Research Articles*

DOI:

10.14658/pupj-phrg-2022-2-2

How to cite:

Munarini, M. (2022) 'New perspectives on the mitigation of gender bias in AI by EU regulations', *Peace Human Rights Governance*, 6(2), 111-136.

Article first published online

December 2022

*All research articles published in PHRG undergo a rigorous double-blind review process by independent, anonymous expert reviewers

PADOVA UP

PHRG 6(2), December 2022

New perspectives on the mitigation of gender bias in AI by EU regulations

Monique Munarini

Abstract

Artificial intelligence is part of people's lives, but most are unaware of its presence. The so-called artificial intelligence revolution is one of the challenges that States need to manage. Nevertheless, there is not even a common definition of what can be considered artificial intelligence. Gender bias is an ancient human societal problem that can now be found in AI-driven technologies. The consequences of gender bias in AI are even worse, considering the amplification of its impact on the decision-making process. For a long time, scientists held a belief in the objectivity of artificial intelligence, ignoring this problem. Now the focus is on how to overcome this issue inherited from humans. The European Union has a leading role in developing an ethical alignment of the development of AI with fundamental rights. However, when it comes to addressing and mitigating gender bias in AI, hard and soft laws have gaps that perpetuate gender inequalities. This article aims to provide some intellectual nourishment to explore these gaps and the possible effects on women's rights protection in the Ethical Guidelines developed by the European Commission and the project for an AI regulation, the AI Act. What emerges is that gender bias in AI is a social and technical problem that must be addressed on those two fronts, not only on the technical side. Therefore, legislators struggle to manage it in a way that some change can be "mis en place". To conclude, this article proposes new perspectives to mitigate gender bias in AI, considering the existing and upcoming legislation on the matter. The article offers new perspectives based on a comprehensive approach involving strengthening stakeholders engagement since, with the AI revolution, gender bias in AI has became a cross-border matter.

Keywords: *Artificial Intelligence, gender bias, decision-making process, AI regulation*

* Independent researcher, MA in Human Rights and Multi-level governance, University of Padova and in Law, Economics and Management, University of Grenoble-Alpes; email: monique.munarini@studenti.unipd.it.

Introduction

Literature created the desire to make real the possibility of having interaction between creator and creature, humans and other forms of intelligence. From Frankenstein to the Wizard of OZ., the possibility of artificially creating beings that could reproduce feelings and be considered intelligent has always intrigued readers.

The term artificial intelligence (AI) has become a synonym of evolution, modernity, efficiency. For this reason, many products (Adams, 2021) and services are announced as AI-driven technologies. Artificial intelligence is spread throughout in our daily lives, from the selection of songs we listen to on streaming platforms to the wording suggestions we accept while writing an email. The question is not whether artificial intelligence will be able to perform a task in a specific field, but how it will do it, what the boundaries of a human and a machine intervention would be. The lack of a standard definition of what can be considered AI and consequently an AI-driven technology leads to an imprecision in tackling the advancements and failures in each region of the globe that is focused on this topic. According to the UK Government for Science, the use of AI is not new in our society. However, it is precisely this mainstreaming and massive use that raises interests regarding its potential, limits and mostly the possible harm of using AI (UK Government for Science, 2015).

Even if AI promises to remove human partiality, giving a more efficient outcome (Council of Europe, 2019), the reliance on AI outcomes created 'the veneer of objectivity' (Raso et al., 2018, 7), which was a barrier to identifying issues such as gender-based bias, as this bias was seen as an exclusively human issue.

There are still some blank spaces in AI trajectory, from the absence of broader regulation to the understanding by experts of the whole process of learning and acting coming from AI. However, it is now known that AI has inherent biases from their creators since they unconsciously put their prejudices when selecting data and coding algorithms used to develop AI.

The use or misuse of artificial intelligence is already on the radar of the European Union. The AI Act was proposed in April 2021, and it will be the first international binding regulation on the matter. Therefore, it is essential to analyse the EU response to gender bias in AI and explore new perspectives to address and mitigate, if not eliminate, gender bias in AI.

This article intends to build an understanding of what was happening with the AI development and deployment in Europe regarding gender bias when the European Union started to focus on AI Governance. The selected use cases will help to verify if possible branches of gender equality and non discrimination principles were thoughtfully addressed in the work that lead to the proposed AI act.

The European Union has a strong core of fundamental values that includes the principles of gender equality and non discrimination. They are protected in a multi-level type of regulations. This research focuses on how these principles were or were not translated in AI Governance. Also, it will discuss in which way the threshold of anti-discrimination regulations across the EU support or not the mitigation of gender bias in AI considering the upcoming legislation regarding artificial intelligence.

To conclude, this paper will address what are the possible gaps are within the AI Governance framework within the EU that dialogues or not with the existing legislation protecting women's rights and what could be done to bridge eventual gaps that could escalate discriminatory outcomes based on gender-based bias in AI.

1. A working definition of AI

There are plenty of examples of artificial intelligence used in machines, from gadgets with voice assistants to humanoid robots. The challenge arrives when it comes to defining what this artificial intelligence is, which allows us to talk and get personalized answers from our phones, while at the same time makes us curious about speeche and other interactions that we once thought would only be a result of human intervention. Mary L. Cummings (2017, 7) argues that all definitions are inherently oversimplified since there is no precise definition of what intelligent behaviour is. The Oxford dictionary (2021) defines artificial intelligence as 'the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.' The working group in AI from the European Union (Annoni et al., 2018, 9) refers to AI as 'machines or agents capable of observing their environment, learning, and based on the knowledge and experience gained, taking intelligent action or proposing decisions'. These are very general definitions.

As a theory, artificial intelligence could be considered a field of computer sciences based on algorithms, trying to reproduce some human capacities. As a computer system, artificial intelligence could be considered a technology applied to understand and replicate some capacities that before required only human intelligence. In this sense, it could be any machine or algorithm that has the capacity to reproduce on some scale the human process of acquiring knowledge (Lawson, 2000). In summary, the first step of this human process is to observe the environment. The second is to learn from this observation. The third, based on the knowledge and experience gathered, results in a proposition for a possible human decision or action taken that could replace human activity. The first and second steps of this process applied to machines are called Machine Learning (ML). In this sense, machine learning is a technique that can be used in AI to achieve a level of intelligence that allows the fulfilment of tasks which before were assigned only to humans (Buiten, 2019). Even if these concepts are connected, they are not the same; while machine learning can be part of artificial intelligence, many technologies use machine learning techniques that are not artificial intelligence.

The difference between machine learning and the classical programming technique is that machine learning is fed with data and answers to create rules that will be applied in the analysis of the most significant amount of data, while the classical technique sets as input the data and the rules to obtain the answers. A voice assistant on a smartphone is filled with vocabulary and answers to different questions. In this way, if someone from India or Argentina asks the voice assistant how the weather looks, the voice assistant will "know" that it was required to share the forecast for the region where the person is. The voice assistant also will learn how to recognise the voice of the smartphone owner better to capture the questions better and provide more precise answers. On the contrary, an Excel programme is filled with mathematical data and equations. When someone types an equation, Excel will calculate and replace the formula with the result required. For years, machine learning was a challenge to scientists who could apply machine learning in technology but could not explain the whole learning process and its outcomes. Recently, this challenge has been overcome. The new grey zone is the subfield from Machine Learning, known as deep learning (DL) (Sartor et al., 2020, 13), which 'learns' from a vast amount of data and easily separates what can be a piece of helpful information. Deep learning is considered an imitation of the structure of the neural system since this neural network functions in a way similar to the human brain. The assumption was not to copy the way humans think but to reproduce artificially the structure used by them for reasoning. This way, the reasoning would 'naturally' come from an artificial neural system. The main issue regarding this machine learning approach is that it is still not possible to explain the outcomes of the work of DL. There are some steps of the process that are possible to determine, but still, the reason for certain responses or action taken is still unknown to scientists creating what is called the black box problem (Bathaee, 2018).

Floridi (2019) explains that a dishwashing machine does a better job doing the dishes than some humans, but it is not possible to claim that because of its efficiency the dishwashing machine is more intelligent than humans.

With this analogy he tries to underline that it is not possible to compare human intelligence and behavior with machines.

The OECD, UNESCO and the European Union have adopted similar definitions of artificial intelligence, even if they are not unanimously accepted by the scientific community. The question is not only about finding an agreed upon definition of artificial intelligence to establish the boundaries of AI Governance but also how humans can explain their decisions based on AI outcomes (Coeckelbergh, 2020, 121). If AI Governance is considered the ways in which stakeholders try to mitigate the negative effect of AI, such as through regulation, self-regulation and advocacy, the AI governance is directly connected with the development of a trustworthy AI (Comandè, 2020). Furthermore, there is a whole new world around artificial intelligence that challenges national and international legal frameworks. However, a working definition of artificial intelligence is needed for this article. As a shortened version of the definitions presented, artificial intelligence can be defined *as an umbrella term for both a scientific field and resources with interdisciplinary implications applied to the development of AI-driven technologies that can use data collected from the environment where it is inserted in to improve assigned tasks, achieve goals or propose human decisions.*

2. The artificial intelligence revolution

Every sector in society makes decisions based on data to optimize their results since every action online and offline has started to be turned into data (Pentland, 2013). Judges use jurisprudence as a source of guidance when making decisions, the business sector uses data to understand what the market wants and needs to be the priority. Even the media content we consume online during our free time is the result of data we produced while using streaming platforms. Everyday tons of data are produced, and its meaningful processing would be impossible without artificial intelligence (AI).

Indeed, artificial intelligence is transforming interactions, environments and reshaping lives into a digital world called by Corinne Cath (2016). The search for efficient outcomes, or optimization, is not an exclusivity of the dominant companies in the tech industry. The democratization (Wang, 2021) of artificial intelligence and its use as a service (Cobbe and Singh, 2021) is an ongoing trend, so that also small and medium enterprises in many sectors are also acquiring AI-driven technologies to be up to date in the market.

The artificial intelligence Revolution (Harari, 2017) with increasing reliance on AI-driven technologies in the decision-making process in all categories of society challenges compliance with international human rights obligations. It raises questions in very diverse sectors, as happened during the Industrial Revolution. In both cases, new tools brought problems and challenges that affected or are affecting, in the case of the AI Revolution, society, the economy, and the political model (UK Government Office for Science, 2015). Taking into consideration the United Nations Sustainable Development Goals, artificial intelligence paves the way for positive impacts related to economic growth (SDG 8), industry, innovation and infrastructure (SDG 9); responsible consumption and production (SDG 12) and partnerships for the goals (SDG 17). However, the cost is the reverse impact on inequalities among nations, peoples and genders (Vinuesa, 2020).

Considering the great potential of artificial intelligence, McKinsey Global Institute released a report attempting to predict the impact of artificial intelligence on the global economy. The result was that at the same time that artificial intelligence could contribute to the economy on a large scale, it could also increase gaps between developing and developed countries or between companies and workers. The additional economic growth was predicted to be around thirteen trillion dollars by 2030, translating into an increase of 1.2 per cent in GDP growth per year (Bughin, J., Seong, J., Manyika, J., Chui, M.,Joshi, R., 2018).

This reasoning was considering the non-linear growth of artificial intelligence that started as the new scientific promise from the latter half of the twentieth century but faced many technological struggles by the end of the cold war, gaining the attention of investors only after the rise of the internet and big data. The report Artificial Intelligence Global Market 2020-30 (Research and Markets ltd.,2021) stated that COVID-19 was an even more significant opportunity for the artificial intelligence market considering the overall development of technologies. It was suggested that the global artificial intelligence market was expected to face an economic growth of almost 45%, jumping from 28 billion dollars in 2019 to around 40 billion dollars in 2020.

The increasing reliance on AI-driven technologies in the decision-making process in all categories of society challenges compliance with the human rights legal framework. The core principles of social justice and equality are connected to the idea that all must enjoy human rights. Even if countries treat their data differently, they must be accountable for the effects this data processing can make when used in the design or training of AI since the positive and negative outcomes will not be equally distributed among different peoples. This directly affects human rights protection.

Artificial intelligence is already affecting the enjoyment and protection of human rights, such as freedom of expression when it comes to social media algorithms or non-discriminatory practices coming from decisions made by or with the support of AI in the judiciary. Considering the numerous types of AI-driven technologies, the level of their impact on human and fundamental rights, positive or negative, depends on several factors such as the use or misuse of this technology, their complexity, effects, scale, and accuracy (Council of Europe, 2019).

According to the study Artificial Intelligence & Human rights: opportunities & risks, from the University of Harvard (Raso, F., 2018), all international human rights framework is challenged by the deployment of AI since there are positive and negative effects in the enjoyment of civil, political, economic, cultural, and social rights with the use of AI. Also, determining those impacts is not easy since AI-driven technologies are applied in the real world, which is not a neutral environment or one with full respect for those treaties.

As highlighted by the UN Human Rights Chief, Michele Bachelet, AI can be a force for good to help society evolve, but it can also have catastrophic consequences regarding human rights violations. For this reason, she believes that the higher the risks in the deployment of AI-driven technology, the stricter the legislation should be (Bachelet, 2021).

However, there are scientists who argue that the use of AI-driven technology is out of control, not because there is no definition of AI, or an agreed upon standard to monitor it, but mostly because it is being used everywhere and for a wider range of purposes that do not require such advanced technology. Bern Carsten Stahl (2021, 24) stated that there are two main purposes for developing AI: improvement of efficiency, also known as optimisation, for-profit maximisation, and social control. He proposes the development of AI for human flourishing to balance the two purposes mentioned above to help individuals build a 'good AI society'. The concept of human flourishing comes from Aristotle and his virtue ethics (Stahl, 2021, 22). As announced by the High-Level Expert Group on AI, the idea is that AI is a tool to increase human flourishing, enhancing 'individual and societal well-being and the common good, and bringing progress and innovation' (AI HLEG, 2019, 4). Since AI is a powerful tool, the ones who connect AI with human flourishing argue (Charles, M., Clark, A., and Gevorkyan, A. V, 2020) that AI should be used to help or at least not to interfere in the development of human functioning, in this way also protecting people from human rights violations. Instead of giving an ordinary functioning to AI, give a meaningful one for human development.

Cath et al. (2016, p. 5) examined three police papers from the United States, European Parliament and the United Kingdom. They concluded that there was a common agreement that States could define a 'good AI society' as one in which digital and real-world can evolve together.

However, even though artificial intelligence can bring about this revolution in our society by improving processes and actions, this optimisation in how our society is performing comes with old problems rooted in society or an extension of our existing culture (Bryson ,2017). Hence, bias can be considered one of the most problematic concerns regarding artificial intelligence because beyond just fixing the algorithms, society needs to be fixed to effectively mitigate this problem.

3. Gender bias in artificial intelligence

Bias is a natural response from the human brain. Humans receive tons of information every second, and in order for the brain to be able to make a decision, it uses these judgments created by one's cultural, social and family experiences. It would take a long time to consider all the elements of a situation in order to act. Bias is not necessarily a fundamentally negative concept, but it is a way that the human brain is found to process more information. The human neural system is able to consciously process only 405 pieces (bits) of information out of the 11 million that are received every second (Markowsky, 2017). Gender bias can be considered as one of these judgements made by our brain, which influences our decisions and is based on gender perceptions obtained from our cultural experiences and societal beliefs. These factors create unwanted preferences against one group considering their unique attributes resulting in the basis of systematic discrimination. These prejudices or unwanted preferences are reproduced daily until they become rooted in the culture of a society or organisation.

A gender-biased decision, however, is not that easy to identify, and it is difficult to change the pattern. The Council of Europe (2021) explains that bias is challenging to overcome because it is easier to fail to notice or not realise something that is the opposite of what we understand as right. After the first judgement, human brains intentionally try to find justifications to motivate gender prejudice rationally. It also happens that 'if contradicted by facts, we would rather deny the facts than question them ('but he is not a real Christian'; 'she is an exception)' (Council of Europe, 2021). Artificial intelligence was introduced in some part of the decision-making processes as a response to human issues. AI could improve results by being faster and more impartial, unbiased, since it would not display preferences or stereotype as humans do. For many years scientists believed this, and the market replicated this belief. In the same way that artificial intelligence can improve actions and decisions that were mainly relegated to humans, however it can also expand the damages.

The analysis of modern AI automation's impact on society requires interdisciplinary work from STEM (Science, Technology, Engineering, Mathematics) areas regarding the economy, psychology, political science, and law since it also has an interdisciplinary impact on society. Stahl named this system of interdisciplinary interaction as the AI ecosystem (Stahl, 2021, 84), considering that many diverse stakeholders are involved in a complex relationship within the AI lifecycle[1]. One of the most expressive cases of gender bias in AI was the one from Amazon (Dastin, 2018). The company developed an AI-driven technology to pre-screen applicants' resumes which understood that women were not suitable to work in the company because of their gender. When the AI created by Amazon established its reasoning, it started to apply the same outcome to all the resumes, and this created an escalation of gender bias. The escalation of gender bias connected with AI-driven technologies' outcomes was helpful to confirm that the source of the problem was not only the algorithm behind the development of AI-driven technologies but intrinsic to society, or to the organisational environment where the AI was inserted in. The AI ecosystem helps to understand that technologies, in this case AI, are not only functional tools, but they also perform a cultural and social role when producing and reproducing meanings (Coeckelbergh, 2019). Therefore, gender bias in artificial intelligence can be seen as a social-technical issue (Stahl, 2021, 22) that needs more than technical solutions from the industry that produces it.

Since the beginning of the internet, the idea of searching for information has been associated with the worldwide web. Google played a crucial role in web search, dominating this field and becoming one of the biggest companies in technology (Gebru, 2020). The quality of this customisation of Google's services started to be questioned when biased results were verified. In 2015, scientists from Carnegie Mellon University created a tool called AdFisher to study Google's Ads settings in order to identify how Google distinguishes groups when showing job advertisements. This study analysed four topics: non-discrimination, transparency, effective choice, and ad choice. The idea was to check if changing a protected attribute, such as gender, in users would result in a biased treatment. The AdFisher tool was responsible for analysing more than 600.000 advertisements, and the scientists conducted twenty-one

[1] The AI lifecycle is a concept refined by Coeckelberg (2020, 121) to establish three key moments when searching for answers regarding artificial intelligence: design, test and application. These moments are related to the data set selection for training and testing the AI; to the algorithms, to the workforce involved in the design of AI and in its application.

experiments with the analysis results (Datta, A., Tschantz, M.,C., Datta, A., 2015).

They found that users identified as females were exposed to fewer high paying jobs than males. Considering the intellectual property involved in designing Google Ads, it was not possible for them to identify if the source of the discrimination was rooted in Google, in the advertiser or the relation between them. It was interesting to notice that in their conclusion, they did not state any violation of non-discrimination laws since it was not possible to find who was responsible for this result. The researchers mentioned that, since it was not possible to verify if it was caused by a joint effort from the advertiser and Google, or from one or another, they left this research as a base for internal audits. However, their prediction was that it was more likely that Google lost control of their AI. The study was conducted in 2014 and released in 2015, and by that time, the authors stated that 'we hope future research will examine how to produce machine learning algorithms that automatically avoid discriminating against users in unacceptable ways' (Datta, A., Tschantz, M., C., and Data, A., 2015, 110).

In 2016, Microsoft released chatbot a named Tay on their Twitter account. The intention was to simulate a teenage girl and to get information from Twitter users about their preferences. However, Tay started to receive discriminatory messages and reply to them based on information collected in social media from previous tweets. The result was that Tay stated that Hitler was right, Jewish people were responsible for 9/11, and feminism was a disease. Tay was shut down less than twenty-four hours after it was launched (The Guardian, 2016). Peter Lee, Microsoft's spokesperson, stated that they were victims of a coordinated attack of people who 'exploited a vulnerability in Tay' (Microsoft, 2016).

Even if in recent years companies have started to fail in mitigating gender bias in artificial intelligence, it was only in 2019 when UNESCO released a policy about the digital gender gap, exposing gender bias in Apple's voice assistant, Siri, that this problem became a key concern regarding the use of artificial intelligence. The policy paper 'I'd blush if I could' denounced how gender bias affected digital assistants, from using female digital assistants as default reinforcing gender stereotypes to the need for the inclusion of women in the development of AI-driven technologies to avoid biased outcomes. The title was a reference to Siri's answer when a user tells her 'Hey Siri, you are a b**!' the answer given by the voice assistant was 'I'd blush if I could' (UNESCO, 2019). The digital assistant was created in 2011 and it was only after the exposure in UNESCO's report that the response was fixed.

Bias and discrimination are societal problems that existed before and indepedently of AI, but these social issues impact AI. The way artificial intelligence is designed and programmed is conceived as a technological problem because it can be solved with rational solutions, but this does not mean that they alone can eliminate gender bias in AI. The absence of proper monitoring and auditing of the deployment and creation of AI also leads to uncertainty since, in most cases, gender bias was only identified when already impacting women's rights.

According to Coeckelbergh (2020, 128), application in the real world might escalate bias when people rely mainly on AI outcomes to make a decision instead of trying to get the big picture. Therefore, when creating regulation for AI that addresses the gender bias problem, it is crucial to verify if this legislation embraces all sections in the AI lifecycle. As already mentioned in this article, the veneer of objectivity in which scientists relied on AI outcomes made problems related to gender bias in the past look like a technical issue. However, in all cases, it was not a matter of only fixing the programming since all AI that presented problems of this nature were discontinued, as in the presented examples of Amazon and Microsoft.

The first international regulation to AI was proposed by the European Commission (2021) on April 21, 2021, the AI Act. In this regard, the European Commission (2020) recommended that AI should be developed in a bias-sensitive way, otherwise it would exacerbate existing stereotypes and biases, increasing the negative social and economic impact. However, these biases were not created out of nowhere, they are a reproduction of societal biases in the environment where AI was trained and deployed. In order to be considered trustworthy (HLGE, 2019), AI needs to be lawful, ethical and robust. The robust component is not only from a technical part, but also social to avoid harm.

4. Milestones in AI Governance within the EU

There is no legislation in the European Union framework specifically regulating AI in respect of the principles of equality and non-discrimination. This section intends to highlight some of the milestones in the AI Governance that contributed to bringing to light the fact that AI should be developed, deployed and regulated according to the promotion and protection of fundamental rights such as equality and non-discrimination.

For some years, while artificial intelligence had restricted use and was not connected to services for the general public, soft laws initiatives were enough. The three laws of Robotics were designed in 1950 by Isaac Asimov (Tardif, 2021) and can be considered the first ever governance proposal for AI (Marchant, 2019). Those were simple rules to maintain the integrity of robots without creating any harm to humans. The first law stated that a robot could not harm a human being, even if by the omission of help. The second law was created to keep robots obedient to humans, except when this violates the first law. The last one was regarding the protection of the robot itself when this would not conflict with the two previous laws. Since AI is connected with these fields, those laws were also applied to it (Tardif, 2021). When the widespread use of artificial intelligence started, companies tried to regulate themselves by monitoring their competitors, but this was not enough, and society started to demand some action from governments. Corine Cath (2018) compared eight articles from leading experts in the field about governing artificial intelligence. There were clear advantages about having ‘open norm-setting venues’ as soft law developed by private and public sectors, but all experts indicated in their work that it was vital to establish hard law regulations about AI.

Since 2016, AI researchers from Apple, Amazon, Google, Facebook, IBM, and Microsoft created the partnership on AI (PAI) to ‘study and formulate best practices on AI technologies, to advance the public’s understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society.’ (PAI, 2021). They became a multi-stakeholder organisation with support from companies, researchers, and civil society organisations, holding the safety-critical AI as one of their working pillars to propose an ethical alignment of AI actors. However, Thilo Hagendorff (2020) analysed twenty-two ethics guidelines from companies working in the development of artificial intelligence, some of them also members of PAI, and he concluded that these documents are not taken into account in the decision-making processes of AI developers precisely because of the absence of consequences in case of violation of their norms. These corporate ethical guidelines and best practices strategies are criticised for being developed and updated not to target a ‘good AI society’ but to cover companies’ own flaws. For this reason, Floridi (2019) advocates for digital ethics and explains that self-regulation creates other problems such as ethics-washing, where companies set their standards but do not follow them. In addition to ethics-washing, the self-regulation of artificial intelligence can also lead to ethics-shopping (Wagner, 2018) when companies select what values they want to use to set their standards according to what they are already doing instead of effectively trying to improve their actions with innovative ethical standards. Thus, private self-regulation to achieve an ethical alignment with fundamental rights is relevant to AI Governance, but the establishment of hard law regulations is vital to enforce remedies in case of violation of fundamental rights.

The next major development in AI Governance within the EU came in 2019 when the High-Level Expert group from the European Commission released the 'Ethics guidelines for trustworthy Artificial Intelligence', which resulted from State and non-state actors' cooperation through public consultation. Respect for ethical principles and values is a key aspect for AI in order to be considered trustworthy. The guidelines are based on seven key elements that AI should respond to, and one of them is 'Diversity, non-discrimination and fairness'. It specifically addresses the problem of avoiding unfair bias and the possibility of exacerbating discrimination. Another requirement is 'Accountability'. It states that there is a responsibility not only for using AI systems but also for their outcomes. However, this responsibility is a human characteristic that is connected to decision-making. In other words, AI does not need to be accountable, but the human who benefits from its outcome must be. Coeckelbergh (2019) shared two main critiques of the Ethics Guidelines. The first one was the absence of concrete measures in case the principles prescribed there are violated, which is a problem shared with corporate ethical guidelines, as shown above. It is essential to give users a way to contextualize discrimination based on AI-outcome. There is a different impact on different stakeholders, but still there is no standard for algorithm disclosure. The second one is the use of the document for public relations purposes. The Ethics Guidelines also connects the idea of trustworthy AI with the principle of explicability. This principle intends to fix the 'black box problem' by allowing decisions to be contested and outputs explained for the ones directly or indirectly involved. The guidelines also divided the concept of transparency regarding AI in three elements: explainability, traceability and communication. Explainability is the ability to explain technical processes related to AI in a way that human beings can understand and trace. For this reason, traceability enables the identification of the reasoning of an AI-decision by accessing the records of the data gathered. Also, communication is the right that individuals have to be able to identify that they are interacting with AI systems. According to Coeckelbergh (2020, 123), explainability or explainable AI (XAI) are more than a moral requirement to explain the reasoning behind a decision but a condition necessary for accountable behaviour. Patrice Bertail et al. (2019, 23) understood that one of the reasons why explainability is connected to the task of opening the black box of AI is not only because there is the need for compliance with law and regulation, but also the possibility to challenge the result in case of discrimination, for example. After the guidelines, the European Commission, through their High-level experts on AI, launched a public consultation from February to June 2020 to collect public opinion from stakeholders considering the document called 'The White Paper on

AI' (European Commission, 2020). Awareness of the need for stakeholder engagement to regulate artificial intelligence is a positive sign of the understanding that a single level solution is not the answer to mitigate bias and other issues related to AI. After the White paper from the European Union, other international actors released their ethical guidelines on AI, such as the OECD and UNESCO. Following the White Paper, on April 21, 2021, the European Commission submitted the first draft of the European Regulation for AI, the Artificial Intelligence Act (European Commission, 2021). The regulation divides three types of AI Systems prohibited AI – the ones against human rights e.g. social scoring systems – heavily regulated, considered High Risk, and less regulated, considered as other AI systems.

The first international regulation to AI was proposed by the European Commission (2021) on April 21, 2021, the AI Act. The regulation divides three types of AI Systems: a) prohibited AI, the ones against human rights e.g. social scoring systems; b) heavily regulated, considered High Risk, and c) less regulated, considered as other AI systems. This risk-based approach might set the standard for the rest of the world, as suggested by the United Nations High Commissioner for Human Rights (2021). It will require AI industries to comply with transparency rules keeping recorded documents and information that can be translated to regular users. The regulation is not only applicable to industries that develop artificial intelligence, but also addresses the responsibility of the ones that sell and deploy AI as a 'post-market monitoring system' (AI Act, article 61). Following the GDPR jurisdiction, the Artificial Intelligence Act covers companies working in the EU but also companies where the output is targeting the European market. The so-called "Brussels effect" could be responsible for the extension of the application of this legal regulation even in cases where data from EU citizens is used to teach the AI in whatever part of the world.According to this proposal, AI systems for recruitment purposes such as advertising vacancies, screening or filtering applications and evaluation of candidates will be considered high-risk. High-Risk AI systems will need to comply with a conformity label to verify their accuracy to avoid harmful effects for citizens. However, this conformity assessment (AI Act, articles 19 and 48) is an internal process that the AI industry should do before putting the product in the market. Since it is an internal check-off, however, it will not count with external surveillance and accountability which contradicts the whole transparency and explainability principles developed since the agreement for cooperation in AI. This regulation is the first of its kind, and will affect the standard for future regulations. Friederike Reinhold and Angela Muller (2021) stated that regarding transparency of AI systems, this regulation seems very promising, however, they criticised the transparency of concepts and criteria to classify AI systems. Also, according to them, the current proposal ignores the perspective of those affected by AI outcomes regarding the possibility of challenging AI results. A simple search on the document gives an overview that bias, biased outputs or discriminatory outputs are present in the part regarding impact assessment. Nevertheless, these mentions are vague and do not guarantee that people potentially affected by AI outcomes would have access to those impact assessments of bias.The Joint Research Centre from the European Commission expressively attested that the European Union's values and the ones present in the Charter of Fundamental Rights of the EU were considered to develop ethical guidance and openly discuss emerging challenges in the development of AI along with stakeholders and other key actors in the AI market (Annoni et al., 2018).

From soft law to hard law, all the European Union strategy to regulate artificial intelligence uses a human-centric approach, or at least intends to use one. This means that artificial intelligence should be developed considering human well-being as the centre of interest (European Commission, 2019). However, the current proposal of the AI Act is criticized by its possible lack of effectiveness regarding fundamental rights protection because it is focused on prediction and prevention but not on concrete measures that can be taken by citizens if their rights are violated by AI-driven outcomes (Veale and Borgesius, 2021; Chander and Jakubowska, 2021).

These selected milestones have the intention of summarizing how in less than ten years the concerns about regulating artificial intelligence to protect society from potential harms went from almost non-existent to an essential topic on the agenda. They also focus on the transition from self-regulation in the private sphere to public soft law under the concept of ethics regulations pending future international binding legal frameworks. It is clear that there is an awareness of the impact of AI on fundamental rights and concrete efforts are being directed to mitigate eventual negative ones. Notwithstanding, it is essential to bring light to the ones that are being left behind in those initiatives.

5. Whose fundamental rights?

The European Union has the protection of fundamental rights among its core values and the provision that they should be applied in all of their regulations. According to article 2 of the Treaty of the European Union (TEU), the principles of equality and non-discrimination are part of the foundation of the European Union, alongside respect for human rights. In the establishment of the internal market, the promotion of equality between women and men is perceived in article 3 (3) as one of the main targets to the sustainable development of the internal market in which the respect of these principles should prevail. The principle of non-discrimination is present in article 14 of the European Convention on Human Rights (ECHR) and protocol n. 12 extends the interpretation of this principle to a general prohibition of discrimination. The Charter of fundamental rights of the European Union (CFREU) in its article 20 addresses the principle of equality before the law, while article 21 the principle of non-discrimination. At a secondary law level, there are Directives addressing the protection and promotion of gender equality and non-discrimination in different sectors, such as employability (Directive 200/78/EC), occupation and social services (Directive 2006/54/EC), and goods and services (Directive 2004/113/EC). The first level has a broader perspective of the necessity of promoting gender equality and protecting women from gender discrimination. The secondary level identifies the possible victims and targets of specific areas according to each directive where those principles should prevail. This legislation works to protect women against direct and indirect discrimination in the public and private sectors. However, there is no provision against structural discrimination[2], and that's exactly the one present in the AI ecosystem. Simone de Beauvoir (1949, 20) already discussed this under the concept of social discrimination, attesting that this type of discrimination is the hardest to fight against because it is blurred under formal gender equality.

In the past years, the EU machinery has been working to develop a comprehensive legislation to prevent AI-driven technologies from causing any negative impact on society that could potentially violate fundamental rights. Thus, these new regulations will work in a complementary way to the legal framework already in place. Apart from the criticism oh whether the chosen human centric approach can be effective in the task, there are also some concerns that it won't prevent algorithm discrimination towards different genders.

Law, ethics and technology are the three guiding forces of AI Governance (Cath, 2018), but it is known that neither science nor technology are neutral fields regarding genders (Haraway, 2016). The human default is usually associated with men, as well as in questions of regulating emerging technologies. Moreover, Simone de Beauvoir (1949) stated that

[2] Structural discrimination, also found in the doctrine as systemic discrimination, is related to biased decisions and actions reproduced so many times they become accepted behaviour or part of the organisational cultures in the private or public sector. According to the International Labour Organisation (2017), this is one of the factors that contributes to gender pay gap because in some institutions women are paid less, and this is reproduced until a point that it becomes inherited, and no one questions it any longer.

the representation of the world was the work of men described from their own perspective, which ended up not being entirely accurate because they did not consider the other half of the world population. Also, declining to include the perspective of women is a great driver of gender bias that tends to pass as gender-neutral (Perez, 2020). It impacts the understanding of artificial intelligence when making decisions regarding women who usually do not fit into this human standard model, as demonstrated by cases cited in the previous section.

Considering that gender-biased AI-driven outcomes might lead to gender-based discrimination, it is not possible to assure that it will effectively violate the discrimination threshold that would fit into EU anti-discrimination regulations by favouring certain biases. In order to file a claim alleging direct or indirect discrimination under EU law, it is imperative to establish *prima facie* discrimination by demonstrating an immediate or occurred harm; a disproportion of this harm and that this harm has higher chance to be manifested in a group or individual that have protected attributes. Watcher (2020) clarifies that the EU jurisprudence shows a heterogeneity in the interpretation and application of EU non-discrimination legal framework, creating a contextual equality according to each case, this leads to a problem of defining standards of non-discrimination to be applied to AI driven-technologies.

Furthermore, the General Data Protection Regulation (GDPR) established two important rights connected to artificial intelligence and explainability. The first one is the right to explanation (article 15) and the right to not be subjected to an exclusively automated decision (article 22). This means that, regardless of the AI outcome, the decision is made by a human. As the use cases shown, the reliance on AI-driven technology outcomes is almost the standard now. When there is a gender-based bias displayed in an AI outcome, it is sometimes not possible for the human exposed to it to identify because this person might have the same bias. However, these biases were not created out of nowhere, they are a reproduction of societal biases in the environment where AI was trained and deployed. This is the trickiest aspect of gender-based bias, the socio-technical problem needs more than technical solutions to address this matter. AI-driven technologies are developed and deployed to give more efficient outcomes in less time. The absence of an agreed definition of concepts such as explainability and transparency just make concepts such as 'responsible AI' and 'trustworthy AI' buzzwords with no agreed upon standards to be followed which makes it even harder to monitor and audit AI.

Algorithm discrimination is a gap in EU law that must be fixed. The case of Google Ads exposing female users to fewer high paying jobs than male users can be an example of this gap. The scientists involved in this study, using the AdFisher tool to analyse job advertisements on the website, could not affirm that discrimination laws were violated because the incidents did not exactly fit into the requirements prescribed by anti-discrimination laws even if it was a clear violation of fundamental values such as gender equality (Datta et al., 2015). Also, even if there are plenty of cases involving gender-based bias in AI outcomes that effectively created an unbalanced treatment, there are few cases where the EU Courts ruled on it, and this can also be considered a reason why there are no preliminary rulings procedures, art 267 TFEU, on the topic (Lutz, 2022). In this regard, the Italian Court, Consiglio di Stato, has decisions involving AI-driven decision-making such as sentences n. 2270 (from 8 April 2019) and n. 7891 (4-25 November 2021) but those are rare exceptions. In addition, the Court of Justice of the European Union (CJEU) has not ruled on gender discrimination rooted in biased AI-driven outcomes, even if there are disputes involving gender equality under the Directive 2006/541[3]. Therefore, the legal gap is not only caused by the absence of specific legislation about artificial intelligence to complement the already existing EU legal framework regarding equality and non-discrimination, but also the way that this legislation is built.

Resolution 2017/3016 from the European Parliament of 17 April 2018 on empowering women and girls through the digital sector underlines the importance of ensuring that gender mainstreaming strategies are part of digital policies (European Parliament, 2018). One of the purposes of gender mainstreaming is to contribute to structural changes in society, taking the perspective that some genders are more affected than others (European Commision, 2004). Gender mainstreaming is enounced in the articles 3 of the Treaty on the European Union (TEU) and 8 of the Treaty on the Functioning of the European Union (TFEU) as the official approach of the European Union to combat discrimination and promote equality since it was endorsed by the European Union at the Fourth World Conference on Women in 1995 (UN Women, 2020). In simple terms, it recognizes the impact of gender in the development of policies, regulations and other activities within the socio-political and economic spheres. Gender mainstreaming entails the introduction of a gender sensitive approach, or gender lens, in all areas related to policy-making, not only those regarding the creation of policies specifically related to women's rights. For this reason, it cannot be

[3] Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation.

considered a policy goal, but it is a tool used to advance equality and social inclusion based on gender analysis (Shreeves, 2019).

The EU legal framework has valuable multi-level instruments focused on the protection of the principles of non-discrimination and equality that directly impact women's rights. Furthermore, there are clear efforts in AI Governance to address problems related to AI-driven technologies that could challenge these principles. However, there are gaps in the current and future regulations that allow gender-based bias in AI to continue to be a problem. There are powerful instruments, already recognised in EU Law, such as gender mainstreaming strategies to foster a gender perspective in EU regulations about artificial intelligence that could have a potential on impact the effective mitigation of gender bias in AI, but they are not applied by policy-makers.

Conclusion

The artificial intelligence revolution increased the demand and use of AI-driven technologies in our society. As a consequence, the escalation of gender bias by AI-driven technologies is not new, but the concern about its impact on the enjoyment of fundamental rights is increasing as the years goes by. Even if the AI ecosystem recognises the need to mitigate gender bias in artificial intelligence, there are plenty of use cases showing that it keeps happening. Moreover, the absence of an agreed upon definition of artificial intelligence itself increases the uncertainty around other key concepts in the AI ecosystem such as explainability, transparency, responsible AI and even trustworthy AI. This lack of standards affects the urgent need to develop proper monitoring and auditing tools for AI-driven technologies in order to avoid negative impacts on society such as gender-based bias outcomes.

Artificial intelligence has just overexposed a problem that for many years has affected women's rights. This can be used as an opportunity to bring this discussion to the table since it is no longer possible to continue ignoring biased outcomes as has been happening for the past decade.

Even if artificial intelligence is a cross-border matter, this article focuses on the European Union normative framework considering its unique position as a possible standard for the rest of the world regarding AI law (OHCHR, 2021) and the possible complementary work that the AI Act can bring to EU anti-discrimination regulations. From soft law to hard law, all the European Union strategy to regulate artificial intelligence uses a human-centric approach, which means that artificial intelligence should be developed considering human well-being as the centre of interest. However, the human default is usually associated with men. The lack of gender perspective in the development of regulatory frameworks, even if prescribed as necessary by EU resolutions such as 2017/3016, renders attempts to mitigate gender bias in AI without much practical impact. It is known that gender bias in AI is a socio-technical issue that needs more than technical solutions to be fixed. The efforts to address gender bias in EU regulations, from soft and hard law, can be seen as an AI ethics-washing compilation since they are not followed by concrete commitments from the business sector and safeguards from legislators to address this issue in a way that facilitates effective changes in society. Thus, the absence of looking at the problem as rooted in multiple sources from start to finish might trap scientists in a continual game of technical whack-a-mole (Hoffmann, 2018).

In summary, soft and hard laws play a key factor in regulating artificial intelligence. The European Union is taking a leadership role in regulating artificial intelligence according to their human and fundamental rights values. The future legislation on artificial intelligence (The AI Act) is already considered a standard to be followed by other countries and international organisations. In any case, the public and private sectors need to work together to mitigate gender bias in its origin from a social and technical approach, and not only chase the problem after it starts to affect women's rights, as is happening today.

For this reason, bringing gender mainstreaming strategies to understand the problem as more than only a technical matter is essential. The potential of artificial intelligence to improve people's lives is undeniable. However, there is an urgent need to establish boundaries and a holistic oversight to avoid this piecemeal addressing of women's rights protection which is unsustainable in the long term.

References

Adams, R. L. (2021). '10 Powerful Examples of Artificial Intelligence In Use Today', retrieved from:https://www.forbes.com/sites/robertadams/2017/01/10/10-powerful-examples-of-artificial-intelligence-in-use-today/ (accessed: 05/12/2021)

Annoni, A., Benczur, P., Bertoldi, P., Delipetrev, B., De Prato, G., Feijoo, C., Fernandez Macias, E., Gomez Gutierrez, E., Iglesias Portela, M., Junklewitz, H., Lopez Cobo, M., Martens, B., Figueiredo Do Nascimento, S., Nativi, S., Polvora, A., Sanchez Martin, J., Tolan, S., Tuomi, I. and Vesnic Alujevic, L. (2018). Artificial Intelligence: A European Perspective, Craglia, M. editor(s), EUR 29425 EN, Luxembourg: Publications Office of the European Union.

Bachelet, M. (2021). 'Urgent action needed over artificial intelligence risks to human rights', retrieved from: https://news.un.org/en/story/2021/09/1099972 (accessed: 05/12/2021).

Bathaee, Y. (2018) 'The artificial intelligence black box and the failure of intent and causation', Harvard Journal of Law & Technology, 31, no. 2 (Spring 2018): 890–938.

Bertail, P., et al., (2019), 'Algorithms: Bias, Discrimination and Equity', retrieved from: https://www.telecom-paris.fr/algorithms-bias-discrimination-and-equity (accessed: 09/12/21)

Beauvoir, S., (1949). The Second Sex. Parshley.

Bloomberg, (2016) , 'Artificial Intelligence Has a "Sea of Dudes" Problem', retrieved from: https://www.bloomberg.com/news/articles/2016-06-23/artificial-intelligence-has-a-sea-of-dudes-problem (accessed: 09/12/21)

Buiten, M. C., (2019) 'Towards Intelligent Regulation of Artificial Intelligence', European Journal of Risk Regulation 10, no. 1, 41–59.

Bughin, J., Seong,J., Manyika, J., Chui, M., and Joshi, R., (2018), 'Modeling the Global Economic Impact of AI | McKinsey' retrieved from: https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy (accessed: 06/12/21)

Bryson, J. (2017), 'From the RSPE Equity and Access Committee', Research School of Physics Event Horizon 43 issue 29, 7-11.

Cath, C. et al., (2016), 'Artificial Intelligence and the "Good Society": The US, EU, and UK Approach', Rochester, NY: Social Science Research Network

Chander, S. and Jakubowska, E. (2021), 'EU's AI law needs major changes to prevent discrimination and mass surveillance', European Digital Rights (EDRi), retrieved from: https://edri.org/our-work/eus-ai-law-needs-major-changes-to-prevent-discrimination-and-mass-surveillance/ (accessed: 01/10/22)

Charles, M., Clark, A., and Gevorkyan, A. V, (2020). 'Artificial Intelligence and Human Flourishing', American Journal of Economics and Sociology 79, no. 4, 1307–44

Coeckelbergh, M., (2019), 'Artificial Intelligence: some ethical issues and regulatory challenges', Technology and regulation, 31–34

Coeckelbergh, M. (2020). AI Ethics. Cambridge, MA: MIT Press.

Comandé, G. (2020) 'Unfolding the Legal Component of Trustworthy AI: A Must to Avoid Ethics Washing'. Rochester, NY: Social Science Research Network.

Council of Europe Commissioner for Human rights, (2019), 'Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights' retrieved from: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 (accessed: 09/12/21)

Council of Europe (2021), 'Discrimination and Intolerance', retrieved from https://www.coe.int/en/web/compass/discrimination-and-intolerance (accessed 07/12/21)

Craglia, M., coord., (2018), 'EU Declaration on Cooperation on Artificial Intelligence', retrieved from: https://ec.europa.eu/jrc/communities/en/node/1286/document/eu-declaration-cooperation-artificial-intelligence (accessed: 09/12/21)

Cummings, M. L., 'Artificial Intelligence and the Future of Warfare', retrieved from: https://www.chathamhouse.org/sites/default/files/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings-final.pdf (accessed: 05/12/21).

Dastin, J. (2018), 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women', Reuters, retrieved from:https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G (accessed: 07/12/21)

Datta, A., Tschantz, M.C., and Data,A. (2015), 'Automated Experiments on Ad Privacy Settings', Proceedings on Privacy Enhancing Technologies, 92–112

Edwards, L., and Veale, M.,(2018), 'Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking For', Duke Law & Technology Review 16 (2018 2017)

European Commission, (2021), 'Proposal for a Regulation Laying down Harmonised Rules on Artificial Intelligence | Shaping Europe's Digital Future', retrieved from: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence (accessed: 09/12/21)

European Commission, (2019), Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions - Building Trust in Human-Centric Artificial Intelligence, Brussels: European Commision.

European Commission, (2020), White paper On Artificial Intelligence - A European approach to excellence and trust. Brussels: European Commission.

European Court of Human Rights (2011). 'Al Khawaja and Tahery v United Kingdom', retrieved from: https://justice.org.uk/al-khawaja-tahery-v-united-kingdom/ (accessed:06/12/21).

European Parliament (2018). 'European ethical Charter on the use of artificial Intelligence in judicial systems and their environment 'retrieved from: https://www.europarl.europa.eu/cmsdata/196205/COUNCIL%20OF%20EUROPE%20%20European%20Ethical%20Charter%20on%20the%20use%20of%20AI%20in%20judicial%20systems.pdf (accessed: 06/12/2021)

European Parliament. (2018). *Resolution 2017/3016 (RSP) on Empowering women and girls through the digital sector.* Adopted on 17 April 2018. Strasbourg.

Floridi, L. (2019). 'What the near future of artificial intelligence could be'. Philosophy & Technology, 32: 1-15.

Gebru, T. et al. (2021) 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?', in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency , Virtual Event Canada: ACM, 2021)

GeekWire, (2020). 'Microsoft President Brad Smith Calls for AI Regulation at Davos', retrieved from:https://www.geekwire.com/2020/microsoft-president-brad-smith-calls-ai-regulation-davos/ (accessed: 09/12/21)

Haenlein, M., and Kaplan, A., (2019), 'A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence', California Management Review 61, no. 4 (1 August 2019): 5–14

Hagendorff, T., (2020). 'The Ethics of AI Ethics: An Evaluation of Guidelines'. Minds and Machines 30, no. 1 (1 March 2020): 99–120

Harari, Y.,N. (2017). 'Reboot for the AI Revolution', Nature News 550, no. 7676.

Haraway, D.J. (2016). A Cyborg Manifesto. Minnesota, US: University of Minnesota Press.

High-Level Expert Group on AI (AI HLEG) (2019a), Ethics Guidelines for Trustworthy AI. Brussels: European Commission.

__________. (2019b). 'A Definition of Artificial Intelligence: Main Capabilities and Scientific Disciplines'. Brussels: European Commission.

Hoffmann, A., (2018), ‘Data Violence and How Bad Engineering Choices Can Damage Society’, retrieved from:https://medium.com/s/story/data-violence-and-how-bad-engineering-choices-can-damage-society-39e44150e1d4 (accessed: 09/12/21)

International Labour Organisation (2017), ‘Breaking barriers: Unconscious gender bias in the workplace’, retrieved from: https://www.ilo.org/actemp/publications/WCMS_601276/lang--en/index.htm (accessed: 01/10/22)

Lawson, A., E. (2000). ‘How Do Humans Acquire Knowledge? And What Does That Imply About the Nature of Knowledge?’. Science & Education 9, 577–598

MacCarthy, M., and Propp, K., (2021), ‘Machines Learn That Brussels Writes the Rules: The EU’s New AI Regulation’, retrieved from: https://www.brookings.edu/blog/techtank/2021/05/04/machines-learn-that-brussels-writes-the-rules-the-eus-new-ai-regulation (accessed: 09/12/21)

Manyika, J., Silberg, J., and Presten, B, (2019), ‘What Do We Do About the Biases in AI?’, Harvard Business Review, retrieved from: https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai (accessed: 07/12/21)

Marchant, G (2019), ‘Soft Law Governance Of Artificial Intelligence’, AI Pulse, 25 January 2019, retrieved from: https://aipulse.org/soft-law-governance-of-artificial-intelligence/ (accessed: 01/10/22)

Markowsky, G.. (2017) ‘Information theory.’ Retrieved from https://www.britannica.com/science/information-theory (accessed: 07/12/21)

Muhammad, Y. (2019), ‘Research Challenges of Big Data’, Service Oriented Computing and Applications 13, no. 2 (1 June 2019): 105–7.

Oxford Dictionary. (2021). ‘artificial intelligence’. Retrieved from: ‘Artificial Intelligence’, Oxford Reference, accessed 31 January 2021, https://doi.org/10.1093/oi/authority.20110803095426960 (accessed: 05/12/21).

Pentland, A. (2013). ‘The data-driven society’. *Scientific American*, 390 (4), 78-83.

Perez, C. C.,(2020). Invisible Women: Exposing Data Bias in a World Designed for Men. 1st edition. Vintage

Raso, F. et al., (2018) ‘Artificial Intelligence & Human Rights: Opportunities & Risks’, Berkman Klein Center for Internet & Society Research Publication, Harvard University’s DASH repository, retrieved from: http://nrs.harvard.edu/urn-3:HUL.InstRepos:38021439 (accessed: 07/12/21)

Reinhold, F., and Müller, A., 'AlgorithmWatch's Response to the European Commission's Proposed Regulation on Artificial Intelligence – A Major Step with Major Gaps', retrieved from:https://algorithmwatch.org/en/response-to-eu-ai-regulation-proposal-2021 (accessed: 09/12/21)

Research and Markets ltd, (2020), 'Artificial Intelligence Global Market Report 2020-30: COVID-19 Growth and Change', retrieved from: https://www.researchandmarkets.com/reports/5050683/artificial-intelligence-global-market-report-2020 (acessed: 06/12/21)

Rodrigues, R., (2020), 'Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities', Journal of Responsible Technology 4 (1 December 2020): 100-105

Sartor, G. et al., (2020). 'The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence: Study', retrieved from: http://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf. (accessed: 05/12/21)

Sood, A., and Tellis, G. J., (2005) 'Technological Evolution and Radical Innovation', Journal of Marketing 69, no. 3, 152–68.

Stahl, B. C., (2021). Artificial Intelligence for a Better Future An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies. Leicester, UK: Horizon 2020 Framework Programme.

Stankiewicz, K. (2020), 'IBM Chief Calls for "precision Regulation" on AI That Weighs Privacy against Benefits to Society', retrieved from: https://www.cnbc.com/2020/01/22/ibm-ceo-ginni-rometty-calls-for-precision-regulation-on-ai.html (accessed: 09/12/21)

Tardif, A. (2021), 'How Asimov's Three Laws of Robotics Impacts AI', retrieved from: https://www.unite.ai/how-asimovs-three-laws-of-robotics-impact-ai/ (accessed: 09/12/21)

The Guardian, (2016), 'The Racist Hijacking of Microsoft's Chatbot Shows How the Internet Teems with Hate' retrieved from: http://www.theguardian.com/world/2016/mar/29/microsoft-tay-tweets-antisemitic-racism (accessed: 07/12/21)

The official Microsoft blog, (2016) 'Learning from Tay's Introduction' retrieved from: https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/ (accessed: 07/12/21)

The Partnership on AI (PAI), (2021), 'About', The Partnership on AI, retrieved from: https://www.partnershiponai.org/about/ (accessed: 09/12/21).

Tung, L., (2021) ‘Google CEO Sundar Pichai: This Is Why AI Must Be Regulated’, retrieved from: https://www.zdnet.com/article/google-ceo-sundar-pichai-this-why-ai-must-be-regulated/ (accessed: 09/12/21).

Turing, A., (1950) ‘Computing Machinery and Intelligence’, Mind, LIX/236, 433-460.

UNESCO. (2019). ‘I’d blush if I could’, retrieved from: https://en.unesco.org/news/women-tech-id-blush-ificould#:~:text=I’d%20blush%20if%20I%20could%20is%20the%20title%20of,’re%20a%20bi***.%E2%80%9D (accessed: 09/12/21)

United Kingdom Government Office for Science. (2015). ‘Artificial Intelligence: Opportunities and Implications for the Future of Decision Making’, retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf (accessed: 05/12/2021).

United Nations High Commissioner on Human Rights. (2021), ‘The right to privacy in the digital age’. Presented in the 48th session of the Human Rights Council, 1-20.

Veale, M. and Borgesius, F. Z. (2021), ‘Demystifying the Draft EU Artificial Intelligence Act’, Computer Law Review International, CRi 4/2021, 97-112.

Vinuesa, R., et al. (2020), ‘The Role of Artificial Intelligence in Achieving the Sustainable Development Goals’, Nature Communications 11, no. 1

Wagner, B., (2018). ‘Ethics As An Escape From Regulation.: From “Ethics-Washing” To Ethics-Shopping?’, in Being Profiled, ed. Emre Bayamlioğlu et al., Cogitas Ergo Sum: 10 Years of Profiling the European Citizen, Amsterdam: Amsterdam University Press, 84–89

Wachter, S., Mittelstadt, B., and Russell, C. (2020), ‘Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI’, SSRN Scholarly Paper, Rochester, NY: Social Science Research Network, 3 March 2020.