



LUND
UNIVERSITY

What's harm got to do with it?

The framing of accountability and harm in the EU Artificial Intelligence Act Proposal

Yasmin Omar

Lund University
Sociology of Law Department

Master Thesis (SOLM02)
Spring 2022



Supervisor: Jannice Käll

Examiner: Sherzod Eraliev

Abstract

The European Union released an Artificial Intelligence (AI) regulation proposal in April 2021 aimed at laying down harmonised rules for AI circulating the Union market. The purpose of this study is to critically examine how accountability and individual, collective, and social harm was approached and framed by the proposal. Previous research has shown that AI and algorithmic bias and discrimination is a widely known concern and there is a pressing need for regulation that protects against various types of harms. The theoretical framework chosen is built up of Boven's public accountability theory as well as interjections from the field of critical algorithm studies. The study conducted a critical discourse analysis as designed by Fairclough of both the regulation proposal as well as other articles written in response to it, in order to see the relationship between the text, the discursive practices and the social practices.

The results showed that the regulation proposal contained several empty promises regarding its intent and was commercially minded. The text showed a clear balance between innovation and development of AI and protections of fundamental rights, but then failed to deliver in terms of mechanisms established to uphold these promises. The regulation also made several exceptions for both groups such as law enforcement and the military, as well as AI systems. There was no involvement of individuals both prior to and within the regulation as well as no protections or rights for the public. As such, accountability and individual, collective, and social harms were not sufficiently considered by the regulation proposal.

Key words: artificial intelligence, European Union, regulation, accountability, harm, critical discourse analysis

Table of contents

| | |
|--|----|
| 1. Introduction | 4 |
| 1.1 Aim..... | 8 |
| 1.2. Research Questions | 8 |
| 1.3. Delimitation | 9 |
| 1.4. Disposition..... | 9 |
| 2. Background | 10 |
| 2.1. Previous resolutions | 11 |
| 2.2. AI Act | 12 |
| 3. Literature review | 14 |
| 3.1. AI bias and discrimination..... | 15 |
| 3.2. AI regulation..... | 18 |
| 3.3. Knowledge gap | 19 |
| 4. Theoretical framework | 21 |
| 4.1. The rise of accountability | 22 |
| 4.2. “Public” accountability and other types | 24 |
| 4.3. Algorithmic accountability | 27 |
| 5. Methodology | 31 |
| 6. Analysis | 38 |
| 6.1. Introductory sections | 38 |
| 6.1.1. <i>Explanatory Memorandum</i> | 38 |
| 6.1.2. <i>Structure of the proposal</i> | 40 |
| 6.1.3. <i>Defining the problem of AI</i> | 42 |
| 6.2. Human rights, bias and discrimination | 45 |
| 6.2.1. <i>Fundamental human rights</i> | 45 |

| | |
|--|----|
| 6.2.2. <i>Ethical implications</i> | 50 |
| 6.3. Trust and Conformity | 52 |
| 6.3.1. <i>Harmonisation</i> | 52 |
| 6.4. Legality and obligations | 56 |
| 7. Conclusion | 60 |
| 7.1. Future recommendations | 63 |
| 8. Bibliography | 64 |

1. Introduction

Definitions are the backbone of law-making; the purpose of laws is to define rules and regulations, which further define the behaviours in society (Määttä, 2007). The law, according to sociolegal scholarship, is heavily influenced by social, political, and cultural contexts and vice versa (Darian-Smith, 2013). While society is not only ruled by formal laws but also through informal rulings and norms, the same applies to them. Nonetheless, the importance of formal laws should not be ignored as most current society is controlled by state laws that have been established by an authority. Furthermore, laws are not constant: as society changes, so do the rules in place for governing because laws are influenced by society and must therefore be analysed as such (Darian-Smith, 2013). Old laws are amended or reformed, while new laws are created for new situations that requires to be regulated. This is the case for artificial intelligence.

In April of 2021, the European Union (EU) released their regulation proposal with the aim of laying down harmonised rules across the EU regarding the creation and usage of artificial intelligence (AI) (European Commission, 2021). It was the first time that the EU made such a large-scale attempt at AI legislation and was highly relevant considering the consistent growth of AI in various spheres of life and society. However, due to it being both brand new and the first of its kind, it was arguably difficult to ensure that it covers any and all aspects of AI development and use. Furthermore, this author is especially interested in how this legislation proposal has considered accountability and various harms such as individual, collective, and social, which is the aim of the thesis. However, first one must look into the details of AI and the multifaceted landscape of issues that exists with it.

Artificial intelligence (AI) is a widely used term for various technologies and algorithms created to “reproduce complex human capabilities such as language use, vision and autonomous action” (Caldwell et al, 2020, p. 3). These technologies have been integrated into society’s different levels and spheres. We are increasingly

surrounded by AIs everywhere we go, right down to our own pockets. Indeed, much effort is being put into making machines learn what they need to learn in order to act and think like humans (Caldwell et al, 2020). However, as we let AI take a larger part in our lives, we need to be careful with how to proceed. With AI becoming smarter and increasingly more independent, the technology does not require as much human intervention but rather works on its own based on past information. This unfortunately also means that bias and discrimination is a problem that keeps on growing.

As society develops and city populations grow denser than ever before, it is not strange that AI is becoming imperative to have and use when dealing with city functions (Falco, 2019). “Smart cities” is a well-known concept involving digitalisation in order to improve operations and standard of living, whether it be surveillance, automated vehicles, infrastructure development and so on (Duarte & Ratti, 2018; Yin et al, 2015). Where urbanisation has caused great pressure on the governing forces, the introduction of AI and the internet to deal with those pressures have alleviated the problem and allowed for further city improvements (Yin et al, 2015). However, due to AI having shown to reiterate inherent human biases in automated form, it is crucial to focus on the social responsibility that comes with AI being used on such a large scale (Falco, 2019), and how accountability is involved in the process.

Smuha (2021) identifies three types of harm that AI may cause: individual harm, collective harm, and social harm. Whereas individual harm refers to when one person’s interests are being hurt for example through biased facial recognition due to race or ethnicity, and collective harm concerns groups of individuals being discriminated against through, for example, predictive policing on certain locations, social harm in relation to AI refers to the larger interests of society being harmed. This type of harm is rarely talked about as it is not seen as urgent as the previous two, despite the fact that it deals with important topics like equality, democracy, and the rule of law. Furthermore, Smuha (2021) also writes that European Union

(EU) policymakers need to have a societal perspective when creating regulation proposals rather than an individualistic one as societal interests are just as imperative yet easier to forget, and in terms of AI, it becomes crucial to consider the social harm that it may cause.

Multiple cases in the past have decidedly proven that machine bias is real and that algorithmic systems can be harmful whether intentionally or not. In 2016, the non-profit organisation ProPublica released an investigative report about the United States risk assessment tool COMPAS, Correctional Offender Management Profiling for Alternative Sanctions, used by the American criminal justice system to handle court cases. COMPAS uses an algorithm to systematically assess each case for potential risk of recidivism, however, ProPublica discovered that the software was racially biased (Angwin et al, 2016). It seemed that the software, which depended on previous data fed by humans, had been racially categorising black defendants as “high risk” of re-offending, and white defendants as “low risk”. Since then, the U.S. has not ceased using this software when determining the outcome of court cases, likely due to its efficiency (Angwin et al, 2016; Eckhouse et al, 2019).

At Amazon Inc., automation was the key to their success, including the hiring process of future employees. It was a revolutionary project, one that could have been a game-changer, had it not been giving discriminatory results. In 2015, people remarked that the algorithm used for vetting job applications had been hiring male workers more frequently than female workers. The team behind the system disbanded but the problem did not lie with them. The issue was that the algorithm had been fed data concerning previous job application processes and resumes over the last 10 years, which turned out to be heavily male dominated, meaning the bias had existed long prior to the introduction of the algorithm. Thus, the algorithm and the team of creators were not to blame because they did not initiate the problem, they only continued the cycle (Dastin, 2018). Facebook went through something similar with their advertisement’s suggestions, where their algorithm would base

their choice of an advertisement on the gender of the user, thus showing careers in science, maths, and technology to male users but not female users (Lambrecht & Tucker, 2015). These are just some examples of when algorithmic systems fail to work without bias, and the question of accountability is raised.

Generally, accountability refers to explaining and justifying one's behaviour in front of an audience, and accepting whatever consequences that may arise (Bovens, Schillemans & Goodin, 2014). In order to do this, one needs to be transparent and allow the audience to know the full scope of the occurrence so that they may give an appropriate response in return. Algorithmic accountability should be no different in either theory or practice, yet the issue lies with the practical aspect. Some challenges with algorithmic accountability include "human resources, legality, ethics, and the role [of] transparency..." among others (Diakopoulos, 2014, p. 26). According to Neyland (2007), accountability needs to have transparency in order to truly function as it is a sign of taking responsibility, especially for those in charge of the algorithm. AI and algorithms are for example often plagued by the "black box" issue where the system, which involves an input and an output, does not show what happens within it, meaning the processes between the input and output are unknown (Diakopoulos, 2014; Pasquale, 2015). Because of this, the biases may be blamed on the "black box" issue and transparency becomes null and void.

Above all else, the lack of legislation to deal with these issues is what is most concerning. Members of the EU are bound by both national regulations and EU legislation. The EU does have non-discrimination laws, for example, The Treaty on the Functioning of the European Union (TFEU) which prohibits discrimination based on nationality but also allows for actions to be taken against discrimination on the grounds of sex, race, or ethnicity, religious or spiritual belief, disability, age, or sexual orientation (EUR-lex, n.d.). Furthermore, the European Convention of Human rights (ECHR) which was drafted in 1950 with the intent to protect human rights and freedoms binds 47 countries, of which only 27 are EU members, and have in most Member states been incorporated into domestic law (EUR-lex, n.d.;

Woods, Watson & Costa, 2017). However, when it comes to AI, there is a need for specific laws regarding the topic in terms of human rights violations and discrimination, which is why the EU has now started creating regulation proposals like the AI Act (Veale & Borgesius, 2021). The AI Act was released April of 2021 with the aim of laying down harmonised rules across the EU regarding AI that would both ensure fundamental rights as well as AI development.

1.1 Aim

The purpose of this thesis is to investigate how AI and algorithmic accountability is seen and implemented by the European Union (EU) through this new AI regulation proposal (European Commission, 2021). This regulation is intended to protect society while simultaneously promoting the innovation and development of AI. This regulation needs to be exploring the impact of AI on all levels: whether it be individual human rights and protections on smaller scales such as surveillance and biometric systems, collective issues such as predictive policing and racial discrimination, where AIs are involved in important decisions for development on “group” level, or the influence of AI on larger societal impacts, involving governmental agencies and equality, democracy, power, and more. This thesis will explore how this proposal has been written and for which intents and purposes. The method used for this thesis is a critical discourse analysis as created by Fairclough (2013) as it relates to not only the text itself but also how it connects with the discursive and social practices. The analysis will further be connected to the theoretical framework of Boven’s public accountability theory as well as interjection of studies from the field of critical algorithm studies.

1.2. Research Questions

1. How has accountability been framed by the discourse in the new AI regulation proposal of the European Commission?

2. How has the new regulation proposal approached the issue of individual, collective and social harm?

1.3. Delimitation

The initial focus of this thesis was to look at AI-related crimes and how legislation has attempted to counteract this. However, it seems that most legislation attempts so far have been focused on preventing individual, collective and social harms rather than specific crimes caused by various AI. While this current proposal has for example explored AI systems with varying risk threats, it is still focused on how to ensure both innovation and furthering the development of AI systems and protecting fundamental human rights at the same time, which is why the aim of the thesis has changed to investigating the ways accountability and various harms has been framed through the discourse of the proposal. Furthermore, while an international focus would have certainly been interesting, the current proposal for the regulation of AI by the EU was of special interest to this author who is a citizen of the EU and thus will be the focus of this study.

1.4. Disposition

This thesis will be structured as follows: after this first chapter including the introduction, aim, and research questions, chapter two will explain the background of AI legislation in the EU, its developments, and previous attempts at regulating it. Chapter three will contain a review of relevant literature for this study that explores the issues with AI. The fourth chapter contains the theoretical framework developed with a discussion of relevant concepts used. Chapter five will explain the methodological decisions made and a discussion surrounding the strengths and limitations of the chosen method. Chapter six will present the analysis of the material and discuss it, followed by a conclusion and some future recommendations in chapter seven.

2. Background

The EU consist of several important institutions with varying functions. The European Commission (EC) was established in 1958 and is made up of 27 members, one from each Member state. The EC is the only EU institution that can propose new laws for the European Parliament (EP) and the European Council to adopt (Woods, Watson & Costa, 2017; European Union, 2022). The President of the European Commission (EC), Ursula von der Leyen wrote in her political guidelines for the 2019-2024 Commission, A Union that Strives for More, that she aims to “put forward legislation for a coordinated European approach on the human and ethical implications for Artificial Intelligence” within the first 100 days of her election as President (Von der Leyen, 2019, p. 13).

Following this, on February 19th, 2020, the EC released the White Paper on AI, A European approach to excellence and trust, which aimed to set policy alternatives that would enable the promotion and development of AI while also taking into account the possible risks associated with it (European Commission, 2020). This paper was a response to the numerous requests from the European Parliament (EP) and the European Council urging for there to be legislation regarding AI and a working internal market for AI systems. Furthermore, they emphasised that as AI becomes more involved in society, citizens fear that their rights and protections are in danger which is why the Commission focused on the seven requirements that the High-Level Expert Group of AI had established regarding trustworthy AI: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental wellbeing; and accountability (European Commission, 2020). Through this, they would ensure that any European regulation would have the citizens in mind while improving the internal market and continuing the development and implementation of AI in Europe.

2.1.Previous resolutions

In October 2020, three resolutions were approved by the EP that were also related to AI covering ethics, civil liability, and intellectual property (European Commission, 2021). One resolution set a framework of ethical aspects of AI, robotics, and related technologies, with an explicit focus on preserving ethical protection when exploring the possible benefits of AI. They, for example, urged that AI should be “human-centric, human-made and human-controlled”, as well as emphasised that the technologies be safe, transparent, and accountable (European Parliament, 2020a, p. 3). Additionally, AI and other technologies should be environmentally sustainable, remedied of biases and discrimination and adhere to privacy of the use of biometric recognition. Essentially, this was to certify that the ethical principles are not ignored in the quest for AI development (European Parliament, 2020a).

The resolution for civil liability regime for AI was created with the expressed focus of liability. It states that “there is no need for a complete revision” of the previous liability regimes of the EU but that due to AI systems being so complex, some definitive modifications are needed in order to fully cover those systems as well (European Parliament, 2020b, par 6). It also outlines liability considerations regarding the operator and that everyone involved in the operation of AI systems must be held accountable for potential harms or damages it may cause. Finally, it covers that there must be different liability rules for different risks of AI and that liability coverage is unmistakably an important part of future successes for AI (European Parliament, 2020b).

The third resolution concerns AI technology development and intellectual property rights (IPR) where the EP wished for the EC to put forward regulations where creators of AI are protected by balanced IPR protection (European Parliament, 2020c). They stressed the importance of a unified legal framework for the EU where the sharing of data is possible across the EU and for the fostering of an environment

of innovation and creativity. Patent law must be assessed, and IPR, as well as copyright and originality, should be protected (European Parliament, 2020c).

A long-awaited fourth resolution was published in October 2021 specifically regarding AI in criminal law and its use by the police and judicial authorities in criminal matters (European Parliament, 2021). Due to the numerous instances where the use of AI in the criminal justice system has had an effect on fundamental human rights, it was deemed necessary to create a resolution in order to protect rights such as the right to a fair trial, non-discrimination and equality, privacy and personal data risks, freedom of expression and so on. The key points described in the resolution involved transparency which is a known issue regarding AI systems, power imbalance which exists with both public and private actors that use and produce these systems, biases and discrimination that are exacerbated by the use of AI and cybersecurity issues. Conclusively, this resolution urges law enforcement and judicial authorities in Member States to be aware of these issues and actively work against them (European Parliament, 2021).

2.2. AI Act

Each resolution was created for the explicit purpose of highlighting and regulating the use of AI in various sectors such as criminal matters, intellectual property rights, ethical considerations, and civil liberty. However, not only were they specifically considering a certain aspect, but they were also only resolutions which means they were not legally binding for the Member States to follow, but rather functioned as recommendations. Additionally, some Member States took these recommendations as inspiration when creating their own national regulations. This all led up to the AI Act proposal, an attempt at horizontally laying down harmonised rules for the EU regarding regulation for AI development and usage (European Commission, 2021), which would be legally binding and all-encompassing. Furthermore, as it is relatively new and a first-time endeavour at legislating AI, it is most likely not free of faults. Therefore, this thesis will attempt to analyse the new proposal from a

socio-legal perspective with a focus on accountability and individual, collective, and social harm.

3. Literature review

In this chapter, the process for the thematic literature review will be explained and presented. A literature review is crucial in order to discover the pre-existing research into the topic in question and to understand different findings and theories, and draw conclusions (Banakar, 2019). Furthermore, it will reveal the knowledge gap in previous literature that this thesis will attempt to identify and give suggestions as to how to bridge this gap.

This literature review concerns AI and legislation, from a socio-legal standpoint which includes individual, collective, and social harm and accountability. As this thesis will specifically be looking into the new EU regulation proposal, it is first important to get a holistic view of other regulations from the past and efforts at regulating AI. Furthermore, studies regarding biases and other types of harms that AI has caused, and what the consequences were, is also of particular interest here. Thus, to begin with, a simple Google search was performed, just to get an insight into the field. A method of snowball sampling was used to allocate various literature across disciplines (Patton, 2015).

Thereafter, both the Google Scholar search engine and the Lund University library database, LUBsearch, were used with a combination of the following keywords and phrases: “artificial intelligence”, “AI”, “algorithm”, “machine bias” “legislation”, “regulation”, “bias”, “discrimination”, “accountability”, “harm”, “social harm”, “individual harm”, “collective harm”, “transparency”, “responsibility”, “obligations”, “EU”, “European Union”, “European Commission”. Other limitations were using only peer-reviewed articles that were published between the years 2000 and 2022, because despite the AI regulation attempts being fairly recent, issues concerning algorithmic biases and discrimination have existed long prior to any legislation endeavour, which is why the chosen time frame goes back to the year 2000. Additionally, several books were used for inspiration such as “Smart Technologies and the End(s) of Law” by Mireille Hildebrandt and “Regulating

Artificial Intelligence: Binary Ethics and the Law” by Dominika Ewa Harasimiuk and Tomasz Braun.

The most relevant literature was chosen for this review and will be presented down below, divided into different themes.

3.1. AI bias and discrimination

AI is developing rapidly and is steadily being implemented into various aspects of society. However, not only is the widespread use being underestimated, but it is also presenting several unique legal issues. Each development in technology is another step towards improving factors such as efficiency, fairness and removing human biases, yet there is a tendency to not expect the potential for machine and algorithmic biases as well as a furthering of discriminatory practices (Caulders et al, 2021; Kammerer, 2022; Moss, 2020; Nwafor, 2021; Raub, 2018).

Several studies have found that one sector where AI bias is magnified rather than minimised is the employment sector (Dastin, 2018; Houser, 2019; Kammerer, 2022; Kim & Bodie, 2021; Moss, 2020; Nwafor, 2021). Discriminatory factors may be related to aspects of race and ethnicity, gender, sexual orientation, religious and spiritual beliefs and several more. Moss (2020) studied the disparity between automated recruitment and people with disabilities. According to Moss (2020), individuals with disabilities, more often than not, were rejected by various automated hiring tools due to biases regarding matters such as nonverbal communication. This community has its fair share of disproportionately high rates of unemployment, around 11,2%, in comparison to able-bodied people with 6,1%, especially those with developmental disabilities such as autism in which the AI would struggle with reading micro expressions and signs of self-expression (Moss, 2020). The nature of discrimination has changed because unlike traditional forms of discrimination, automated discrimination has the tendency to be more subtle and abstract, making it harder to detect (Watcher, Mittelstadt & Russell, 2021).

Gender is another factor that has continuously shown itself to be a problem for AI. Giant tech companies such as Google and Facebook have had several controversies regarding gender-based discrimination, whether it be in the working force such as women only making up less than 15% of employees, or within machine translation tools and chatbox such as Google Translate automatically switching gender-neutral pronouns to gender-stereotypical ones (Nwafor, 2021). One of the main issues is that the labour force has a long history of patterned professional segregation, especially in terms of race and gender. It is not strange that these patterns are reflected in AI and automated tools, thus depending on these tools to make future decisions with regard to past events increases the possibility of repeating human biases and discrimination if one is not careful (Kim & Bodie, 2021).

However, Kimberly Houser (2019) instead advocates for more AI involvement regarding job recruitment in order to protect women. She states that “the tech industry and legal system have failed women miserably” and that women are grossly discriminated against with regard to employment possibilities within the tech industry (Houser, 2019, p. 351). As mentioned before, the tech industry is a very male-dominated sector and job employment has been biased against other genders for quite some time. According to Houser (2019), despite the numerous issues regarding AI such as the “black box” mystery of how the algorithmic outcome came to be, and the risk of reproducing human biases in automated form, she still believes that the employment sector would benefit from more AI decision-making as long as it is used responsibly. This author agrees that rather than leaving it up to human decision-making entirely due to AI possibly being harmful, it should still be used but further developed and updated in order to achieve minimal social and individual harm.

Racial discrimination through the use of AI and algorithms has been notably criticised since the beginning (Nwafor, 2021). Other than the previously mentioned employment sector, racial discrimination is prominent within sectors such as the

criminal justice system (Livingston, 2020; Završnik, 2020), health care (Omar, 2020), surveillance (Katyal, 2019), online visual representation (Makhortykh, Urman & Ulloa, 2021; Schlesinger, O’Hara & Taylor, 2018), and more. In the US, AI is frequently used within the criminal justice system, carrying out tasks such as setting bail, detecting fraud, and making suggestions for case investigations (Livingston, 2020). Predictive policing software is often based on “data of past locations of police responses to predict future locations of illegal activity” but this data is also over-representative regarding areas predominantly filled with people of colour (Livingston, 2022, p. 2). The famous COMPAS controversy case, as investigated by ProPublica, is another example of racial discrimination through AI where the automated risk assessment tool would unfairly label black defendants as repeated criminals almost twice as often as it would white defendants (Angwin et al, 2016; Eckhouse et al, 2019; Livingston, 2020). The issue is not only the fact that this occurred, but that the tool was still being in use after the investigation under the premise of efficiency. It is not uncommon that software or automated tools are used despite being faulty (Angwin et al, 2016; Eckhouse et al, 2019)

In 2017, Stanford University released a study that claimed that a computer algorithm could identify the sexual orientation of a person through facial detection technology (Nwafor, 2021). With this study came a lot of controversies, particularly from the LGBTQ+ community as the algorithm based its results on discriminatory stereotypes of gay and lesbian characteristics. It would for example claim to be accurately differentiating gay and straight men 81% of the time and 74% for women (Levin, 2017; Nwafor, 2021). Furthermore, not only was there outrage about the methods and results of this study, but many pointed out the potential risks this could have for the community, and how this type of research from an esteemed university such as Stanford could be used by cruel regimes in order to harass and hurt queer people.

3.2. AI regulation

Much research has been conducted regarding algorithmic accountability and its importance in the field of AI (Diakopoulos, 2014; Eckhouse et al, 2019; Neyland, 2007; Schuppli, 2014; Shah, 2018; Watcher, Mittelstadt & Russell, 2021; Wieringa, 2020). There are several issues with algorithmic accountability that make it difficult to exist in practice. According to Shah (2018), some of the issues of algorithms are that they reflect back the human biases, that there is limited transparency, there is not enough monitoring of the resulting impact, particularly with regard to vulnerable communities, there is a lack of right to challenge or redress, and a need of better governance. Shah also suggests that not only is there a need for more work regarding regulation with each regulator focusing on the “implications of machine learning algorithms in their area of work” but also a more diverse workforce in order to reduce biases (Shah, 2018, p. 3).

According to Harasimiuk and Braun (2021), AI is not a new concept but has yet to be regulated properly. There are different types of regulating, and algorithmic governance falls in line with a design-based regulation, meaning that it needs regulation “adjusted to the design of the entire regulated system” (Harasimiuk & Braun, 2021, p. 3). Furthermore, AI regulation attempts so far have been concerned with socio-political processes: as the AI field is increasing innovation at a rapid speed, the large tech companies such as Google, Apple and Amazon, have taken it upon themselves to create the Partnership on AI, with the goal to make the “best practices, research and public dialogue about AI” (Harasimiuk & Braun, 2021, p. 48) which may be dangerous to democracy if left in the hands of private companies. Thus, it is imperative that we create regulation with both governmental and non-governmental efforts on a larger scale.

In Europe, several EU resolutions and attempts at regulating AI have been made leading up to the current AI Act proposal, however, other countries have also moved towards creating a regulatory framework for AI. For example, in the US, the Obama

administration published a report on the future of AI, outlining various policy opportunities regarding AI, and that any regulation attempts should have a risk-based approach (Felten & Lyons, 2016). In 2019, the White House published a draft titled Guidance for Regulation of Artificial Intelligence Applications, with ten principles considering public trust in AI, risk assessment, fairness, and transparency that US agencies are to follow when creating AI regulation (Vought, 2020). Meanwhile, the UK published their plans for AI in their National AI Strategy of the UK, describing a 10-year vision of AI strategies involving three key points: promoting and developing the AI ecosystem while planning for long term needs, supporting the transition to an AI-enabled economy and ensuring the benefits that come with it, and making sure the people's fundamental rights and values are protected through national and international governance (UK Government, 2021).

3.3. Knowledge gap

The literature covered so far has entirely been focused on the dangers of AI, such as bias and discrimination in the workplace, in the criminal justice system, in academia and more, and the biases are often based on race, gender, ableism, sexual orientation and so on. Furthermore, the issues with AI and algorithmic accountability have been related to transparency, the lack of follow-ups of the system to ensure they are working properly, prioritising efficiency over fairness by knowingly using faulty AI systems, and a lack of diversity among AI creators.

There have been many national attempts at creating or planning to create legislation for AI yet for the EU in particular, this AI Act proposal is the first endeavour at regulating AI across a continent, with the aim of outlining harmonised rules for all Member states to adhere to (European Commission, 2021). However, since it is fairly new and the first of its kind, there is not much research regarding this specific proposal and how it frames accountability and social harm, which is the intent of this thesis. Some research have analysed and discussed the AI proposal at length with various aims such as sociological and legal (Schwemer, tomada and Pasini,

2021; Smuha et al, 2021; Sovrano et al, 2022, Stuurman & Lachaud, 2022; Veale & Borgesius, 2021), which will be further discussed in the analysis later. By analysing the proposal through critical discourse analysis, this essay will attempt to highlight how the EU have considered fundamental rights and values while ensuring the innovation and continuous development of AI in the EU.

4. Theoretical framework

This chapter will present the theory and concepts used when analysing the AI regulation proposal by the EU. The main theory used is Mark Bovens's public accountability theory, with interjection from various studies in the field of critical algorithm studies. The reason for choosing a public accountability theory rather than socio-legal theory is because this theory is better suited for a study heavily focused on accountability. One of the main issues with AI regulation is the aspect of accountability due to the fact that AI itself is often vague and confusing, even to its creators. AI, once created, is out of the hands of creators and developers and functions independently outside of them, which makes accountability difficult in situations where AI goes wrong. Socio-legal theories such as living law or law in books versus law in action may potentially offer good insight into the topic but this author felt that using Boven's theory would fit better with the aim of this thesis. Algorithms are everywhere, surrounding us in various shapes and forms. They are in our phones, our computers, our coffee machines, our cars and so much more. They are increasingly becoming of great importance to our day- to-day lives but are also flawed (Kemper & Kolkman, 2019). Whether it is regarding our online presence, our working lives or example, algorithms may fail to do the job it was intended for. This is why algorithmic accountability is needed, to counteract for example bias and discrimination resulting from algorithms (Wieringa, 2020), but before one tackles algorithmic accountability, the concept of accountability needs to be explored.

Bovens's theory of public accountability is a well-known and accepted theory that grounds itself on the idea that accountability may only be achieved if the behaviour is justified in front of an audience, which in turn has the opportunity to deliberate and decide the outcome of this behaviour. The responsibility of one's action needs to be appraised by an appropriate audience in order to condone it or determine potential consequences (Bovens, Schillemans & Goodin, 2014). The purpose of using this theory in regard to AI is to explore the ways accountability, and especially responsibility, is tackled by the EU in their new regulation proposal because AI and

algorithms are generally not transparent enough, thus explaining and justifying the behaviour of an AI or algorithmic system becomes difficult for both actor and audience. Additionally, the interjection of critical algorithm studies is useful for giving contextual analysis regarding accountability as well as suggestions for how to achieve algorithmic accountability, which is much needed for this thesis.

4.1. The rise of accountability

“Accountability is the buzzword of modern governance” (Bovens, Schillemans & Goodin, 2014, p.1). The concept of accountability may be traced back to the reign of William I, who shortly after the Norman conquest of England in 1066, demanded that each inhabitant who owned property would give a count of everything they owned. The collected information was then gathered in the Domesday Books with the purpose of organising a “foundation of the royal governance” (Bovens, 2005, p. 183) and ensuring that all property holders in this book were loyal to the king. This became the earliest recognition of holding someone accountable for something.

Meanwhile, the term accountability has been traced back to the early 1800s but did not have a cultural significance until the 1960s and 1970s, most likely due to its ties with literal accounting (Bovens, 2005; Dubnick, 2014). In the beginning, the term itself was connected to accounting and bookkeeping and did not hold any significant importance outside of this realm. Since then, this term has exploded with cultural significance and as such, the meaning and purpose of it had changed as well. When before the idea was that the people be held accountable by the leading forces in the country, it was now authorities who needed to be held accountable by the civilians (Bovens, 2005; Dubnick, 2014).

Due to the recent growth of accountability, Melvin Dubnick performed a scan of all legislation proposals given to the US Congress and found that with each two-year cycle, approximately 50 to 70 proposals included the word accountability in the title

(Bovens, 2005; Bovens, Schillemans & Goodin, 2014). This has of course led various institutions and organisations worldwide to incorporate accountability into their work as well, to the point of accountability in public discourse becoming a theme within academic scholarship and debates (Bovens, Schillemans & Goodin, 2014).

Historically, due to the idea of accountability having its relationship with accounting, a plethora of academia have tried to incorporate the term into their scholarship which has led to confusing definitional differences, making it difficult to reach a consensus on what it actually entails. In social psychology, the term has a “relational and communicative core” where it is defined as “the expectations that one may be asked, often by an authority or one’s superior, to justify one’s thoughts, beliefs or actions” (Bovens, Schillemans & Goodin, 2014, p. 4). This definition is shared by most social psychologists; though some of them do not necessarily agree with using its formal definition, the essence of it inevitably shows itself in their research and experiments.

Furthermore, even some accountancy literature agrees with this definition, however, where social psychology directs attention at the communicative action between actor and audience, accountancy instead connects this definition to the practical aspects of accounting such as reporting, book-keeping, audits, and reviews. Meanwhile, literature on public administration, the study of the organisation and implementation of government policies, disagrees; where the above-mentioned disciplines focus on non-public informal forms of accountability, public administration shifts attention to the public and formal kind of accountability by taking it to a structural level, and so does international relations which spotlight the governmental agents, NGOs, politicians and more. Political science often views accountability as an issue of power, where the focus lies on the relationship between politicians and their voters (Bovens, Schillemans & Goodin, 2014).

Each discipline and its following scholarship described above have a different focus and approach to the concept of accountability, yet they share a general theme. They all view accountability as explaining and giving answers to one's behaviour. It describes a situation where an actor has to answer for something to an audience. Therefore, Bovens's (2007, p. 450) official definition of accountability as "a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgement, and the actor may face consequences", is the definition adopted for this thesis as it gives a holistic view of accountability.

4.2. "Public" accountability and other types

The reason public accountability theory has been chosen is that it is fairly broad. Concerning the purpose of this study is how the EU frames accountability and social harm in their new AI regulation proposal, one cannot simply pick one type of accountability since there could be signs of legal accountability, political accountability, social accountability, economic accountability and more. A broad, blanket theory that may be narrowed down once the analysis of the material is completed is a fitting strategy for this thesis since no predictions can be established beforehand.

Since the definition has been established, there are a few concepts that need to be explained. The "public" in public accountability, for example, is not just the notion that it involves the public domain, though that is one aspect of it. Public, according to Bovens, Schillemans & Goodin (2014), is connected to transparency and openness. Oftentimes, accountability is upheld in private and done regularly: every day we are held accountable to others such as to our parents, our friends, our bosses, and our neighbours, but public accountability regards involving the general public. If an actor does something that they need to be held accountable for, the explanation of the conduct and the following debate is to be open for the public forum (Bovens,

Schillemans & Goodin, 2014). What qualifies as an actor or a public is related to the sphere in which the accountability is raised; a religious person may be held accountable by the deity they believe in, the CEO of a company may be held accountable by his employees, or the president may be held accountable by his people. The idea of “public” is built on whoever the actor has a relationship with and thus feel an obligation towards them (Bovens, 2005).

Public accountability has several functions; the main function is democratic control. In order to have democratic governance, it is crucial to make sure that the democratic process has public account, thus ensuring that the agents with power of, for example, a country, are controlled by the power of the voters who are able to judge their work and vote them out if necessary. The second function is to “enhance the integrity of public governance” where it is expected that through public accountability, behaviours such as corruption, nepotism, and abuses of power are hindered as the forum will oversee the behaviour of the actors (Bovens, 2005, p. 193). It is unclear how to proceed if the forum itself is also displaying the same behaviour as the actor such as racism or sexism: supposedly, there may be another actor-forum relationship on a larger scale, for example if a smaller collective group such as a hate group is the approving forum in one situation, they may be the actor in another where they have to justify their beliefs in front of the nation instead.

The third function is prevention and improvement, in which not only is behaviour controlled but public accountability is also meant to adjust norms and behaviours for the future, ensuring that the cycle does not continue. Finally, the fourth unofficial function that is birthed through the first three functions is to cultivate and improve the legitimacy of public governance. Through public accountability, the gap between representatives and residents should be bridged through transparency, receptivity, and liability (Bovens, 2005).

Depending on the discourse, accountability can be viewed in two ways: accountability as a virtue and accountability as a mechanism. As a virtue, accountability is an advantageous element of one's character, meaning most people will inevitably strive for it. Bovens, Schillemans and Goodin (2014) give examples of various legislations in the past such as the Syria Accountability Act and the United Nations Voting Accountability Act that do not specify the components with which to enact the law but rather describes what the purpose of it is. Legislations like those are not about detailing the how but rather the what and why, where accountability is connected to a responsibility to comply, making it clear for the agent what "good" conduct and behaviour is. Accountability as a mechanism, however, deviates from the focus on the behaviour of agents, to how institutional situations control and guide the behaviour of agents. From this perspective, what is important is how the agent can be held accountable by the forum post-conduct (Bovens, Schillemans & Goodin, 2014).

Interestingly, Bovens, Schillemans and Goodin (2014) argue that there is a difference between how accountability is viewed as either a virtue or a mechanism in terms of location. Accordingly, Americans have a tendency to view it as a virtue whilst Europeans see accountability as an instrument or a mechanism (Yang, 2012). For the EU, however, there is a tendency to create regulations that are broader in order to work transnationally for each Member State, thus it is impossible to predict whether they would approach accountability in their regulations as a virtue or as a mechanism. This will be further explored later in the thesis.

There are several types of public accountability, but they can be determined by these questions: who is to be held accountable, for whom and for what, by which standards and finally why? (Bovens, 2007; Bovens, Schillemans & Goodin, 2014). Whoever the accountable actor is can be easily identified in individual circumstances but becomes difficult when it concerns public organisations and institutions. For instance, incorporate accountability, the actor is the organisation

as a legal body, while in hierarchical accountability it is the leader, the CEO for example, of the organisation, and in collective accountability, each member of the organisation is equally accountable. The accountability forum in political accountability are political forums such as voters, parties and so on, while in administrative accountability they're administrative entities and agencies.

The nature of the conduct in constitutional law regards accountability for due process, while in financial accountability it concerns financial matters. The standards of accountability are formed by expectations that may be based on legal norms, professional norms, or political claims (legal, professional, or political accountability). Finally, the nature of the obligation regards the relationship between actor and forum, why an actor feels obligated to give an account. Mandatory accountability concerns the power of the forum released onto the actor, while voluntary accountability gives the actor the choice of whether or not to account for their conduct (Bovens, 2007; Bovens, Schillemans, & Goodin, 2014).

One type of accountability that sparks curiosity for this thesis is algorithmic accountability; as transparency and debate are essential aspects to accountability generally, it is intriguing how such polar opposites are put together in the same term. The next section will explore the field of critical algorithm studies and the concept of algorithmic accountability.

4.3. Algorithmic accountability

The various types of accountability described in the previous section give the reader examples of how the actor, the forum, and the relationship between the two, change depending on the perspective. Furthermore, they all involve the public domain in some way, as it requires transparency between actor and forum. However, transparency is of special concern to this author as when it comes to algorithms and AI, it is an aspect many have considered to be challenging to achieve.

Many different initiatives have attempted to regulate the situation of algorithmic accountability, for example, the Automated Decisions systems Task Force was put in place by the city of New York to investigate and assess the algorithmic systems, and in the Netherlands, the Dutch Open Government Action Plan incorporates a section on open algorithms (Wieringa, 2020). Furthermore, much academic literature has shyly approached the topic of algorithmic accountability, but the core issue remains: what truly is algorithmic accountability?

According to Wieringa (2020), algorithmic systems are not just technical objects but socio-technical systems which are embedded into culture. Just as with Bovens's theory (2007), these systems may be seen differently depending on the perspective. Wieringa (2020) explains that the algorithmic accountability 'cube' is three dimensional where the relationship goes through three phases: information-giving, discussion, and consequences, which greatly resembles the public accountability theory (Bovens, Schillemans & Goodin, 2014). However, she adds that there are underlying normative perspectives as well that make up this cube; the democratic perspective, which controls governmental actions and evaluates how accountability helped encourage better behaviour within the executive branch; the constitutional perspective which concerns the regulation of power and abuses of power in the executive branch, and the learning perspective where public offices and agencies are given feedback to boost efficiency (Wieringa, 2020).

She concludes with the argument that algorithmic accountability is dependent on the socio-technical system and the different stages of its life cycle. Rather than having one actor and one forum, there are several actors that have all in one stage or another made an impact on the system which means they are all responsible for the account. There are also multiple forums by which each actor is to be held accountable for either a section of the system or the system entirely (Wieringa, 2020). Frank Pasquale (2015) explores the idea of power through purposeful secrecy and the knowledge gap in his work *The Black Box Society* and uses the term "black box" to illustrate the mystery that happens within a system that we do

not or cannot know its inner workings. In certain situations, the secrecy is warranted for example when the matter revolves around national security, but there are too many occurrences where one may wonder why the secrecy is even present. As technology advances and the digital world expands, the data and process of algorithms become increasingly less transparent while the rest of the world gets more curious (Pasquale, 2015).

The issue of transparency becomes very difficult to solve. Due to the recent surge in big data usage and the complexity of algorithms, identifying one single actor's impact on the algorithmic system is challenging enough, let alone every actor involved. Kemper and Kolkman (2019) have explored this problem and claim that algorithmic transparency may only be acquired through having a critical audience. One guideline created by academia is the FAIR guiding principles, where FAIR is an acronym for findable, accessible, interoperable, and reusable. Through these principles, data should be easily identified globally, accessible by both humans and machines, interoperable when machine-actionable, and reusable later on. However, the FAIR principles are meant to improve data management within academia. Another guideline is the FACT principle, Fairness, Accuracy, Confidentiality and Transparency, often used for data science and storage and analysis of data in a broader spectrum (Kemper & Kolkman, 2019).

These two guidelines are examples of how different fields have set up guidelines for how to handle data, but transparency remains the key guiding factor for both. Still, Kemper and Kolkman (2019) have concluded that without a critical audience, there is no way transparency would ever be achieved. Firstly, any measures for transparency will remain as "empty signifiers" if there is no "critical and unbiased engagement" from an audience (Kemper & Kolkman, 2019, p. 2092), although this raises the question of whether there may ever be such a thing as unbiased engagement, particularly when it comes to who the audience is, their relationship with the system that is to be analysed, and perhaps even their general stance on

algorithms and advanced technology. Secondly, even if one has an audience, there are still issues with algorithmic complexities and the constant changing of the technological sphere with new models being released at a rapid pace. Thus, there is a pressing need for more research regarding algorithmic accountability in order to create guidelines that function well for all types of systems (Kemper & Kolkman, 2019). As mentioned in the introductory chapter, Smuha (2021) pointed out social harm as one of the various harms that is often neglected, especially by policymakers, thus it is even more important that accountability and individual, collective and social harms are all considered in any AI regulation.

5. Methodology

The methodological position for this analysis will be described in this chapter. The focus is the EU proposal for regulation regarding AI, with the aim of analysing how the regulation proposal views accountability and individual, collective, and social harm. There are many forms of discourse analysis that may be applied over a range of disciplines; however, the critical discourse analysis (CDA) by Fairclough (2013) is best suited for this paper considering its ties with a sociological perspective. By applying this method of analysis to the regulation proposal, the hidden social practices and beliefs may be revealed, and whether or not this proposal was written with the intention of preventing harm or if it was written with other intentions, as well as the framing of accountability in those cases. This paper will analyse the proposal with a top-down approach with a deductive approach through various concepts such as order of discourse, ideologies, textual analysis, situational context, intertextuality and more.

5.1. Fairclough's Critical discourse analysis

The purpose of discourse analysis (DA) is to study language use in relation to its social context. There are many different variations of discourse analyses, some that analyse the texts in detail and others that focus more on other aspects such as the social practice. Within the field of social sciences, discourse analysis is often heavily affected by Foucault and therefore tends to not concentrate much on the textual or linguistic aspects (Fairclough, 2003). Regarding socio-legal methodology, other methods could have been useful in order to study this phenomenon such as context analysis, a close-reading of texts or perhaps a combination of several methods. However, critical discourse analysis byways of Fairclough (2003; 2013) has been chosen as it situates itself between social theory and linguistics, effectively bridging the gap and observing the effect they have on each other. Due to this thesis's focus on accountability and social harm in the new EU AI act proposal, it seemed fitting to use an interdisciplinary approach.

A discourse may be defined in different ways, one of which is as a practice that systematically organises the subjects it refers to. From this perspective, all discourses are context-based and socially determined (Määttä, 2007). Fairclough (2013) claims that CDA contains three attributes: it is relational, dialectical, and transdisciplinary. By relational, he means that the focus should lie with social relations and that a discourse is just as complex, both within a discourse and between discourse and other objects in the world such as people, power, and institutions.

Discourse cannot be identified and understood by itself, but can only be understood through analysing its relations, which are in turn dialectical. By this, he means that dialectical relations are different from each other yet not entirely separate. He gives the example of power: power is discursive and depends on the language in order to be executed but may also be expressed through physical force. Power cannot be boxed in as a discourse alone but depends on the situation in which you encounter power, which is why Fairclough (2013) argues that elements such as power must be analysed through its dialectical relations with other objects. Finally, it is transdisciplinary in the sense that the purpose of CDA is not to analyse discourse itself but rather the relations, thus it is an interdisciplinary form due to relations being situated in various disciplines such as sociology, linguistics, and politics, and because of this, it demands a transdisciplinary methodology as well (Fairclough, 2013).

5.2. CDA method

A transdisciplinary methodology does not only concern the literal methods of analysis but “a theory-driven process of constructing objects of research” (Fairclough, 2013, p. 5). Through this, the researcher has to design a theoretical framework that converges different theories in order to help pinpoint issues with

translations between the concepts in CDA and the topic at hand. For instance, this study will be looking into accountability within the EU AI law proposal and according to Määttä (2007, p. 166), "...law, by definition, represents, shapes, and codifies the values and ideologies of a society: law is the central site of power and regulates all discourse". This perspective is the positivist understanding of law where law is society's guiding force. Of course, one could argue that this understanding can be questioned as not all laws weigh as heavily as others. Perspectives other than law as society's guiding force may see law as culture, law in context or law as rules; there is also an entire field dedicated to comparative law (Örücü, 2007). Thus, it becomes evident that looking into how the different fields identify a particular type of discourse is crucial before trying to analyse it.

When it comes to the actual CDA method, Fairclough's three-dimensional model (Chouliaraki and Fairclough, 1999; Winther-Jorgensen & Philip, 2002) describe every moment of language use as a communicative event, and suggests looking at textual analysis, discursive practice, and social practice as separate but interconnecting dimensions. On the micro-level, a textual or linguistic analysis is done, looking at for example word choice, syntactic analysis, use of various language devices and so on. Then on the meso level, the discursive practices revolve around the production and consumption of the text and firmly establish themselves as the mediators between texts and social practice. At this stage, under investigation is the author(s) of the text, who was involved in the production, what readers were included or excluded from the target audience, and more. Finally, on the macro level are the social practices, where the researcher is focused on the intertextual elements, looking at aspects such as power relations, hegemony, ideologies and which traditional or cultural values are embedded in the text (Chouliaraki & Fairclough, 1999).

Furthermore, Chouliaraki and Fairclough (1999) also describe the two main focal points in any analysis is the intertextual and situational contexts. The situational

context is related to the discursive practices, meaning that in order to do a situational context analysis, one must focus on when, what, how, who and why aspects which may be done by looking at the communicative event, such as a newspaper article or an interview video, and the order of discourse. The order of discourse is “the configuration of all the discourse types which are used within a social institution or a social field” (Winther-Jorgensen & Philip, 2002, p. 67).

Intertextuality, meanwhile, refers to the ways that texts relate to each other, whether it be explicitly, implicitly, or otherwise. It is a common analytical tool within CDA methodology to use when investigating how earlier communicative events influence current events, where we often repeat the same language, phrases and texts. While discourses do change, there is still an effect of earlier discourses on our language today (Chouliaraki & Fairclough, 1999).

5.3. Data collection and analysis

Since the aim is to investigate how the EU has approached the issues of accountability and various types of harm in relation to AI through the proposed AI Act itself, the main material used in this analysis is the AI Act itself. First published in April 2021, this proposal came after several EU resolutions and discussions regarding AI legislation and how one would go about its creation. As such, other materials used (although not heavily focused on) in the analysis are the four EP resolutions on AI covering aspects like ethics (European Parliament, 2020a), civil liability (European Parliament, 2020b), intellectual property rights (European Parliament, 2020c), and AI usage in criminal matters (European Parliament, 2021). Furthermore, news articles and research articles that were published after the release of the AI Act proposal as a response to it, were also briefly analysed as part of the general discourse about the regulation proposal (Schwemer, Tomada & Pasini, 2021; Smuha et al, 2021; Sovrano et al, 2022; Stuurman & Lachaud, 2022; Veale & Borgesius, 2021).

Each document came out between 2020 and 2022 as AI regulation in the EU is a recent happenstance, this proposal being the first attempt at an all-encompassing AI legislation in the EU. While the other documents were only briefly analysed, the proposal itself was deeply analysed using the Fairclough CDA method. First, the structure of the texts was examined and how the arguments were organised, essentially looking at the text on a macro level. This was followed by examining discourse fragments, meaning individual statements that share a specific code, such as “human rights” or “risks” in order to understand which “truth” has been assigned to this code. Thereafter, the linguistic features and rhetoric of the text was analysed, looking at word groups, grammar features, direct and indirect speech and so on (Schneider, 2013). This is all considered the first step of Fairclough’s CDA which is the textual analysis.

As the discursive practices concern the production and consumption of the text, the next step was to analyse and identify the cultural references and how the context affects the arguments as well as how the text refers to other texts, meaning intertextuality. Furthermore, a background check was performed as part of understanding the production process and how this text came to exist. The social, historical, and political context was examined and how this text is part of a broader debate about AI, the response to the text, and so on. Finally, elements of power, ideologies and hegemony was searched for in the text in order to see if any social practices or values were embedded in the text (Schneider, 2013; Chouliaraki & Fairclough, 1999).

5.4. Methodological limitations

The research design chosen for this thesis fits well with its purpose but does come with some limitations that one must take into consideration. One factor is that CDA as a method is a great analytical tool but heavily depends on the researcher’s ability to conduct it properly. Researcher bias is a common issue as the conducted analysis

is vulnerable to the researcher's own beliefs and ideas. As such, this researcher has attempted to combat this by first analysing the text of the main material, the EU AI proposal, without looking into any other discourse surrounding it as it may influence how one would read the proposal.

Furthermore, for the remaining documents involved in the discourse, they were found using a non-probability sample and were not randomly picked. In order to further avoid bias, the documents and articles chosen showed varying opinions as a response to the EU proposal to ensure that differing opinions were highlighted. Although, one may consider the idea that full objectivity is never completely possible in any type of research; Haraway (1988) discussed the concept of objectivity and bias and came to the idea of situated knowledge. According to Haraway (1988, p. 576), feminists have been “trapped by two poles of a tempting dichotomy on the question of objectivity” where on one end is the traditional scientific perspective on objectivity and on the other is a strong social constructionist perspective that claims that all processes towards knowledge are done with the motive of power rather than finding out truths. By situated knowledge, she aims to show that rather than being so concerned with biases and how to eliminate them, we should instead take responsibility for them and engage with them (Haraway, 1988). Hence, even though the research here has applied methods to combat subjectivity, it should be noted that full objectivity is never possible in any form of research.

5.5. Ethical considerations

This thesis follows the ethical guidelines provided by the Swedish Research Council (2017). Accordingly, there are four ethical demands that needs to be fulfilled in order to conduct a good research practice, namely information, consent, confidentiality, and use of data. As this study is focused on analysis of EU regulation, which are all publicly accessed and available, the need for informing the creators, obtaining consent, and ensuring confidentiality is not necessary. The final

guideline is fulfilled considering that the data will only be used for research purposes. Thus, this methodology and analysis were conducted according to the recommendation provided by the Swedish Research Council (2017) and does not infringe upon any ethical aspects.

6. Analysis

The analysis of the EU AI Act proposal will be presented in the following manner: part one will go through the attached explanatory memorandum, the structure of the proposal, and the definitions used for various concepts such as AI. Part two will present how human rights, bias and discrimination have been approached through the proposal and what measures have been taken against it. Part three concerns how the EC has framed trust and conformity, both in the ways to achieve harmonisation across the EU and the support and protection of providers and users of AI. The fourth part will go through the legal aspects and the obligations put upon the creators and providers, through concepts such as transparency and sanctions.

6.1. Introductory sections

6.1.1. Explanatory Memorandum

The 21st of April 2021, the EC introduced the long-awaited EU AI regulation proposal. After long discussion and debate, the EU finally attempted to lay down harmonised rules regarding the development, placement, and usage of AI systems in the Union market. The regulation begins with the attached Explanatory Memorandum and should be highlighted as it attempts to explain the circumstances in which the proposal was created. The sections explained are the context, the legal basis, results of evaluations and stakeholder assessments as well as budgetary implications (European Commission, 2021). It was clearly intended that this memorandum would explain the reasoning of the EU behind the proposal, and according to Chouliaraki and Fairclough (1999), context is crucial when conducting a CDA. However, this section raised more questions regarding the intentions of this proposal with regards to what it would truly demand from providers and users of AI.

Firstly, the context (section 1.1.) outlines that the reasons for the regulation were “to strive for a balanced approach” which entails making sure that the economic and societal benefits of AI are reached while simultaneously ensuring that fundamental human rights and values are protected (European Commission, 2021).

However, it is only vaguely explained in which ways those reasons would be met, or even what they truly are. This section was heavily focused on highlighting the urgent need for legislation regarding AI to protect both the creators and providers, as well as the users, but they never outlined any examples of the possible dangers of AI. The analysis showed that the language choice in this section was deliberately trying to emphasise the benefits the EU Union market would receive with AI while remaining strong on the idea that fundamental human rights would be protected at all costs. In fact, every paragraph contains both positive phrases such as “benefits”, “opportunities”, “trustworthy” and “safe” while also showing consciousness of the negatives such as “risk-based”, “high-risk”, “consequences” and “problems”, and that they were going to “mitigate”, “minimise” and “restrict” each negative (European Commission, 2021). The balanced approach used in the proposal is without a doubt also present in its choice of words.

Secondly, the regulatory framework of this proposal outlined four specific objectives:

1. Ensuring that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values.
2. Ensure legal certainty to facilitate investment and innovation in AI.
3. Enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems.
4. Facilitate the development of a single market for lawful, safe, and trustworthy AI applications and prevent market fragmentation (European Commission, 2021).

Once again, both the positives and the negatives were highlighted, with nearly equal focus on encouraging AI development and innovation, and the strengthening of protection of human rights. These objectives are economically, socially, and legally committed and bring up themes such as trust, safety, respect, rights and values,

productivity and efficiency, and the law. Through these objectives, this proposal aims to have a balanced framework where AI regulation covers all sectors and all dangers. This is not the case, as will be explained further along in the essay.

Finally, prior to the existence of this proposal, the EC have reportedly completed “extensive consultation with all major stakeholders” (section 3.1.) and 1215 entities were involved, of which 352 were companies or business organisations, 406 came from individuals, 152 from academic/research institutions and so on. The largest group were the 406 individuals (European Commission, 2021), yet it is unknown who these individuals are, where they are from, and how they were chosen for the consultation. Furthermore, from civil society, there were 160 opinions contributed but these were mainly consumer organisations, non-governmental organisations, and trade unions. It is unclear whether the public had a fair say in the consultation as most of the stakeholders were inevitably businesses and organisations that would most likely benefit from AI legislation being more lenient than harsh. As such, this paragraph inadvertently displayed a sense of inequality and power imbalance due to its obscurity concerning who was involved in influencing this regulation proposal, especially since many of those who were involved seemed to be those who would benefit from an unrestricted AI regulation. In terms of accountability, the audience seemed to only be partly represented by the stakeholders, which consisted of the beneficiaries, indicating a type of power imbalance.

6.1.2. Structure of the proposal

As previously mentioned, the proposal begins with an Explanatory memorandum, containing several sections such as context and legal basis, and a section describing the structure of the regulation. The actual proposal is divided up into specific provisions to easily navigate through the areas of necessary AI regulation, beginning with general provisions (Title I) where the scope and definitions of AI are defined including defining who the key participants are, meaning the providers and users of AI in both public and private sectors. Title II covers prohibited AI practices, as in the most unacceptable dangers of AI, compiling a list of AI systems

that go against Union values by for example violating human rights or manipulating and exploiting people in vulnerable groups. Title III details the rules regarding high-risk AI systems that threaten health and safety, fundamental human rights, and more. This is the most developed section in the regulation, containing several chapters clarifying classification rules and legal requirements and setting obligations on providers of high-risk AI systems (European Commission, 2021).

Title IV explains the transparency obligations for certain AI systems, specifically, those that interact with humans, detect emotions and social categories via biometric data, or generate or manipulate content such as “deep fakes.” Title V describes the measures in support of innovation. Title VI, VII and VIII concern governance and implementation, where the first title clarifies the established governance systems at both Union and national level, the second means to aid monitoring work of the Commission and national authorities through the creation of an EU-wide database for high-risk AI, and the third explains providers obligation to monitor and report AI systems and AI-related incidents and malfunctioning (European Commission, 2021).

Codes of conduct is referred to in Title IX where a framework is set up to encourage providers of non-high-risk AI to take initiative and freely apply the obligatory rules for high-risk AI systems. Finally, Titles X, XI and XII set out to clarify the all-round obligation for the confidentiality of information and data and exchange of information, the approach regarding how to ensure effective implementation of the regulation, delegation and implementing powers, as well as the obligation for the Commission to routinely check, evaluate and review the regulation for needs of improvement (European Commission, 2021).

Through the analysis, the structure of the proposal seemed to have been given due consideration, beginning with the understandably needed general definition of AI, then going through all the types of risks, followed by measuring, governing, and evaluating AI. It was easy to follow, the headings and subheadings were clear and

precise, and despite being only a summary of the proposal's provisions, it gave the intent that it would cover all the important aspects that this regulation needed to cover. When considering the historical context of the EU attempting to create legislation for AI which showed itself through various resolutions, it seems through the textual analysis that this proposal was thoroughly reflecting its history. The leading issue with regulating AI was how to define it, which later led to identifying which AI had the most or least risks, and finally how to measure, govern and follow up.

6.1.3. Defining the problem of AI

One of the most prevalent issues with AI regulation is the definition: throughout history, there have been several different types of definitions for AI (Harasimiuk & Braun, 2021) which made it even more difficult to attempt to regulate it. This current regulation proposal aims to set clear-cut definitions regarding not only AI, but other concepts used in the proposal through Article 3. The EU's attempt begins with Article 3 (1) where it states that an AI system refers to:

software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with (European Commission, 2021).

Annex I then provides a detailed list of various technologies that are recognised as AI, for example, machine learning (supervised, unsupervised and reinforcement learning, as well as deep learning) and “logic-based and knowledge-based approaches, including knowledge representation, inductive logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems” and “statistical approaches, Bayesian estimation, search and optimisation methods” (European Commission, 2021). This definition is openly broad: it does not cover all AI software but attempts to, at the very least, cover *some* aspects of

most AI. The issue with this is the same issue that has existed for years back: how can you efficiently regulate AI when what constitutes as AI is only vaguely explained? As the first definition of AI was already broad, the supporting Annex I is meant to clarify which software is included yet continues to remain obscure and expansive.

Furthermore, by having such a broad definition, there is a tendency to have loopholes come up along the way, especially if the development of AI is rapidly occurring at this very moment. With each new AI system, it is impossible to say if this definition will continue to apply.

The definitions for provider and user are just as vague, whereas according to Art. 3 (2) and (4):

‘provider’ means a natural or legal person, public authority, agency, or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge.

‘user’ means any natural or legal person, public authority, agency, or other body using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity (European Commission, 2021).

According to the definition, one is only considered a provider if they are putting the AI system in the market or in service under a name or trademark. This excludes any other AI developer that does neither: according to Smuha et al (2021) one group that the definition of provider does not account for is the academic researcher. It is unclear because an academic researcher does not fall under the two aforementioned categories yet may be a natural person who develops AI for first use directly to the

user or for own use, free of charge. Another group that is curiously not mentioned in the regulation is national security and intelligence agencies. Smuha et al (2019) remarked that Annex III (6, 7, 8) made sure to exclude AI systems developed for law enforcement, migration-related administrative agencies, and judicial authorities, but did not specify national and international security or intelligence agencies exemptions or involvements. It is unclear whether the EU is going to hold those groups accountable for harmful AI production and uses when they are not covered by the proposal.

Regarding user, this definition excludes people who are using AI privately which means that the regulation would not apply to private users. This begs the question of how the protection of fundamental human rights is going to affect private users as well as individuals who are not users but have been subjected to the AI system by a private user. Both aforementioned definitions seem to be focused on groups of people who provide or use AI for business purposes only: what would happen if a private AI user got access of an AI system on the market and use it non-professionally but ends up harming an individual or a group through for example discrimination? This regulation does not cover instances such as this which is a cause for concern. This is comparable of the private user exemptions in copyright and patent law; laws that are specifically meant to protect the rights of authors and creators' content but then allows for people to use and copy that content freely if for private consumption. This is something that is difficult to regulate in the current digital age, especially regarding European copyright law which is not binding for Member states who are also allowed to create national legislation on private copying (Cohen, 2005; Helberger & Hugenholtz, 2007).

Furthermore, 'putting it into service' is later defined by Article 3 (11) as "the supply of an AI system for first use directly to the user or for own use on the Union market for its intended purpose" (European Commission, 2021). Another definition for 'intended purpose' is written thereafter, and so on. With each definition as a response to a previous definition, it is difficult to say if the intent is actually to

define or if it is to confuse the reader. Naturally, the language in this proposal is legal and political, but the general audience reading this is not only from the legal and/or political spheres because the regulation supposedly concerns everyone from the individual to the user to business owners and creators. Yet, in Article 3 alone, there are 44 definitions that cover both important factors such as AI, provider, and user, and some less than important ones such as ‘intended purpose’, ‘serious incident’ and ‘performance of an AI system’ which according to Art.3 (18) states “...means the ability of an AI system to achieve its intended purpose” (European Commission, 2021). Thus, when trying to understand each definition, the reader may only get more misled rather than informed.

6.2. Human rights, bias and discrimination

6.2.1. Fundamental human rights

The EU AI Act proposal aims to protect fundamental human rights and values by addressing the various risks of AI and through a “clearly defined risk-based approach” (section 3.5 in the Explanatory memorandum). This includes aspects such as human dignity, respect for private life and personal data protection, non-discrimination and equality, freedom of expression, rights to a fair trial and more (European Commission, 2021). According to Smuha et al (2021), the three pillars of legally trustworthy AI are: the suitable allocation of responsibility for wrongs and harms in order to protect fundamental rights, the establishment of a framework that strengthens legal rights and secures rule of law, and the securing of transparency, accountability, and rights of public participation.

Harm is approached in various ways in this proposal, both in terms of different types of harm such as individual and collective harm, physical and psychological harm, harm of human rights and so on, and the harm regarding innovation and creation. Regarding Smuha et al’s (2021) pillars, one issue that comes up is the definition of AI systems under Annex III. As previously mentioned, the software covered by Annex III are restricted to three categories including machine learning software, logic-based and knowledge-based approaches and statistical approaches.

This definition, which is simultaneously both too broad and exclusionary, does not cover all systems that may harm human rights.

Strangely enough, the beginning of the regulation states that the purpose is to create “harmonised rules for the placing on the market, the putting into service and the use of artificial intelligence systems (‘AI systems’) in the Union” (Art. 1 (a)), which infers that this regulation will cover all AI systems (European Commission, 2021), yet entirely ignores low to medium risk AI. The regulation only covers prohibited AI systems in Title II and high-risk AI systems in Title III, and there is a distinct lack of other AI systems ever being mentioned. This is concerning because it is not clear what qualifies as a low or medium risk AI, or the impact it has on society. Moreover, there is no clear distinction for high-risk AI systems either; rather than attempting to define or classify, the regulation proposes a list of high-risk AI systems which this proposal covers, meaning there is no coverage for all the other high-risk AI systems that exist or will exist.

Veale and Borgesius (2021) analysed the AI Act proposal as well and found that the threshold for the harm requirement in the proposal was low, as well as that “manipulative AI systems appear permitted insofar as they are unlikely to cause an individual (not a collective) ‘harm’” (Veale & Borgesius, 2021, p. 99). Accordingly, this leads to increased opportunity for loopholes; for example, the proposal does not account for cumulative harm, as there are certain AI systems that may not explicitly cause harm in one event but rather slowly builds up harm over time (Veale & Borgesius, 2021). The analysis therefore shows that despite the language choice indicating that this regulation regards all AI, it deceptively only covers a limited number of systems.

Fundamental rights are mentioned in the four general objectives of the proposal, where not only do they specify that AI should be trustworthy and safe, but also that fundamental rights, Union values and safety requirements are to be protected when applied to AI (European Commission, 2021). The text in the proposal is heavily

embedded with the phrase “fundamental rights” and makes the reader feel that consideration of those rights has been appropriately performed such as in Article 53 (3):

[...] Any significant risks to health and safety and fundamental rights identified during the development and testing of [AI regulatory sandboxes] shall result in immediate mitigation and, failing that, in the suspension of the development and testing process until such mitigation takes place (European Commission, 2021).

Yet, when thoroughly looking into the document, any clarification of those rights is ignored, and no operationalisation is done. The EU Charter of Fundamental Rights requires that these rights are not seen as “interests of individuals to be balanced against the interests of others, including collective interests” and thus cannot be treated “as an afterthought” (Smuha et al, 2021, p. 10). This proposal has balance at the forefront, attempting to both lay down harmonised rules for the innovation of AI whilst protecting fundamental rights and values and as such, ends up treating fundamental rights as more of an oversight. The use of a risk-based approach does not help considering that the mechanisms described in the proposal are too technical and one-dimensional meaning they are not harmonious with fundamental rights (Smuha et al, 2021).

Discrimination and bias are two topics regarding AI that have been heavily argued about since before any regulation was attempted. The EU proposal claims under section 1.2. under the Explanatory memorandum that it complies with existing EU law on non-discrimination with a special focus on reducing algorithmic discrimination (European Commission, 2021). One argument that has come up during the analysis is the intent of the proposal. Both this author and other authors have concluded that the AI Act proposal seems to be more commercially minded rather than focused on safety and protection while also making it very difficult to actually regulate AI (Smuha et al, 2021; Stuurman & Lachaud, 2022; Veale &

Borgesius, 2021). One mechanism in the proposal that has been written about in regard to discrimination and bias is biometric systems. Under Article 3 (33), the definition states:

‘Biometric data’ means personal data resulting from specific technical processing relating to the physical, physiological, or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data (European Commission, 2021).

Biometric systems, such as facial recognition software and CCTV networks, may threaten existing fundamental rights established by the Union and have been classified as a prohibited AI practice. However, under Article 5, the use of ‘real-time’ remote biometric identification systems in public areas by law enforcement comes with several exceptions, such as Art. 5 (1d):

The targeted search for specific potential victims of crime, including missing children. The prevention of a specific, substantial, and imminent threat to the life or physical safety of natural persons or of a terrorist attack. The detection, localisation, identification, or prosecution of a perpetrator or suspect of a criminal offence... (European Commission, 2021).

It continues in the next paragraph of Article 5, where more exceptions are laid out with the urge to consider the consequences of the use of the systems, as well as the consequences of not using the systems (European Commission, 2021). It is written in such a way to warn the reader that while using biometric systems may be harmful due to risks of bias and discrimination against certain vulnerable groups, there are also risks in terms of security, particularly national and international security, that may arise with the absence of those systems. Furthermore, other shortcomings identified by Veale and Borgesius (2021) are that the provision does not ban EU vendors from selling biometric systems outside of the EU when they may be illegal

within the EU, only bans ‘real-time’ biometric systems which allow for systems that use identification post-event such as identifying people who were at a previous protest, and finally, it does not ban the usage of biometrics for purposes outside of the law enforcement such as public health and crowd control. These are aspects that are not immediately evident when reading the proposal, especially if one is not an expert at AI technologies, as the discourse in the proposal gives the impression that it is dependable and has the protection of human rights as a core goal. However, fundamental rights are non-negotiable and ignoring them or making exceptions cannot be justified for the sake of ‘national security’.

Besides the limited range of uses of biometric systems being prohibited, there is also a limited range of actors given restrictions (Smuha et al, 2021). Exceptions have been made regarding groups who are allowed to use whatever AI system they desire for their line of work. Not only do law enforcement have more lenient rules regarding their use of AI, but military forces, intelligence agencies and other agencies working with national security are not at all covered in the regulation. The ethical concerns of this will be discussed in the following section, but another concern is the lack of clarity regarding the justification of excluding these groups. Merely naming national security and public health as the reason is not enough to condone zero regulation for AI that may potentially cause other harms to both individuals, collective groups, and society as a whole.

Another contradicting point of the proposal is the lack of individual rights. Despite fervently insisting that fundamental rights are to be protected by the AI Act, there are no individuals mentioned in the proposal whatsoever. The right of redress and the complaint feature is completely missing. The lack of stakeholder information makes it unclear whether individual opinions have been represented in the process and instead, there is a heavy focus on AI provider and user obligations etc. For example, if a person has had their rights affected by an AI system, they have no ways of enforcing action, they cannot seek any type of redress or remedy, nor do they have any legal standing under the proposal (Smuha et al, 2021). Moreover,

there is no complaints system organised, meaning no individual can make a formal complaint with any authority regarding the impeding of their rights through AI. Indeed, the discourse of this proposal is riddled with discrepancies about fundamental rights and its importance in regard to AI.

6.2.2. Ethical implications

Prior to the AI act proposal being published, the EC had created a High-Level Expert Group (HLEG) in 2018, an independent group of experts whose purpose was to establish policy recommendations regarding AI (HLEG, 2019). The following year, the group published their Ethical Guidelines for Trustworthy AI, where they outlined three components that need to be achieved by any AI system throughout its entire life cycle: the AI needs to be lawful, following all laws and regulations relevant, needs to be ethical, following all ethical principles and values, and needs to be robust, technologically, and socially, in order to prevent unintentional harm. Furthermore, the four ethical guidelines identified by HLEG are: respect for human autonomy, prevention of harm, fairness, and explicability (HLEG, 2019).

Ethical principles tend to be reflected upon in most existing laws and regulations as a goal, however, the mechanisms later outlined to uphold the laws need to reflect them as well. It is simple to claim in the regulation that a core aim is to protect against harm but actually complying with it is another question. For example, under Art. 2 (3) of the proposal, AI systems developed for military purposes are exempt from the regulation entirely (European Commission, 2021). This exclusion raises ethical concerns, particularly in relation to lethal autonomous weapon systems, which HLEG have named a critical concern (HLEG, 2019; Smuha et al, 2021). To allow military AI to continue to exist and develop without any restriction is an unsettling thought as the repercussion could lead to “an uncontrollable arms race on a historically unprecedented level” (Smuha et al, 2021, p. 18).

Moreover, to ensure that AI does not undermine or impair human autonomy and fundamental rights, there is a pressing need for human oversight. Article 14 of the proposal outlines the requirements for human oversight (European Commission, 2021). and the language choice suggests that it is of utmost importance that human oversight is both built into the AI system as well as enabling individuals to understand, follow up and be able to intervene in any situation. Yet it is unclear as to who this applies to; is it the provider who has to follow up and ensure human oversight or is it the users? Additionally, is it even feasible to ask either party to be able to keep track of every AI system they provide and/or use, during every facet of the system meaning the development, deployment, and usage? It is difficult enough to ensure human oversight against the backdrop of the ‘black box’ enigma (Pasquale, 2015), which means much of the process inside the system is beyond human comprehension. Indeed, Article 14, as well as several other articles, may seem possible in theory but not in practice.

The discourse surrounding ethics in the proposal in comparison to the other texts analysed are concurrently similar and different. Each text puts ethical implications and concerns at the forefront, highlighting its importance and raising it as a critical issue regarding trustworthy AI. However, in the proposal, it is mentioned heavily in the explanatory memorandum whilst in the regulation itself, it is often put behind other aspects such as market progress and national security. The other texts (Smuha et al, 2021; Veale & Borgesius, 2021; Stuurman & Lachaud, 2022) responding to the proposal have all remarked on this disconnect between what the regulation proposal promises and what it actually delivers. While they commend the EU for attempting to create legislation for AI, they acknowledge the shortcomings and speculate whether there are other reasons behind the regulation, if there are other intentions prioritised over guiding the way for legally trustworthy and safe AI.

This author would have to agree with the responding discourse; one should be able to question whether the EU ignoring certain fundamental human rights is ethical, no matter the justification. The analysis showed that by making exceptions for

certain AI despite the risk level they show, such as biometric systems or military AI, right after promising to protect those very same rights, it is odd to see such contradictions in such an important document. It gives the feeling of negligibility, by sacrificing one type of protection in order to achieve another such as sacrificing the protection of one racial group in order to reach protection against terrorism. The implication taken from this action is that the EU might regard this as acceptable.

6.3. Trust and Conformity

The AI Act proposal spends a lot of time and effort on the concept of ‘trust’. The term is heavily found in abundance, used in multiple ways such as ‘trustworthy’, ‘an ecosystem of trust’, ‘earn [people’s] trust’, ‘establish trust’, ‘promote trustworthy AI’, ‘increase people’s trust’, ‘public trust’ and so on (European Commission, 2021). The ways the proposal uses the word trust can be divided into two; trust in the AI itself and how the EU will achieve trustworthy AI to put on the market which has been discussed in the previous section, and trust in the regulation by the Member States in order to reach conformity.

6.3.1. Harmonisation

The pressing need for harmonisation is evident in the proposal; in the Explanatory memorandum, the EU is fiercely advocating for Unionised ruling and for cooperation from every Member state. There is an acknowledgement that prior to the regulation proposal becoming public, some Member states have already attempted to create regulations regarding AI nationally in order to comply with fundamental rights obligations, and the EU has both commended the effort while simultaneously condemning it. Under section 2.1., it is explained that having national AI legislation would lead to two problems: “a fragmentation of the internal market” and “the substantial diminishment of legal certainty for both providers and users of AI systems”. Thus, in favour of continuing the circulation of AI products and services throughout the EU, there is a need for harmonised legislation (European Commission, 2021).

Through a horizontal approach, the EU is trying to ensure conformity with all Member states in order to protect and improve the internal market. The AI Act proposal is not the only proposal coming out and must be understood together with the Digital Services Act, the Digital Markets Act, the Machinery Regulation, and the Data Governance Act (Veale & Borgesius, 2021). The proposal draft is inspired by Article 114 of the Treaty on the Functioning of the European Union (TFEU) which is centred on the internal market, and this helps set a tone of cooperation being the sole solution, only this time with an emphasis on economic importance. The EU does not only want to create harmony and compliance but also to ensure the development of the internal market. This aspect of the discourse has been met with both appreciation and concern; dealing with AI in a manner that ensures a legally equal arena throughout the EU provides a certain safety for the people regardless of their residence, but it is worrying that throughout the proposal, more attention has been placed on improving the internal market rather than protecting fundamental rights (Smuha et al, 2021; Veale & Borgesius, 2021; Stuurman & Lachaud, 2022; Schwemer, Tomada & Pasini, 2021).

In order to make both the AI and the regulation proposal more trustworthy, the EU has set up a conformity assessment as a measure toward harmonisation. Regarding the AI systems, Article 40 states:

High-risk AI systems which are in conformity with harmonised standards or parts thereof the references of which have been published in the Official Journal of the European Union shall be presumed to be in conformity with the requirements set out in Chapter 2 of this Title, to the extent those standards cover those requirements (European Commission, 2021).

The following articles refer to the process of the conformity assessment but of special interest here is the heavy reliance on self-assessments. The proposal outlines the requirements for providers to perform this assessment which is fitting

considering the AI providers tend to be the most knowledgeable about the AI system (European Commission, 2021). This will be further explained in a later section; however, the issue lies with the concept of notified bodies. In order to have a system set up for conformity assessments, there is a need for an outsider party, a body that has to approve the assessment to make it valid. Articles 30 through 39 set the obligations regarding the notified bodies including that their main purpose is to verify the conformity of the provider's AI system as well as uphold the quality management system. Article 33 (4) especially states:

Notified bodies shall be independent of the provider of a high-risk AI system in relation to which it performs conformity assessment activities. Notified bodies shall also be independent of any other operator having an economic interest in the high-risk AI system that is assessed, as well as of any competitors of the provider (European Commission, 2021).

This is important because of the explicit mention of protecting against economic interests and competitors, which is rare for the regulation to do. This paragraph intends to ensure that the role of the notified bodies is to verify the assessment of the AI system without any influence from either the provider's side or the beneficiaries' side. The issue is that despite the inclusion of notified bodies as well as a whole system dedicated to them, there are hardly any mention of circumstances where the notified bodies are expected to perform. There are virtually no expectations from them which makes the designed system null and void. Article 43 (2) says that for all high-risk AI systems (excluding biometric systems): "...providers shall follow the conformity assessment procedure based on internal control as referred to in Annex VI, which does not provide for the involvement of a notified body." (European Commission, 2021). Due to the conformity assessment being reliant on self-assessment, this allows the provider to determine whether the AI system is conforming to the regulation entirely on their own.

Innovation support seems to also take more of a priority rather than individual protection. For example, Articles 53 through 55 specifically concerns measures in support of innovation, including the introduction of AI regulatory sandboxes. Art. 53 (1) says:

AI regulatory sandboxes established by one or more Member States competent authorities, or the European Data Protection Supervisor shall provide a controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan. This shall take place under the direct supervision and guidance by competent authorities with a view to ensuring compliance with the requirements of this Regulation and, where relevant, other Union and Member States legislation supervised within the sandbox (European Commission, 2021).

Much effort has been put into enabling the innovation of AI to continue with utmost support: Article 55 even gives extra priority to small scale providers and start-ups to the AI regulatory sandboxes, ensuring their success in the market (European Commission, 2021). Indeed, the EU makes several promises in terms of offers and protection for providers and users in order to gain conformity, which would then ensure the Member States be pleased with the results and stay allegiant. Furthermore, they also set obligations for the Commission and the EU as well in order to present themselves as trustworthy. Article 56 (2) explains how the EU created the European Artificial Intelligence Board with the purpose of assisting the EC with analysing “emerging issues across the internal market with regard to matters covered by this Regulation”, while Article 60 describes the creation of an EU database for stand-alone high-risk AI systems, including the fact that access to the database is open to the public (European Commission, 2021). The proposal is trying to show that the EU is taking active steps to make sure that they stay vigilant on the matter, monitoring and adjusting accordingly.

6.4. Legality and obligations

The main purpose of this regulation proposal is to create legally trustworthy AI, ensuring both innovation and protection of fundamental rights. As such, there are obligations outlined for the AI providers, users, Member states and the EU (European Commission, 2021). When reading the obligations, as this is a regulation that is supposed to bind all parties to it if it were passed to be law, the linguistic features were found to be strict yet amiable. Rather than using words such as “must” or “need” they say each party “shall” or “may” do something, for example, “Member States shall ensure that an appeal...” (Art. 45). With each obligation comes an explanation of reasoning, either in the same paragraph, the Explanatory Memorandum prior to the regulation or in the following Annexes (European Commission, 2021). However, it seems that the obligations are more often than not placed on the wrong parties.

The regulation does not only outline obligations for the AI providers and users, but also product manufacturers (Art. 24), authorised representatives (Art. 25), importers (Art. 26) and distributors (Art. 27). As previously mentioned in this chapter, several other groups have been excluded from the regulation such as law enforcement, the military and national security and intelligence agencies which is a legitimate concern because no responsibility for any harm has been assigned and there is no accountability if the harms occur. The AI providers and users who are covered in the regulation are given a list of obligations in Articles 16 and 29. In this section, it is explained that providers are obliged to create a quality management system to ensure compliance with the regulation, to draw up technical documentation, do a conformity assessment with each system, to name a few (European Commission, 2021). At first glance, the language of the proposal once again gives off an air of certainty and attention to detail, which is promising in regard to ensuring that the AI is trustworthy. But with a closer look at the language, it seems to be obscuring what those measures actually entail in terms of responsibility. For example, the regulation claims that AI systems need to meet data quality criteria and be safe regarding fundamental rights in order to be acceptable.

Article 9 goes through the risk management system set up to analyse high-risk AI systems, including the identification, estimation, and evaluation of possible risks as well as the adoption of suitable measures in cases where risks were identified. Art.9 (7) says:

The testing of the high-risk AI systems shall be performed, as appropriate, at any point in time throughout the development process, and, in any event, prior to the placing on the market or the putting into service. Testing shall be made against preliminarily defined metrics and probabilistic thresholds that are appropriate to the intended purpose of the high-risk AI system (European Commission, 2021).

It seems that all responsibility is put on the provider to perform the risk assessment of the AI system, yet simultaneously keeps it vague as to how exactly this assessment shall be done. There is no explanation of any sort of evaluation of the assessment nor where this assessment should be submitted. As previously mentioned, the notified bodies that are supposedly in place to judge the assessment are not actually enforced to do this because the assessment is based on internal control and decidedly up to the provider. If anything, the providers are able to determine entirely for themselves how ‘well’ their AI system has passed the assessment, especially regarding fundamental rights.

Furthermore, this paragraph also refers to testing against probabilistic thresholds that are appropriate to the intended purpose, which means that the provider needs to have the full knowledge and training to conduct the assessment whilst also knowing the user’s level of knowledge and training when using the system. Smuha et al (2021, p. 28) also noted this and concluded that the proposal leaves “an unduly large amount of discretion to the provider” in terms of the performing of the risk assessment. Since there is no authority checking to see if the assessment was done correctly, the providers may judge their systems’ performance as they wish.

This proposal relies heavily on conformity assessment as well, as articles 17, 19 and 43 have proven. Considering that providers already have too much power in assessing the risks of their systems, the enforcement of the proposal is also up to the provider's responsibility. This is a clear failure of the proposal to ensure accountability because it all comes down to the provider's judgement, who is the one who wants the AI system on the market or in service as soon as possible. It is also a clear sign of failing to secure transparency. Article 13 states:

High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent [...] An appropriate type and degree of transparency shall be ensured, with a view to achieving compliance with the relevant obligations of the user and provider... (European Commission, 2021).

The word "appropriate" is used often in the proposal, with no clear understanding of what is and what is not appropriate. Due to the responsibility being fully allocated to the providers, the transparency obligations written in articles 13 and 52 (European Commission, 2021) are essentially inconsequential. Using the term appropriate seems like a conscious choice, knowing that it is entirely up to the providers yet seemingly attempting to establish thought and concern regarding transparency for the sake of trustworthiness.

Regarding consequences, articles 70 through 72 explain the terms of confidentiality of information, penalties for failing to adhere to the regulation and administrative fines on Union institutions, agencies, and bodies as well. Art. 71 (1) claims:

In compliance with the terms and conditions laid down in this Regulation, Member States shall lay down the rules on penalties, including administrative fines, applicable to infringements of this Regulation and shall take all measures necessary to ensure that they are properly and effectively

implemented. The penalties provided for shall be effective, proportionate, and dissuasive (European Commission, 2021).

It seems that all responsibility is placed upon the Member States to create penalties as they choose, which is strange considering that the main goal of this regulation is harmonised rulings across the EU. The only requirements given are that the penalties must be effective, proportionate, and dissuasive which are relatively subjective terms. One Member State may consider only using financial sanctions while another may combine it with imprisonment. Furthermore, the following paragraphs explain the infringements which are administrative fines of up to 30 million EUR when there is non-compliance with AI prohibitions and requirements established in the proposal. There are too many possible discrepancies among the Member States, not to mention no sanctions given for the Member States should they not enforce the penalties at all.

7. Conclusion

The aim and research questions of this thesis was to investigate how the AI regulation proposal had approached and framed accountability and individual, collective and social harm. The results of the analysis of this regulation proposal were continuously two-faced: on one hand there were promises of harmonisation, transparency, safety, and protection for all, while on the other hand there was obscurity, misallocation of responsibilities and unequal distribution of power. The linguistic features of the proposal contained word choices that were uplifting, reassuring, and credible, in order to make the readers believe that they were dependable on the matter at hand. It had clear connections of intertextuality, constantly referencing other regulations and directives of the EU as well as showing consideration of issues brought up by previous resolutions, giving it a holistic perspective of all aspects of AI that needed to be covered by this proposal.

The other side of the proposal was fraught with inconsistencies and hidden segments that were overshadowed by the positive language. The definitions were too broad, confounded or sometimes missing entirely, fundamental rights were not prioritised ahead of innovation, several exemptions were made for parties such as law enforcement despite their issues in the past regarding AI and too much responsibility was given to the AI providers. There were clear signs of power imbalances and hegemony both in terms of giving certain groups more power to wield, as well as giving more weight to innovation and development ideas over safety and protection.

This thesis aimed to explore how accountability and individual, collective, and social harm were framed in the AI act proposal through a critical discourse analysis, and the results have shown that they were close to hollow shells wrapped in favourable promises. The concept of accountability in this proposal has been approached in several ways. Firstly, AI providers have been given the sole responsibility of ensuring that the AI systems are safe to place on the market or put into service. However, there is no mechanism in place that would guarantee that the

AI providers are doing as required, meaning there is no audience or forum in place to make sure that the actor is held accountable. There are also no requirements whatsoever concerning low or medium risk AI systems; the proposal has only laid down rules for prohibited and high-risk AI, and even then, only covered the AI systems in the pre-established list.

Secondly, in order to have significant accountability, there is a pressing need for transparency, particularly because AI and algorithms are often mysterious in how they process data. The proposal has attempted to ensure this through transparency obligations where AI systems need to be transparent enough that those subjected to them are fully aware and informed that they are being subjected to them. The AI providers have to be transparent by informing the users of every feature of the AI, and finally, an EU database has been created for stand-alone high-risk AI that is fully accessible by the public. The issue is that these obligations are not enforced in any way: the AI systems may be transparent but that does not inherently mean they will protect the individuals subjected to them, AI providers may inform about every feature of the AI but are only forced to do so with users of the AI, not the people that could be affected by them, and the EU database that has been established does not have all the information needed for the public to understand or even question the AI systems (Smuha et al, 2021). Therefore, every attempt at securing transparency for AI has not been thoroughly examined and at best, these obligations are just there as a formality for the readers of the proposal to feel like the EU has considered transparency.

Thirdly, accountability is concerned with three aspects: information-giving, discussion, and consequences (Wieringa, 2020). As mentioned, AI providers are obliged to give information to the users of the AI but also to conduct risk management and conformity assessment. Because these are self-assessments, there is no enforcement in place that guarantees that the providers are truly performing the assessments as intended. Furthermore, the sanctions outlined in the regulation are essentially determined by the Member States meaning differing sanctions may

occur for similar crimes. Nevertheless, the main concern with the aspect of consequences is the lack of individual rights. There is virtually no involvement of

individuals or the public in terms of questioning and contesting the impact of the AI systems on their lives. There is no complaint feature or any right to redress which implies that individual and collective harm is not of importance. As previously mentioned, this proposal seems to be more commercially minded and this is evident in the lack of input from the people.

Social harm is therefore not given sufficient thought. This type of harm tends to be forgotten behind individual and collective harm but is just as important, and it seems that this proposal has not prioritised all three. A societal perspective is crucial when creating regulations as they inevitably are for society's benefit. Yet because of the lack of intervention from the public, both prior to releasing the proposal through stakeholder consultation, and through individual rights in the regulation itself, there is a lack of democracy and a power imbalance present that may cause future displeasures and harms. Moreover, the exemption of law enforcement, the military and national security and intelligence agencies implies that their production and use of AI is necessary no matter the harm it can cause people. Despite the issues with for example predictive policing and biometric systems used by authorities for safety purposes despite biases and discrimination, the EU has supposedly prioritised public safety over individual or social harm. It gives the impression of an evil circle; in order to protect society, they will do everything they can even if it hurts society.

Consequently, the textual analysis showed that the language was clearly deliberate and aimed at presenting the regulation proposal as secure, considerate, and thorough. It was positive and uplifting whilst balancing both innovation and protection. The discursive practices showed clear intertextuality with several references to other resolutions, regulations and EU directives and situated this proposal among other proposals coming out regarding automation and digital

governance. Finally, the social practice which regards the implications of this regulation, showed power imbalances between providers and users, a lack of individual rights and protections, as well as prioritisation of innovation over fundamental human rights.

7.1. Future recommendations

The results of this thesis have shown that the regulation proposal lacks a societal perspective, and that accountability and individual, collective, and social harms were not given priority or thorough consideration during its creation. This study was conducted from a socio-legal standpoint and has been an enlightening process. For future studies, it would be interesting to see other perspectives such as international relations or public administration. This EU AI proposal is relatively new, just over a year old and as such, much research has not been conducted on what this proposal means for various aspects. Perhaps there could be research regarding the exclusion of low and medium risk AI and its implications, or how other EU institutions or officials perceive this regulation proposal. There are many paths one can take when examining this proposal and hopefully the EU will take any comments or criticisms as constructive and aim for improvement of the AI Act.

8. Bibliography

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. There is software that is used across the country to predict future criminals. And it is biased against blacks. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [Accessed 2022-01-19].

Banakar, R. (2019). *On Socio-Legal Design*. Lund University.

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European law journal*, 13(4), 447-468.

Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.). (2014). *The Oxford handbook public accountability*. Oxford University Press.

Calders, T., Ntoutsis, E., Pechenizkiy, M., Rosenhahn, B., & Ruggieri, S. (2021). Introduction to the special section on bias and fairness in AI. *ACM SIGKDD Explorations Newsletter*, 23(1), 1-3.

Caldwell, M., Andrews, J. T. A., Tanay, T., & Griffin, L. D. (2020). AI-enabled future crime. *Crime Science*, 9(1), 1-13.

Chouliaraki, L., Fairclough, & N. (1999). *Discourse in late modernity: Rethinking critical discourse analysis*. Edinburgh: Edinburgh University Press.

Cohen, J. E. (2005). The place of the user in copyright law. *Fordham Law Review*, 74, 347-374.

Darian-Smith, E. (2013). *Laws and societies in global contexts: Contemporary approaches*. Cambridge: Cambridge University Press.

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. In *Ethics of Data and Analytics* (pp. 296-299). Auerbach Publications.

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital journalism*, 3(3), 398-415.

Duarte, F., & Ratti, C. (2018). The impact of autonomous vehicles on cities: A review. *Journal of Urban Technology*, 25(4), 3-18.

Dubnick, M. J. (2014). Accountability as a Cultural Keyword. In Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.). *The Oxford handbook public accountability*. (pp. 23- 38). Oxford University Press.

Eckhouse, L., Lum, K., Conti-Cook, C., & Ciccolini, J. (2019). Layers of bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behaviour*, 46(2), 185-209.

European Commission. (2020). *White Paper on AI - A European approach to excellence and trust*. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf. [Accessed 2022-02-15]

European Commission. (2021). Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final, *EUR-Lex*, April 21, 2021, <https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX%3A52021PC0206>.

EUR-lex. (n.d.). *European Convention on Human Rights (ECHR)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:eu_human_rights_convention [Accessed 2022-02-24]

EUR-lex. (n.d.). *Non-discrimination (the principle of)* https://eur-lex.europa.eu/summary/glossary/nondiscrimination_principle.html [Accessed 2022-02-24]

European Parliament. (2020a). *European Parliament resolution of 20 October 2020 on a framework of ethical aspects of artificial intelligence, robotics and related technologies*, 2020/2012(INL).
<https://oeil.secure.europarl.europa.eu/oeil/popups/printficheglobal.pdf?id=710002&l=en> [Accessed 2022-02-15]

European Parliament. (2020b). *European Parliament resolution of 20 October 2020 on a civil liability regime for artificial intelligence*, 2020/2014(INL).
https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html#title1 [Accessed 2022-02-15]

European Parliament. (2020c). *European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies*, 2020/2015(INI).
https://www.europarl.europa.eu/doceo/document/A-9-2020-0176_EN.html#_ftn11 [Accessed 2022-02-15]

European Parliament. (2021). *European Parliament resolution of 6 October 2021 on artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters*. 2020/2016(INI)
https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_EN.html [Accessed 2022-02-15]

European Union. (2022). *European Commission*. https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/institutions-and-bodies-profiles/european-commission_en [Accessed 2022-04-18]

Fairclough, N. (2013). *Critical discourse analysis: The critical study of language*. Routledge.

Falco, G. (2019). Participatory AI: Reducing AI bias and developing socially responsible AI in smart cities. In *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)* (pp. 154-158).

Felten, E., & Lyons, T. (2016). *The Administration's Report on the Future of Artificial Intelligence*.

<https://obamawhitehouse.archives.gov/blog/2016/10/12/administrations-report-future-artificial-intelligence> [Accessed 2022-04-18]

Harasimiuk, D. E., & Braun, T. (2021). *Regulating artificial intelligence: binary ethics and the law*. Routledge, Taylor & Francis Group.

Helberger, N., & Hugenholtz, P. B. (2007). No place like home for making a copy: Private copying in European copyright law and consumer law. *Berkeley Technology Law Journal*, 22(3), 1061-1098.

High-level expert group on artificial intelligence. (2019). *Ethics guidelines for trustworthy AI*. <https://42.cx/wp-content/uploads/2020/04/AI-Definition-EU.pdf> [Accessed 2022-04-25]

Houser, K. A. (2019). Can AI solve the diversity problem in the tech industry: Mitigating noise and bias in employment decision-making? *Stanford Technology Law Review*, 22(2), 290-354.

Kammerer, B. (2022). Hired by a Robot: The Legal Implications of Artificial Intelligence Video Interviews and Advocating for Greater Protection of Job Applicants. *Iowa Law Review*, 107(2), 817-850.

Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, 22(14), 2081-2096.

Kim, P. T., & Bodie, M. T. (2021). Artificial Intelligence and the Challenges of Workplace Discrimination and Privacy. *Journal of Labor and Employment Law*, 35(2), 289-315.

Lambrecht, A., & Tucker, C. E. (2015, December). Can big data protect a firm from competition? Available at SSRN 2705530: <https://ssrn.com/abstract=2705530>

Levin, S (2017, September 9). *LGBT groups denounce 'dangerous' AI that uses your face to guess sexuality.* The Guardian.
<https://www.theguardian.com/world/2017/sep/08/ai-gay-gaydar-algorithm-facial-recognition-criticism-stanford> [Accessed 2022-04-13]

Livingston, M. (2020). Preventing Racial Bias in Federal AI. *Journal of Science Policy & Governance*, 16(2).

Makhortykh, M., Urman, A., & Ulloa, R. (2021). Detecting race and gender bias in visual representation of AI on web search engines. *In International Workshop on Algorithmic Bias in Search and Recommendation* (pp. 36-50). Springer, Cham.

Moss, H. (2020). Screened out Onscreen: Disability Discrimination, Hiring Bias, and Artificial Intelligence. *Denver Law Review*, 98(4), 775-806.

Määttä, S. K. (2007). The foundations of Europe: A critical discourse analysis of the EU constitution. *Critical approaches to discourse analysis across disciplines*, 1(1), 166-178.

Neyland, D. (2007). Achieving transparency: The visible, invisible and divisible in academic accountability networks. *Organization*, 14(4), 499-516.

Nwafor, I. E. (2021). AI ethical bias: a case for AI vigilantism (Allantism) in shaping the regulation of AI. *International Journal of Law and Information Technology*, 29(3), 225-240.

Omar, R. A. (2020). Unabashed Bias: How Health-Care Organizations Can Significantly Reduce Bias in the Face of Unaccountable AI. *Denver Law Review*, 98(4), 807-838.

Raub, M. (2018). Bots, bias and big data: artificial intelligence, algorithmic bias and disparate impact liability in hiring practices. *Arkansas Law Review*, 71(2), 529-570.

Schlesinger, A., O'Hara, K. P., & Taylor, A. S. (2018). Let's talk about race: Identity, chatbots, and AI. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 1-14). [https://doi-org.ludwig.lub.lu.se/10.1145/3173574.3173889](https://doi.org.ludwig.lub.lu.se/10.1145/3173574.3173889)

Schneider, F. (2013) How to Do a Discourse Analysis. *Politics East Asia*. <http://www.politicseastasia.com/studying/how-to-do-a-discourse-analysis/> [Accessed 2022-03-22].

Schuppli, S. (2014). Deadly Algorithms: Can Legal Codes Hold Software Accountable for Code that Kills? *Radical philosophy*, 187, 2-8.

Schwemer, S. F., Tomada, L., & Pasini, T. (2021). Legal AI Systems in the EU's proposed Artificial Intelligence Act. In: *Proceedings of the Second International Workshop on AI and Intelligent Assistance for Legal Professionals in the Digital Workplace* (LegalAIIA 2021).

Shah, H. (2018). Algorithmic accountability. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 376(2128), 1-6.

Smuha, N. A. (2021). Beyond the individual: governing AI's societal harm. *Internet Policy Review*, 10(3).

Smuha, N. A., Ahmed-Rengers, E., Harkens, A., Li, W., MacLaren, J., Piselli, R., & Yeung, K. (2021). *How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act*. Available at SSRN.

Sovrano, F., Sapienza, S., Palmirani, M., & Vitali, F. (2022). Metrics, Explainability and the European AI Act Proposal. *J*, 5(10), 126-138.

Stuurman, K., & Lachaud, E. (2022). Regulating AI. A label to complete the proposed Act on Artificial Intelligence. *Computer Law & Security Review*, 44, 105657.

UK Government. (2021). *National AI Strategy - HTML version - GOV.UK*. <https://www.gov.uk/government/publications/national-ai-strategy/national-ai-strategy-html-version> [Accessed 2022-04-18]

Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act - Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4), 97-112.

Von der Leyen, U. (2019). A Union that strives for more. *My agenda for Europe. Political guidelines for the next European Commission, 2024*, 2019.

Vought, R. T. (2020). *Re: Guidance for Regulation of Artificial Intelligence Applications*. Available online at: https://www.acr.org/-/media/ACR/Files/Advocacy/Regulatory-Issues/acr_comments_draft-OMB-memo_3-12-2020.

Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*, 41, 105567.

Wieringa, M. (2020). What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 1-18).

Winther-Jørgensen, M., & Phillips, L. (2002). *Discourse analysis as theory and method*. Lund: Studentlitteratur

Woods, L., Watson, P., & Costa, M. (2017). *Steiner & Woods EU Law*. Oxford University Press.

Yin, C., Xiong, Z., Chen, H., Wang, J., Cooper, D., & David, B. (2015). A literature survey on smart cities. *Science China Information Sciences*, 58(10), 1-18.

Završnik, A. (2020). Criminal justice, artificial intelligence systems, and human rights. *ERA Forum*, 20(4), 567-583.

Örücü, E. (2007). Developing Comparative Law. In Örücü, E., & Nelken, D. (Eds.) *Comparative law: a handbook*. London: Bloomsbury Publishing.