Πάντειον Πανεπιστήμιο Κοινωνικών & Πολιτικών Επιστημών

Panteion University of Social and Political Sciences

Ensuring Human Rights in AI Regulation: A Study of the EU AI Act

Athanasia T.Orfanioti

2025

MA in Digital Transformation

Specialization in Digital Law

Master's thesis

Academic Supervisor

Professor of EU Law and Policy

Panteion University of Social and Political Sciences (Athens) Greece

Vassilis Hatzopoulos

Table of Contents

1.INTRODUCTION

AI systems or automated data processing tools and services have become essential in our everyday lives. They have opened a whole new world of possibilities, like helping us to select a movie, or a book or a new music track that fits our preferences and personal styles. AI systems, also assist governments to control tax evasion, conduct police investigations, companies in making employment decisions, universities in accepting new students and the health sector to detect probable new diseases and cures. But on what basis can a person's profile match that of a potential criminal or terrorist? Which profile is more suitable for obtaining an enrollment in a prestigious university or an employment at a renowned firm? What kind of advertisements or information can we see first on our monitors or our social media newsfeeds? Which kind of requirements sgould someone fulfill to get a visa or a refugee status or a house loan? As we can realize, AI systems cannot only be used, by internet users in order to access new information and services, but can also be used in significant, even life changing, decision-making processes, that previously could only be in control of humans. AI systems can take decisions concerning people automatically, on behalf of them, and that creates great challenges, let alone serious dangers for society. Subsequently, AI systems can and indeed have a great impact on fundamental human rights, as we will analyze in this thesis, such as the right to life, to privacy, to freedom of expression, to a fair trial, to workers and students' rights, to free elections and can cause distortions in democracies, such as discriminations of minorities, inequality and the protection of human dignity,overall. The main reservations have to do with the fact that humans have no control over the use of them and far more they don't understand the technical details of their design and the extend of the sectors where they are being applied. In reality, many times they never learn that an AI System was involved in a third party's decision, that changed in any shape and form their lives and this is as frightening, as it can be.

In this sense ,we will firstly analyze,the legal framework that the recently promulgated Artificial Intelligence Act(thereof AI Act or the Regulation), of the European Parliament and the Council of Europe, designed and implemented concerning the AI technology, the risk based approach that it took concerning the deployment of AI systems in regard to human rights protection and how this approach is supposed to prevent and address in a significant and successful way, any possible violation. So,our analysis will be centered on the articles and recitals of the Act, that directly or indirectly engage with the human rights violations element, we will try to examine which fundamental rights are already being challenged by the AI systems use,establish their negative impacts,while at the same time presenting a series of well-known case studies where AI systems, violated.Through this examination, we aim to illustrate how the Act not only identifies risks but also operationalizes safeguards to ensure the ethical deployment of AI technologies.

<u>2.The EU AI Act:A first glance</u>

The European Union Artificial Intelligence Act,is a Regulation of the European Parliament and of the Council of Europe,that was adopted and published in the Official Journal of the European Union on the 12^{th} of July 2024 ,under the protocol number 2024/1689, which entered into force on the 1^{st} of August 2024 and it seems to mark a global milestone in AI regulation. Among its main objectives is to ensure that AI systems, put in the European Union market are used in a safe, lawful way that facilitates and enhances innovation in AI while at the same time, they are respecting human rights and values. It is a market regulation that, we can say, aims to harmonize AI system use across EU member states while addressing socioeconomic, ethical, and human rights implications. It actually puts human rights at the center of the discussion. It’s not just a market tool. It’s more like a human rights law, embedded in a market tool, that tries to protect society and its people towards the risks and harms of the new era, the era of AI. (Butt,2024).

As it is stated, in the first article of the European AI Act:“*The purpose of this Regulation is* ***to improve the functioning of the internal market*** *by laying down a uniform legal framework in particular for the development, the placing on the market, the putting into service and the use of artificial intelligence systems (AI systems) in the Union,* ***in accordance with Union values, to promote the uptake of human centric*** *and trustworthy artificial intelligence (AI)* ***<u>while ensuring a high level of protection of health, safety, fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union</u>*** ***(the ‘Charter’),*** *including democracy, the rule of law and environmental protection,* ***to protect against the harmful effects of AI systems in the Union, and to support innovation.****”*

In this paragraph, the idea is very clear. It calls explicitly upon the EU market and the Charter of Fundamental Rights of the European Union.These are the basic two elements, the market and EU values and principles concerning its citizens. The Regulation aims to be a market tool, strongly subordinated to the values of the Charter(Staghl et al,2022,Edwards,2021).

The Recital (2) paragraph, emphasizes more on the values of the Charter:

"***This Regulation should be applied in accordance with the values of the Union enshrined as in the Charter,*** *facilitating the protection of natural persons, undertakings, democracy, the rule of law…."*

Recital (5), mentions explicitly that AI use "…*depending on the circumstances regarding its specific application, use, and level of technological development,* ***AI may generate risks and cause harm to public interests and fundamental rights that are protected by Union law****. Such harm might be material or immaterial, including physical, psychological, societal or economic harm*."

Recital (6), digs deeper in the issue of human rights, reiterates the notion of "**Union values**», the "**fundamental rights**», the "**Charte**r" and for the first time, introduces the term "**human centric technology**», by the people, for the people, for society, for the wellbeing of us all. More specifically it states that, *«Given the major impact that AI can have on society and the need to build trust, it is vital for AI and its regulatory framework to be developed* ***in accordance with Union values as enshrined in Article 2 of the Treaty on European Union (TEU), the fundamental rights and freedoms enshrined in the Treaties and, pursuant to Article 6 TEU, the Charter.*** *As a prerequisite, AI should be a* **human centric technology**. <u>*It should serve as a tool for people, with the ultimate aim of increasing human wellbeing*."</u>

Recital (7), sets the importance to set guardrails in order to protect human rights, the public interest and mitigate the risks, again according to the "Charter disposing that "*In order to ensure a consistent* <u>*and high level of protection of public interests as regards health, safety and fundamental rights,*</u> *common rules for high-risk AI systems should be established.* <u>*Those rules should be consistent with the Charter*</u>*, nondiscriminatory……"*

Recital (8) of the Preamble, continues to highlight fundamental rights, this time in order to explain why a legal framework is requiring, in order to control possible harmful AI use by claiming that : "<u>***A Union legal framework***</u> *laying down harmonized rules on AI* <u>***is therefore needed to foster the development, use and uptake of AI***</u> *in the internal market* <u>**that at the same time meets a high level of protection of public interests, such as health and safety and the protection of fundamental rights, including democracy, the rule of law and environmental protection as recognized and protected by Union law.**</u> <u>*Those rules should be clear and robust*</u>

in protecting fundamental rights…. *By laying down those rules Regulation supports the objective of promoting the* ***European human centric approach to AI."***

All in all, the first article and the first recitals of the Regulation, set the scene very clearly. The Union aims to adopt the AI use and services, but only in a beneficiary way for the society. It wishes to promote innovation and technology but that will come with rules and prerequisites. The rules will be based on the unnegotiable human rights, the principal values of European culture and will secure a human centric, ethical and moral approach to AI technology.(Hristozova,2024,Pham &Davies,2024)

2.1. Definitions-Glossary

2.1.1 AI system

An AI system is defined in Article 3(1) of the Regulation, as "*a machine based system that is designed to operate with varying levels of autonomy and that may exhibit addictiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments;*"

As it is well known, Artificial Intelligence is based on algorithms, which can be defined as "a series of steps undertaken in order to solve a particular problem or accomplish a defined outcome " (Diakopoulos , 2015).AI systems are composed by a set of instructions or rules that take data as inputs, process them mathematically and return "outcomes" or results, many times simple words ,such as "approved», or "denied" "rejected"(Goodman & Flaxman,2017).An AI system can be exclusively hand coded by a human or generated automatically from "learned" data or metadata or from a number of other AI systems, as in machine learning. In other words, the creation of AI systems can be a manual process where each step is coded by humans or an automated one where AI tools generate themselves through machine learning., that we will analyze in a moment(Pasquale,2019).

As the preamble of the Regulation, states in recital (12)", a key characteristic of AI systems is their capability to infer. This capability to infer, refers to the process of obtaining the outputs, such as predictions, content, recommendations, or decisions, which can influence physical and virtual environments, and to a capability of AI systems to derive models or algorithms, or both, from inputs or data. The techniques that enable inference, while building an AI system, *include* <u>*machine learning techniques that learn from data on*</u> *how to achieve certain objectives, as well as logic and knowledge based techniques that infer from encoded knowledge or symbolic representation of the task to be solved*".

<u>2.1.2. Machine learning techniques</u>

Machine learning, is the technology that trains computers to perform a task by learning from examples, data and prior experience or shall we say knowledge. In this sense, machine learning AI systems concern computational methods that enable systems to automatically learn from data without explicit programming, for every specific task (Mittelstadt et al,2016). These AI systems, are basically designed to identify patterns, make predictions, and improve their performance over time through training on large datasets(AIHLEG,2019). There are several types of machine learning AI systems, including supervised learning AI systems, which use labeled data to train models for classification, unsupervised learning AI systems, which analyze unlabeled data to discover hidden structures and relationships, and reinforcement learning AI systems, which optimize decision-making, by learning from rewards and penalties. The effectiveness of these AI systems depends on the quality of the data and the appropriateness of the chosen AI System for the task. As machine learning AI systems continue to evolve, they improve numerous applications such as recommendation systems, autonomous vehicles, and predictive analytics, while also raising concerns about biases, transparency, and accountability in their decision-making processes, as we will see. (Sandvig et al,2016).

In Recital 12 of the Regulation, the machine based technique is defined as a reference "to <u>*the fact that AI systems run on machines. The reference to explicit or implicit objectives underscores that AI systems can operate according to explicit defined objectives or to implicit objectives.*</u>"

AI systems that use machine learning techniques can identify, detect and extract digital data and traces from individuals' internet behaviors and general patterns of activities and produce new information, for new decisions(AIHLEG,2019). Moreover, AI systems, have the ability to make decisions, with no human involvement or with little human involvement. The data, used in AI systems, can be data provided by the individuals themselves, they can be data collected by observing the individual's activity (for example, tracking their location) and lastly they can be inferred data, about the profile of individuals that has been created in the past., as we will see in the "profiling" definition (Article 19 Data Protection Working Party,2017). This extensive data handling allows AI systems to make decisions based on a broad understanding of individual behaviors and characteristics.

3. The Risk based Approach adopted by the AI Act

The AI Act is claimed to follow a risk based approach, one that tailors the choice and design of regulatory instruments based on the level of risk, according to the rule: "the higher the risk, the stricter the rules". By adopting this approach, the EU seeks to safeguard fundamental values and rights without unduly hampering the benefits that AI can bring to society. This means that the Regulators aim to achieve a proportionality principle in the Regulation and that the underlying objective of the AI Act's risk based approach is to succeed an optimal and proportional balance between innovation and the benefits of AI systems on the one hand, and the protection of fundamental values such as safety, health, and fundamental rights on the other

In this regard, Recital (26) AI Act points out that: «*In order to introduce a proportionate and effective set of binding rules for AI systems, a clearly defined risk based approach should be followed. That approach should tailor the type and content of such rules to the intensity and scope of the risks that AI systems can generate. It is therefore necessary to prohibit certain unacceptable AI practices, to lay down requirements for high-risk AI systems and obligations for the relevant operators, and to lay down transparency obligations for certain AI systems*."

To this end, the AI Act distinguishes four risk categories (unacceptable, high, limited, and minimal), defining regulatory requirements based on the risks posed by AI systems. While only unacceptable and high risk AI systems are explicitly named in the AI Act, additional transparency requirements are defined for certain systems posing limited risk. (Kusche, I,2024).

3.1. The Notion of Risk

The global emergence of risk based approaches to regulating AI systems raises, first and foremost, the question as to what policymakers (and the EU in particular) mean when they talk about risk.

Generally speaking, "risk" is the likelihood that a source of danger, of hazard will turn into actual harm. Based on this understanding, Art. 3 No. 2 of the AI Act defines risk as "*the combination of the probability of an occurrence of harm and the severity of that harm*». This means, that the amount risk equals the severity multiplied to the likelihood of a certain harm. Recital 5 of the Act Mentions that: "*Such harm might be material or immaterial, including physical, psychological, societal or economic harm*."

One of the most pressing issues for any risk based approach is how to reach consensus on which risks to select and how serious they are considered to be, either in terms of probability or impact, or both. The AI Act tried to define the criteria for determining when AI poses unacceptable risks to society and individuals and to determine when an AI system should be banned in the sense of a human rights centric approach and rationale. The same with AI systems that should be classified as "high-risk" and therefore allowed on the market, only if certain safeguards are put in place. (Malgieri & Santos,2024)

Limited risk AI systems need to comply with transparency obligations defined in the regulation, while AI systems whose risk is deemed **minimal** can follow a voluntary code of conduct defined in the Act.

3.2. Risk-Based Classification of AI systems

The AI Act classifies AI risk into four levels, namely 1) unacceptable risk, 2) high risk, a 3) limited risk and 4) minimal risk.

1) **Unacceptable risk**

AI systems considered to pose an unacceptable risk, are strictly prohibited due to their potential to violate fundamental rights, compromise safety, or harm societal well-being. These include systems that deploy subliminal techniques to manipulate individuals against their will, exploit vulnerabilities of specific groups, such as children or persons with disabilities, or enable harmful practices like social scoring by public authorities. Additionally, the use of real-time remote biometric identification systems, such as facial recognition in public spaces for law enforcement, is generally banned unless under narrowly defined exceptions, such as counterterrorism or serious criminal investigations, and even then, requires judicial or regulatory authorization. All AI systems considered a clear threat to the safety, livelihoods and rights of people are banned, from social scoring by governments to toys using voice assistance that encourages dangerous behavior.

2)**High risk**

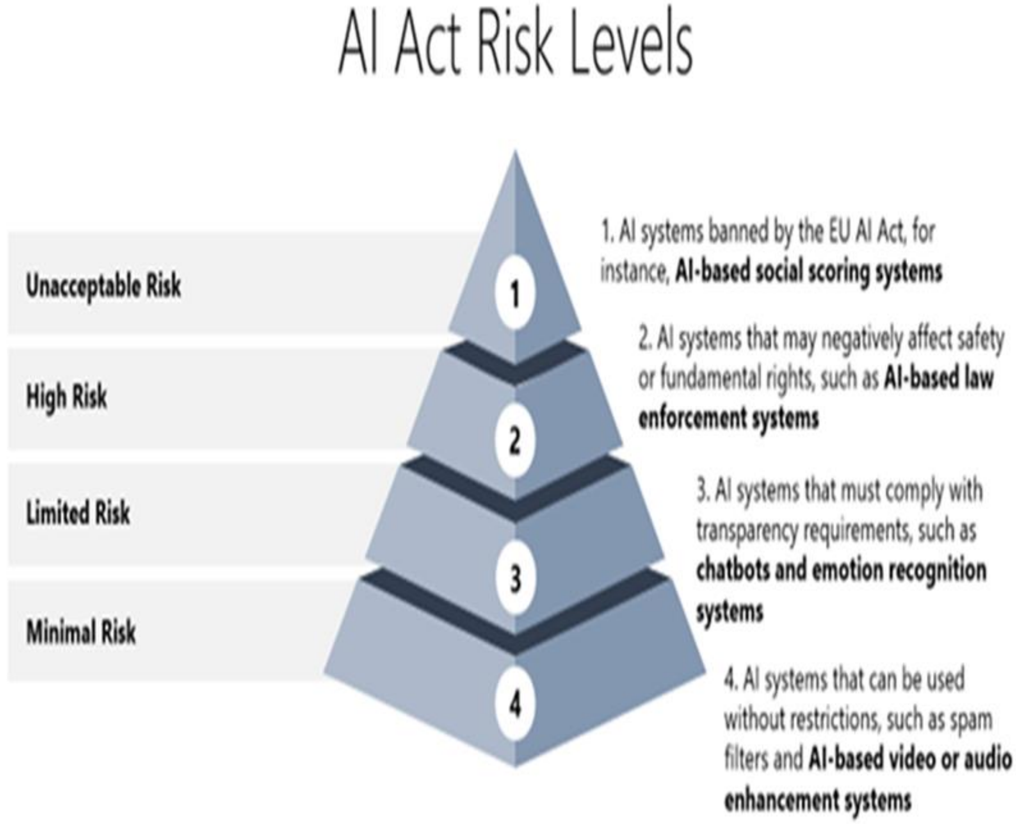
Under the EU AI Act, high-risk AI systems are those deemed to pose significant risks to fundamental rights, safety, or societal well-being, particularly in sensitive domains such as law enforcement, migration, border control, education, employment, and healthcare. These systems are subject to stringent regulatory requirements to ensure their ethical and safe deployment. Deployers must conduct comprehensive risk assessments, such as the Fundamental Rights Impact Assessment (FRIA), to evaluate potential harm to individuals or groups and implement mitigation measures.

3)Limited risk

Limited risk refers to the risks associated with lack of transparency in AI usage. The AI Act introduces specific transparency obligations to ensure that humans are informed when necessary, fostering trust. For instance, when using AI systems such as chatbots, humans should be made aware that they are interacting with a machine so they can take an informed decision to continue or step back. Providers also have to ensure that AI generated content is identifiable. Besides, AI generated text published with the purpose to inform the public on matters of public interest must be labelled as artificially generated. This also applies to audio and video content constituting deep fakes.

4)Minimal or no risk

The AI Act allows the free use of minimal risk AI. This includes applications such as AI enabled video games or spam filters. The vast majority of AI systems currently used in the EU fall into this category.



Source :European Commission

4.Part A’: Prohibited AI Practices and Human Rights Protection Under the EU AI Act

In this part, we will analyze which are the AI systems of unacceptable risk that are banned from the market, which are their characteristics and which fundamental rights are being threatened by them.

4.1. Prohibited AI practices

Prohibited AI practices, are AI applications considered as of unacceptable risk, due to their potential to cause severe harm and violate fundamental rights. They are strictly forbidden under the Regulation, with two exceptions, for specific reasons that we will see.

More specifically, Art. 5(1)(a) of the Act declares that an” *AI system that deploys* ***subliminal techniques beyond a person’s consciousness or purposefully manipulative or deceptive techniques,*** *with the objective, or the effect of materially distorting the behavior of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm.»,* is prohibited.

In particular, Recital 29, expands more by proclaiming that “*in particular (an AI system) having sufficiently important adverse impacts on physical, psychological health or financial interests are likely to occur, are particularly dangerous and should therefore be prohibited*.”

Art. 5(1)(b) bans an AI system that “***exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective,*** *or the effect, of materially distorting the behavior of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm*;”

Art. 5(1)(c) of the Act prohibits *“****systems for the evaluation or classification of natural persons or groups of persons*** *over a certain period of time* ***based on their social behavior or*** *known, inferred or predicted personal or personality characteristics,* ***<u>with the social score</u>*** *leading to either or both of the following:*

(i)*detrimental or unfavorable treatment of certain natural persons or groups of persons in social contexts that are unrelated to the contexts in which the data was originally generated or collected;*

(ii)detrimental or unfavorable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behavior or its gravity;"

Art. 5(1)(d) prohibits "*the use of an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics; this prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity*;"

The prohibition of article 5(1)(d),holds an exception though, which is defined in recital Recital 42 as follows: *«that prohibition does not refer to or touch upon risk analytics that are not based on the profiling of individuals or on the personality traits and characteristics of individuals, such as AI systems using risk analytics to assess the likelihood of financial fraud by undertakings on the basis of suspicious transactions or risk analytic tools to predict the likelihood of the localization of narcotics or illicit goods by customs authorities, for example on the basis of known trafficking routes*."

This exception, is in accordance to the proportionality principle of the Regulation's risk based approach that wishes to support innovation and technology development, wishes to "obtain the greatest benefits from AI systems" while at the same time ensuring fundamental rights.(Almada &Petit 2023)

Art. 5(1)(e) &(f) of the Act ban systems "***<u>that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage</u>****;"* and that "**<u>infer</u>** ***<u>emotions of a natural person in the areas of workplace and education institutions,</u>*** *except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons*;"

Art. 5(1)(g) AIA of the Act prohibits "*the use of (AI) biometric categorization systems that categorize individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation; this prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorizing of biometric data in the area of law enforcement;*"

Still the following, Art.5(1)(h) of the Act poses an exception to these AI biometric categorization systems and illustrates specific examples where they can be lawfully used.

It specifically indicates that "*the use of 'realtime' remote biometric identification* systems *in publicly accessible spaces for the purposes of law enforcement**, unless and in so far as such use is strictly necessary for one of the following objectives:** (i) the targeted search for specific victims of abduction, trafficking in human beings or sexual exploitation of human beings, as well as the search for missing persons; (ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack; (iii) the localization or identification of a person suspected of having committed a criminal offence, for the purpose of conducting a criminal investigation or prosecution or executing a criminal penalty for offences……"*

This means that the prohibition does not apply in cases, where these systems are used for very specific purposes, where criminal offences are involved and the use of these systems can help in protecting the public order and safety. Cases, such as searching for victims of human trafficking or sexual exploitation, or for the prevention of terrorist attacks. This is again, an exception article, that clearly demonstrates the proportionality and the balance that the EU AI Act wants to achieve in regulating AI technology. It confirms that when an AI system is beneficial for society and humanity, then there is no need to prohibit or restrict tithe human rights, the values and the rule of law is the crucial issue always in this Regulation and the exceptions make the rules.(Almada &Petit,2023)

4.1.1. Prohibited AI practices and human rights risk analysis

As cited in Recital (28) of the Act, "*Aside from the many beneficial uses of AI, it can also be misused and provide novel and powerful tools for manipulative, exploitative and social control practices. Such practices are particularly harmful and abusive and should* ***be prohibited because they contradict Union values of respect for human dignity, freedom, equality, democracy and***

the rule of law and fundamental rights enshrined in the Charter, including the right to nondiscrimination, to data protection and to privacy and the rights of the child. «Generally speaking, the articles above are clearly intended to set the boundaries for any possible misuse of an AI system that will endanger any value, any right, any fundament of the European human centric culture.

The list of prohibited AI practices stated in Article 5 of the AI Act, covers four distinct sets of AI that clearly aim to protect and fortify our fundamental human rights, by completely prohibiting certain AI techniques.

The first one, is subliminal AI techniques, which are AI enabled manipulative techniques which "can be used to persuade persons in unwanted behaviors" and have '*a significant potential to manipulate persons through subliminal techniques beyond their consciousness'*, such as personalized psychographic microtageting that we already saw. The second set, are practices that '*exploit vulnerabilities of specific vulnerable groups, such as children or persons with disabilities'*. Here, the focus is on the rights of the children and the vulnerable individuals, whose vulnerability is a default element due to their age or their state of health. For both these subliminal and exploitative AI practices, it is required that they are used '*in order to materially distort*' a person's behavior '*in a manner that causes or is likely to cause that person or another person physical or psychological harm(Recital 29)*'.The third set, involves the so-called 'social scoring systems', that we already saw infringe in a vast way our fundamental human rights, and which are systems used by '*public or private actors" and that "may lead to discriminatory outcomes and the exclusion of certain groups"(Recital 31) '*.We already analyzed, the Schufa and the Syri case, both cases of AI social scoring. Finally, the fourth category consists of biometric AI systems that involve the 'use of "*real-time «remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement'* (Recital 32). We have mentioned how these AI tools can turn societies in mass surveillance stated, violating most human rights, democracy and societies as a whole.

In the Preamble of the AI Act (28,29,30,31,32 &33), explains the context and reasons for the existence of this category and gives us a clear rationale, connecting the four sets with human rights and their protection

4.1.2. Subliminal AI techniques

The first category of prohibited AI practices involves the so-called subliminal AI practices and techniques which are used for the manipulation of people, either consumers, voters or just citizens. Subliminal techniques are characterized by their 'hidden' element (such as an audio, a video an image stimuli), that is "***beyond human perception***" and which is used as a manipulative tool of influence, without a person being aware or having a clue about its existence. These techniques, clearly violate a person's freedom of choice, freedom of thought, freedom of conscience and autonomy. The use of subliminal messages has been used, mainly in the context of nudging, which has been defined as '***any aspect of the choice architecture that alters people's behavior, without explicitly and directly forbidding any other options***'. It has been confirmed, that subconscious communication, that takes the form of subliminal messages, is often far more effective than seeking to directly address, provoke and cause the desired, economic in most cases, effect. The reason is, that the mind cannot oppose and resist to subliminal messages, since it does not know about the existence of them and, thus, does not recognize them. These findings about subliminal manipulation raise far more serious questions, since neuroscientific findings have demonstrated, that a great portion of cognitive functions is carried out unconsciously rather than consciously. This could mean that subliminal forms of manipulation can distort a larger part of our cognitive functions, that determine our thoughts, choices and actions, without our consent or even our conscious awareness.(Bulgakova,2023)

Synonymous with subliminal AI techniques, are the so-called '**dark patterns**». These are defined, as deceptive elements, that are intentionally designed to make the users do actions that they wouldn't do otherwise. For example, dark patterns allow for the manipulation of a user, by using the aforementioned subliminal stimuli, to purchase more or even unwanted products (unethical marketing), thus violating the Article 38 of the Charter, concerning consumer protection.(Neuwirth,2023,Bulgakova,2024) Other common examples of subliminal techniques, include prechecked consent boxes, confusing opt out mechanisms, and deceptive interface designs that obscure critical information and thus distort any rational decision-making process. In summary, subliminal AI techniques fall into the category of the manipulation of the human mind and behavior and poses serious risks for individuals and societies as a whole as they '*can cause direct harm, violate the target's autonomy and treat persons as things*'

4.1.3.AI Practices exploiting vulnerabilities

The second category of prohibited AI practices concerns those that exploit, "*the vulnerabilities of a person or a specific group of persons, due to their age, disability or a specific* *social or economic situation, that is likely to make those persons more vulnerable to exploitation, such as persons living in extreme poverty, ethnic or religious minorities".*

These AI systems can be used "with *the objective to or the effect of materially distorting the behavior of a person in a manner that causes or is likely to cause significant harm to that or another person…*."

The principal motive behind this provision is '***vulnerability***', illustrated by the examples of specific vulnerable groups, such as children, persons with disabilities, low income population and minorities, the most vulnerable and in need of protection part of a civilized society. As in the case of subliminal AI, the combination of different technologies and big data that makes it easier to detect and exploit vulnerabilities (Malgieri&Rebrean,2024). For instance, data marketing practices have already been identified that combine and rely technologies and apps, that track movements, locations, search engines history and inquiries, use of shopping apps, health profile apps, medical data tracking, in order to bombard individuals with target advertising, personal recommendations, political advertising and other services and products. These technologies make it instantly possible to identify and exploit individual vulnerabilities and there are numerous instances of discriminatory AI systems that have already been found to pose serious threats to individuals, groups and society as a whole based on the exploitation of vulnerabilities. For instance, there are numerous apps and games that have been identified that 'spy on kids' and collect data and personal information about their location or their vulnerability status(runaways, abusive family conditions, abusive environments etc..Other AI powered applications can "nudge" children in sharing, for example sexually explicit content, influence their behavior in certain way, for example overspend on videogames purchases or fall victims of online fraud, exploit them and distort their mental status in such highly sophisticated ways manipulative ways, that pose numerous risks for their own lives, let alone their mental health(E. Lievens et al., 2019).Similarly, other AI systems, for example recently dating apps,deceive elder lonely people in order embezzle money from them.(Neiwirth,2023,Bulgakova,2024)

All in all, AI practices that exploit specifically vulnerabilities, are cases of multiviolations of the fundamental human rights enshrined in the Charter, such as the rights of the child, the the rights of the elderly and of course the right to dignity.

4.1.4. Social scoring systems

The third category of prohibited AI practices, concerns the '**social scoring systems'**. On a general level, personal characteristics are collected from public and private actors through social networks or official registers and documents and are widely used to rate people in fields, ranging from hiring someone for a job, approve a loan, attribute social security benefits, health insurance compensations etc. The key problem is that such scoring systems "*may lead to discriminatory outcomes and the exclusion of certain groups*"(Recital 31) ,thus violating a plethora of rights, disposed in the Charter, such as the right to dignity(Article 1) ,to protection of personal data(Article 8),to engage to work and choose an occupation(Article 15),to education(Article 14),to health care(Article 35),to social security and assistance(Article 34) and of course to equality(Article 20) and nondiscrimination(Article 21).

Another problematic issue for highlighting is the fact that the datasets used in the social scoring systems, are more than often *inaccurate, since they are more often than not collected through social networks, biased* and discriminatory towards gender, race, ethnicity, low income, location of residence and lifestyle. In order to to prevent these detrimental or unfavorable effects of social scoring systems, the AI Act ,rightfully so, prohibits them(Neuwirth,2023).Lastly,another important point, is that all this "social scoring" data, especially those collected from public bodies, are completely "*unrelated to the context in which the data was originally generated or collected*" and thus don't have the consent of the data subjects to be used in AI systems that serve a completely irrelevant purpose, that might even harm them and cause life changing problems. (Kusche, I.,2024).

4.1.5. Biometric identification systems

The fourth and last category of prohibited AI practices relates to so-called **'"real-time" remote biometric identification systems «*which concerns "natural persons in publicly accessible spaces for the purpose of law enforcement*"**. Biometry generally refers to different methods of identifying people, using physiological features. Probably the best known system that uses biometric data is facial recognition technology. However, the AI Act Article 5(e) specifically states that "***the use of AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage***" is a prohibited AI practice that violated multiple human rights, such as the right to privacy and the freedom to assembly.

However, there exist many more biometric systems, including systems that use other features, such as fingerprints, irises, retinas, hand geometry, voices, signatures and palm print features, that are being used in order to identify an individual, to track his movements, his whole life actually. There are also various forms of digital manipulation of faces, such as "deep fake", that can be used to mislead identification systems and which, consequently, may lead to false decisions. The risks to human rights, by this kind of AI biometric systems, include breach of the right to privacy, restriction of the freedom to assembly, violation of the freedom to expression and a general feeling of mass surveillance by the Big Brother of the AI technology.(Leslie et al,2021)

The AI Act addresses these concerns and tries to mitigates the possible risks by qualifying certain biometric systems, namely the use of "real-time «remote biometric identification systems, in publicly accessible spaces, for the purpose of law enforcement' by categorizing them as prohibited AI practices. Recital 32, elaborates more by claiming that the use of these AI systems," is *particularly intrusive to the rights and freedoms of the concerned persons, to the extent that they may affect the private life of a large part of the population…indirectly dissuade the exercise of the freedom of assembly and other fundamental rights*"

The term '"real-time" remote biometric identification system' is defined by the AI Act, as '*a remote biometric identification system whereby the capturing of biometric data, the comparison and the identification all occur without a significant delay*'. The same definition clarifies that 'real time' not only includes *'instant identification, but also limited short delays in order to avoid circumvention'*. Thus, the keywords for the prohibition of these biometric identification systems are that they are operated 'remotely', in 'real time' (including short delays) and in 'publicly accessible places' as well as 'for the purpose of law enforcement'.

At the same time, the provision also features three broad exceptions to the prohibition, namely when the said system is used for: (i) the targeted search for specific potential victims of crime, including missing children; (ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack; (iii) the detection, localization, identification or prosecution of a perpetrator or suspect of a criminal offence.

In sum, we can say that, the use of AI biometric surveillance systems raises serious concerns, as to fundamental rights. Remote biometric recognition is linked to breaches of the right to privacy, people's autonomy, their right to establish details of their identity and psychological integrity. It also negatively impacts freedom of expression, association and freedom of movement. Furthermore, remote biometric identification and predictive tools may lead to discrimination, violate the values of equality and justice, due to biased data sets. (Pehlivan,2024)

4.1.6.AI systems inferring emotions via biometric identification systems

Under Article 5(1)(f), "*the use of AI systems to infer emotions of a natural person in the areas of workplace and education institutions*;" is a prohibited AI practice. Now, the regulators felt the need to last these AI systems since it became obvious that, these techniques were used in work and educational environments, using biometric data, in order to "spy" actually on how a worker feels, if he likes his job or not, how he is going to react, what type of person he is, an introvert or a fitforpurpose good seller extrovert? Is the student lazy is he procrastinating? Is he probably cheating, because he gazed for a few seconds outside the window? For this reason, as stated in Recital (46),"*AI systems identifying or inferring emotions or intentions of natural persons on the basis of their biometric data may lead to discriminatory outcomes and can be intrusive to the rights and freedoms of the concerned persons. Considering the imbalance of power in the context of work or education, combined with the intrusive nature of these systems, such systems could lead to detrimental or unfavorable treatment of certain natural persons or whole groups thereof. Therefore, the placing on the market, the putting into service, or the use of AI systems intended to be used to detect the emotional state of individuals in situations related to the workplace and education should be prohibited*."

These systems,again,using biometric data, apart from the usual violations of privacy and human dignity, pose risks to other fundamental rights too, such as the right to education of Article 14 of the Charter, the right to engage to work of Article 15,the right to fair and just working conditions of Article 31 and of course the right to human dignity, since no one, has the right to spy on anyone, using his facial expressions and generally his expression of emotions ,which among others, vary considerably among different cultures, just for the mere purpose to track their mind, their intentions, their thoughts, their life.

Part B': High-Risk AI Practices and Human Rights Protection Under the EU AI Act

5.1. High-risk AI systems

High-risk AI Practices concern AI systems that pose significant risks to fundamental human rights**, <u>but</u>** can be used legally, if they meet strict regulatory requirements to mitigate risks. They are, as we mentioned, characterized by a high potential for posing a threat to fundamental human rights, but they are not deemed as unacceptable and therefore they are not in need for prohibition, but they are obliged to fulfill certain compliance requirements.

As stated in Recital (46), "*High-risk AI systems should only be placed on the Union market, put into service or <u>used if they comply with certain mandatory requirements</u>. Those requirements should ensure that high-risk AI systems available in the Union or whose output is otherwise used in the Union do not pose unacceptable risks to important Union public interests as recognized and protected by Union law*".

Paragraph (48) of the Recital elaborates more on the issue of the connection between high risk AI systems and the fundamental rights by declaring: "*The extent of the adverse impact caused by the* ***<u>AI system on the fundamental rights protected by the Charter is of particular relevance when classifying an AI system as high risk.</u>*** *Those rights include the right to human dignity, respect for private and family life, protection of personal data, freedom of expression and information, freedom of assembly and of association, the right to nondiscrimination, the right to education, consumer protection, workers' rights, the rights of persons with disabilities, gender equality, intellectual property rights, the right to an effective remedy and to a fair trial, the right of defense and the presumption of innocence, and the right to good administration. In addition to those rights, it is important to highlight the fact that children have specific rights as enshrined in Article 24 of the Charter and in the United Nations Convention on the Rights of the Child, further developed in the UNCRC General Comment No 25 as regards the digital environment, both of which require consideration of the children's vulnerabilities and provision of such protection and care as necessary for their wellbeing. The fundamental right to a high level of environmental protection enshrined in the Charter and implemented in Union policies should also be considered when assessing the severity of the harm that an AI system can cause, including in relation to the health and safety of persons*."

As it obvious almost all fundamental rights of the Charter are being mentioned in this paragraph and since this is the case in the next part we will present them by using specific "use cases" that demonstrate how certain AI systems threatened or violated fundamental human rights. Some "use cases" are cases of multiviolations of different human rights and we will present them last, since they are considered important. All these AI systems are listed in Annex III of the Regulation, as seen in the ANNEX III of the Regulation below.

ANNEX III

High-risk AI systems referred to in Article 6(2)

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

. Biometrics, in so far as their use is permitted under relevant Union or national law:

a) remote biometric identification systems.

This shall not include AI systems intended to be used for biometric verification the sole purpose of which is to confirm that a specific natural person is the person he or she claims to be;

b) AI systems intended to be used for biometric categorization, according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics;

c) AI systems intended to be used for emotion recognition.

. Critical infrastructure: AI systems intended to be used as safety components in the management and operation of critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity.

. Educational and vocational training:

a) AI systems intended to be used to determine access or admission or to assign natural persons to educational and vocational training institutions at all levels;

b) AI systems intended to be used to evaluate learning outcomes, including when those outcomes are used to steer the learning process of natural persons in educational and vocational training institutions at all levels;

c) AI systems intended to be used for the purpose of assessing the appropriate level of education that an individual will receive or will be able to access, in the context of or within educational and vocational training institutions at all levels;

d) AI systems intended to be used for monitoring and detecting prohibited behavior of students during tests in the context of or within educational and vocational training institutions at all levels.

Employment, workers' management and access to self-employment:

.

a) AI systems intended to be used for the recruitment or selection of natural persons, in particular to place targeted job advertisements, to analyze and filter job applications, and to evaluate candidates;

b) AI systems intended to be used to make decisions affecting terms of work-related relationships, the promotion or termination of work-related contractual relationships, to allocate tasks based on individual behavior or personal traits or characteristics or to monitor and evaluate the performance and behavior of persons in such relationships.

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

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

b) AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems used for the purpose of detecting financial fraud;

c) AI systems intended to be used for risk assessment and pricing in relation to natural persons in the case of life and health insurance;

d) AI systems intended to evaluate and classify emergency calls by natural persons or to be used to dispatch, or to establish priority in the dispatching of, emergency first response services, including by police, firefighters and medical aid, as well as of emergency healthcare patient triage systems.

Law enforcement, in so far as their use is permitted under relevant Union or national .law:

a) AI systems intended to be used by or on behalf of law enforcement authorities, or by Union institutions, bodies, offices or agencies in support of law enforcement authorities or on their behalf to assess the risk of a natural person becoming the victim of criminal offences;

b) AI systems intended to be used by or on behalf of law enforcement authorities or by Union institutions, bodies, offices or agencies in support of law enforcement authorities as polygraphs or similar tools;

c) AI systems intended to be used by or on behalf of law enforcement authorities, or by Union institutions, bodies, offices or agencies, in support of law enforcement authorities to evaluate the reliability of evidence in the course of the investigation or prosecution of criminal offences;

d) AI systems intended to be used by law enforcement authorities or on their behalf or by Union institutions, bodies, offices or agencies in support of law enforcement authorities for assessing the risk of a natural person offending or reoffending not solely on the basis of the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680, or to assess personality traits and characteristics or past criminal behavior of natural persons or groups;

e) AI systems intended to be used by or on behalf of law enforcement authorities or by Union institutions, bodies, offices or agencies in support of law enforcement authorities for the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of the detection, investigation or prosecution of criminal offences.

Migration, asylum and border control management, in so far as their use is permitted .under relevant Union or national law:

a) AI systems intended to be used by or on behalf of competent public authorities or by Union institutions, bodies, offices or agencies as polygraphs or similar tools;

b) AI systems intended to be used by or on behalf of competent public authorities or by Union institutions, bodies, offices or agencies to assess a risk, including a security risk, a risk of irregular migration, or a health risk, posed by a natural person who intends to enter or who has entered into the territory of a Member State;

c) AI systems intended to be used by or on behalf of competent public authorities or by Union institutions, bodies, offices or agencies to assist competent public authorities for the examination of applications for asylum, visa or residence permits and for associated complaints with regard to the eligibility of the natural persons applying for a status, including related assessments of the reliability of evidence;

d) AI systems intended to be used by or on behalf of competent public authorities, or by Union institutions, bodies, offices or agencies, in the context of migration, asylum or border control management, for the purpose of detecting, recognizing or identifying natural persons, with the exception of the verification of travel documents.

Administration of justice and democratic processes:

.

a) AI systems intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts, or to be used in a similar way in alternative dispute resolution;

b) AI systems intended to be used for influencing the outcome of an election or referendum or the voting behavior of natural persons in the exercise of their vote in elections or referenda. This does not include AI systems to the output of which natural persons are not directly exposed, such as tools used to organize, optimize or structure political campaigns from an administrative or logistical point of view.

"Source: Official Journal 12.7.2024"

5.2. Key Characteristics of High-Risk AI systems that can challenge fundamental human rights

It is of paramount importance to comprehend the characteristics of AI systems that may potentially, infringe upon fundamental human rights in order to guarantee their responsible and ethical utilization. This section, examines the key aspects of AI systems, including opacity, automation, big data analysis, personalization, social constructs and machine learning. It also considers how these characteristics can affect human rights, which will be analyzed in more detail later in the next section.

5.2.1. Opacity:

AI tools are often described as "black boxes", because their internal workings are not transparent, even to those who design them. The "black box" problem represents a significant challenge in the field of automated decision-making, due to the opacity of complex AI systems. This phenomenon arises from the fact that machine learning AI systems are capable of processing and analyzing vast quantities of data in order to make predictions or decisions. (Pasquale,2015) However, the internal mechanisms through which they operate remain opaque to human understanding, which is why they are often referred to as "black boxes". To provide an example, a deep learning neural network may make accurate predictions regarding medical diagnoses or financial transactions. However, the complex layers of computations and transformations within the network make it challenging to identify the specific reasoning behind these conclusions. This opacity can be particularly problematic in domains where the consequences of erroneous decisions are significant, such as healthcare, finance, and criminal justice. (Edwards& Veale,2017).

The phenomenon of the "black box" introduces a number of additional critical issues. Firstly, it presents a significant challenge to accountability. In instances where an AI system's decision has significant consequences, such as the denial of a loan or the recommendation of a medical treatment, it is imperative to understand the rationale behind the decision in order to assess its fairness and accuracy. In the absence of transparency regarding the decision-making process, individuals or organizations affected by the outcome are left without an understanding of the rationale behind the decision. (Burell,2016)

Furthermore, this opacity challenges the principle of informed consent, a key element of human rights protection. If individuals cannot comprehend the implications of their consent, the legitimacy of that consent, and by extension, the protection of their human rights, is questionable.

As a natural consequence, the opacity in AI systems can result in a loss of trust. Stakeholders, including users, regulators, and the general public, may be reluctant to accept or rely on automated decisions if they are unable to comprehend or verify the processes that underpin them. This lack of trust can impede the wider adoption of AI technologies and limit their potential benefits. In essence, the black box problem demonstrates the conflict between the complexity of contemporary AI systems and the necessity for transparency and interpretability (Grasseger & Krogerus,2019)

5.2.2. Automation

The term "automation" is used to describe the capacity of AI systems to perform tasks with minimal or no human involvement. This process enables computers to replace humans in a variety of scenarios, performing tasks autonomously and even establishing goals that were not initially defined by their designers. The autonomy of AI systems enables them to make decisions and execute tasks based on complex rules and a vast amount of data, allowing them to operate with a high degree of autonomy. Nevertheless, this autonomy gives rise to concerns regarding the system's behavior and its implications when minimal or absent human oversight is present. Moreover, another considerable challenge associated with automation is the lack of transparency in the decision-making processes involved. The internal workings of AI systems are frequently opaque to users and regulators alike. This lack of transparency may impede efforts to identify and address potential biases or errors in decision-making processes. To take one example, if an automated credit scoring system denies a loan, it can prove difficult for the individual concerned to understand the rationale behind the decision. The introduction of automation gives rise to concerns regarding unpredictability. The intrinsic complexity of automated systems, coupled with their capacity to process vast quantities of data, renders their outputs difficult to anticipate. Such unpredictability can prove highly problematic, particularly when automated systems have an impact on critical areas such as healthcare. (AIHLEG,2019)

5.2.3. Big Data Analysis

Big data analysis, is the process of examining large and complex datasets with the objective of identifying patterns, correlations, and trends. The application of big data has become a standard practice across a range of fields, including business, healthcare, and public policy(Binns & Johnosn,2019).The intrinsic complexity associated with the vast quantity of data constitutes a fundamental attribute of big data analysis, since it challenges the accuracy and fairness of the analyses, which must be addressed in order to guarantee the reliability of the results.(Bozdag,2013).Furthermore, the vast volume of data, derived from a diverse range of sources, can impede the ability to ascertain its quality and comprehend the context of the findings. Consequently, the analysis may yield results that are challenging to interpret or act upon. Furthermore, the AI systems utilized in the context of big data analysis have the potential to perpetuate or amplify existing biases. In the event that the data employed is biased or unrepresentative, the resulting analysis may yield inaccurate or unfair results. Furthermore, the lack of justification represents a significant issue. A considerable proportion of big data AI systems function in a manner that can be characterized as opaque, providing only minimal or no insight into the rationale behind their conclusions. This absence of justification presents a significant challenge for individuals seeking to comprehend or contest the results. By way of example, if a large data set identifies specific demographic groups as high-risk for specific behaviors, it is not uncommon for the rationale behind these categorizations to be unclear. This lack of transparency can impede efforts to address potential errors or biases.

<u>5.2.4. Profiling</u>

Profiling, is an automated statistical analysis technique that employs deductive logic to predict human behavior and characteristics of individuals or groups, based on data from multiple sources. This data driven process, categorizes individuals according to statistically similar behavioral patterns and characteristics observed in others, thereby allowing for predictions about an individual's identity, interests, and various personal aspects(Hildebrandt,2014). The essence of profiling lies in its ability to analyze and infer information about a person, including their demographic information, personal preferences, behaviors, and even more intimate details such as religious beliefs or health status(Buttarelli,2016). Since, profiling generates new information and insights from combined datasets, even when individual data points are seemingly no sensitive or irrelevant, the combination of several such data points can lead to the creation of highly sensitive personal data, with a high level of accuracy. Through the collection of data, statistical analysis to identify correlations, and the application of these correlations to individuals, profiling generates new, previously unknown information about a person. This capacity to infer new information from existing data sets is what makes profiling so powerful, yet also potentially problematic. (Algieri & Comande,2017)

Profiling, is integral to AI systems, that we have seen before, especially in scenarios where results or negative outcomes, are entirely or partly based on the results of profiling. For instance, profiling can be used to determine whether an individual has the skills necessary for a particular job, their eligibility for a loan, insurance, or healthcare services, or even their suitability for a visa, refugee status, or university admission. The impact of profiling on these critical, life changing decisions is truly enormous

As it is obvious AI systems and profiling are connected and linked processes, and while AI systems generate decisions, results, outcomes, it is the profiling technique provides the detailed analysis needed for these decisions.

<u>5.2.5. Personalization Capacity</u>

The term 'personalization capacity', is defined as the process of adapting content, recommendations, or services to align with individual preferences based on data analysis(Hildebrandt,2014). Personalization AI systems, frequently make use of self learning mechanisms in order to predict future behaviors or preferences and enhance their performance over time. By analyzing user data and feedback, these AI systems are able to refine their models, thereby improving the accuracy of their predictions or recommendations (Goodman & Flaxman,2016) To provide an example, the creation of echo chambers through the use of personalized search results or news feeds has the potential to manipulate users' political views and decisions, without their explicit awareness. This is due to the fact that such platforms have the potential to reinforce existing beliefs and limit exposure to diverse perspectives. (Tene & Polonetsky,2012)

5.2.6. Social Constructs Around AI systems algorithms

The term **'social construct'** is used to describe the values, norms, and assumptions that are embedded in AI systems. These constructs influence the functioning of AI systems and the outcomes they produce, by reflecting societal biases and values. It is not uncommon for AI systems to reflect the biases and values of their designers or the data on which they are trained. This can result in decision-making processes that serve to reinforce existing societal biases, inequalities or racism. For example, facial recognition AI systems have been the subject of criticism due to their racial bias, which is largely attributed to the lack of diverse training data that fails to adequately represent all demographic groups.

5.3. High Risk AI systems and their Impact on Human Rights

AI systems, seem to affect specific fundamental human rights and pose great risks to individuals and the society in general. Such risks can include discrimination, inequality and stereotyping. These negative effects, affect greatly most of the fundamental human rights, for example the right to a fair trial, the right to privacy, the right of freedom of expression and many others.

In this part, we will try to examine the negative impacts of several AI systems mentioned in Annex III of the Regulation, on individuals and collectively at societal level, by presenting specific examples. At the same time, we will try to link these impacts, with the violation of specific human rights articulated by the European Convention of Human Rights (“ECHR”) and the Charter of Fundamental Rights of the European Union (hereafter “The Charter”). We have to point out, that many negative impacts of AI Systemic systems can potentially violate more than one human right at the same time, and thus their consequences overlap. For this reason, the last part of the charter contains high-profile cases that the misuse of AI systems, abused and violated several fundamental human rights.

5.3.1 Civil rights -The right to a Fair trial and the "due process" clause

AI systems, can violate the right to a fair trial, according to Article 47 of the Charter and article 6 of the ECHR and the due process clause, which includes the presumption of innocence, the right to be informed about the accusation, the right to a fair hearing and the right to defend oneself in person.

5.3.1.a) AI systems used in the justice System-Annex III (8)(a)-Administration of justice

AI systems are widely used today, in the criminal justice system, for assisting and producing custodial and sentencing decisions, mostly in USA and the UK. The US criminal justice system started using, in the early 2000's, AI risk assessment tools in order to predict, how likely it is for an individual to reoffend, or help the judge determine which individual should be granted a bail or even how long a sentence should last, especially in minor cases of low risk offenders, who can be given a shorter sentence or even not being jailed at all. These kinds of risk assessments seem to help with avoiding prison overcrowding and huge workload of courts and judges, but the criteria, through which the scores are generated, are being under scrutiny for mainly racial biases. In 2016, the investigative journalism site ProPublica, published a report which demonstrated that risk assessment AI systems used for the legal process, can be inaccurate and racially biased. Risk scores of 7000 people who were arrested in Florida and assessed, using an AI tool, known as COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), have shown that the AI System which was used was biased against blacks (Angwiin, Mattu, and Kirchner, 2016). That means, that black people were twice as likely to be determined, as "high risk" for committing violent crimes in the future, then white people and at the same time, white people were significantly more often, determined as "low risk", than black people. This observation implies that there should be great caution as to what these AI systems can actually do and under which conditions they can be used, in order not to jeopardize, **the right to a fair trial and the equality before the law.**

In 2016, a defendant, named Loomis who has been given a long prison sentence in the state of Wisconsin, challenged the AI tool decision, claiming that he and his legal team were not able to study and evaluate the COMPAS AI system and that he was prevented from challenging the scientific validity and accuracy of such test that produced a great negative effect to his life. He also alleged that the system in question (COMPAS), violates **due process rights** by taking gender and race into account. The State Supreme Court in the Loomis vs. Wisconsin case, rejected the challenge, claiming that it was not necessary for the defendants to know the criteria of the risk score AI system and that the final outcome of the AI system, was enough in terms of transparency, a statement which of course is not at all right.(Larson et al.,2016) The COMPAS AI tool, apparently lacked transparency and the denial to challenge the decisions made from it, was a **direct violation of the right to a fair trial and the principle of equality of arms**.

Similar AI systems for risk assessment scores are today being used in some Constables in the UK. More specifically, like the COMPAS tool, the Harm Assessment Risk Took (HART), an algorithmic tool, was used by Durham Constabulary to help custody officers decide whether arrested individuals should be offered an opportunity to participate in its 'Checkpoint' rehabilitation program. HART used 'custody event data', given from the Constabulary's custody management IT systems and data from Durham Constabulary's pre-existing records. HART purported to calculate the arrestee's risk of committing an offence in the subsequent two years. Individuals predicted as likely to commit a 'serious' offence (defined as an offence involving violence), a non-serious offence, or no offence, were classified as 'high', 'moderate' or 'low' risk respectively. Only those receiving a 'moderate' prediction was eligible for "Checkpoint». In the case of HART, the problem was that training data was built from arrest data, but to be arrested is no marker of criminality in itself. Consequently, to be given a higher rating in the HART tool was not truly a predictor of criminality, but at best a predictor of likely arrest. Since only those who were given a moderate prediction were eligible for Checkpoint, the algorithmic decision making process is impacting individual's rights of equality before the law, under Article 20 of the Charter, their right to an effective remedy for sure, under article 47, their right to defense under article 48 and violated the principles of legality under Article 49. (Yeung & Harkens)

5.3.1.b). AI systems used in law Enforcement-Annex III (6) Law enforcement &Administration of Justice

Law enforcement agencies are increasingly using predictive policing AI systems to forecast criminal activity, control crime and efficiently allocate police resources. Predictive policing generally describes any system that analyzes available data to predict either where a crime may occur in a given time window or who will be involved in a crime as either victim or perpetrator. These AI systems are built on data which are often racially biased and highly prejudicial against specific ethnic and racial backgrounds. (Angwil et al, Propublica,2016) This happens because the existing, «historical" police databases intentionally or unintentionally are heavily biased. Biases to racial or ethnic background, may not be recognized by police data scientists, when integrated into an AI system and as a result bias may become standardized, stereotyped and are not at all easily identified. The use of these AI systems in the criminal prevention system, greatly affects the right to a fair trial and the "due process" clause, since many subsequent decision-making processes in criminal justice, including those during prosecution, pretrial services, sentencing, parole, and corrections, derive their analysis from policing data inputs, which may be inaccurate and unjust (Richardson et al, 2019, Perry,2013)).

For example, Amazon's Facial Recognition Technology and Surveillance, named "Rekognition", has been sold to law enforcement agencies, enabling real-time identification of individuals in public spaces. This technology has been criticized for its potential to facilitate mass surveillance, infringing on individuals' rights to privacy and freedom of assembly. Notably, studies have indicated that facial recognition systems can exhibit biases, leading to higher misidentification rates for people of color and women, thereby exacerbating issues of discrimination. In response to these issues, Amazon shareholders have expressed apprehension about the company's AI technologies contributing to human rights violations. Proposals have been submitted urging Amazon to assess and address the human rights impacts of its AI systems, particularly those used in surveillance and law enforcement contexts.

5.3.1.c). AI systems used in predictive counterterrorism-ANNEX III-Law enforcement

International terrorism and more specifically a string of brutal terrorist attacks in the center of European and American cities, that has shocked the international community during the last years, causing an earthquake in the fundaments of democracies around the world, has obliged national leaders to ask online social media platforms to use their AI systems, in order to detect potential terrorists and to identify users and accounts that may spread extremist content and radicalization. In this sense, AI systems are used to support the activities of the intelligence and security services. But, the problem is that the AI systems for terrorism prevention might wrongfully classify as terrorists or predict future terrorist attacks, people who are innocent (false positives) or irrelevant actions; and at the same time, classify as non-terrorists (false negatives), very suspicious or even guilty but overlooked terrorists and terrorist plans. (Goel et al,2017)

So, the serious threat to the right to a fair trial and the principle of proportionality of penalties (art. 49 of the Charter), is the fact that these “predictive counterterrorism” AI systems and “profiles” are used, as evidence in court, while the output of any predictive AI System should never constitute a standard of proof in its own right and should in no case, be considered as evidence for any measures or penalties imposed, in the judicial process. The output of any predictive AI Systemic model always represents mere possibility rather than proof and any decision on the basis of predictive AI systems, must be made by a judge who is sufficiently informed about the limits and capabilities of the model being used and who will safeguard proportionality in the process (McKendrick, 2019).

Use cases of predictive counterterrorism that only succeeded in violating human rights comes again from the US, even though China has a much more sophisticated and intrusive such AI technologies in use. In the United States "Operation Iron Bridge" and the Suspicious Activity Reporting (SAR) System, were used by the U.S. Department of Homeland Security (DHS) in order to identify and track potential terrorist threats. The issue was that SAR guidelines included very vague criteria like "acting suspiciously," which led to over reporting and profiling of innocent individuals.Even,activities such as taking photos in public places or buying certain chemicals for legitimate purposes have been flagged.Furthermore,Muslim communities and people of Middle Eastern descent have been disproportionately targeted by the systems. For example, mosques have been monitored based on algorithms that flag patterns of behavior tied to religious practices, a matter that raised concerns about discrimination, inequality and violation of the right of the freedom of religion. There was also a threat of the right to free expression, since minorities, especially the religious minorities mentioned(Muslims) were afraid of being monitored and this fear has led to selfcensorship, particularly among minority groups who feel targeted by these systems.

5.3.2. The fundamental rights to free elections and freedom of expression-ANNEX III-Democratic processes

The use of AI systems in European and global democracies have detrimentally affected political and civil fundamental rights, such as the right to free elections and the freedom of the press, media pluralism and the freedom of expression, as we will analyze below. These violations have a great negative effect on the fundaments of democracies and for this reason we will analyze them in the same section.

The right to free elections is a fundamental civil right and a unique procedure in democracy which is protected by Article 3, Protocol 1 of the ECHR which ensures "the free expression of the opinion of the people, in the choice of the legislature" (Council of Europe, 1952). However, it is also a right, greatly violated these days by AI systems and their implementation, in the elections' procedures worldwide.

Social bots, psychographic microtargeting, fake news and voting advice applications that allegedly "help" people determine their political identity; have a great negative impact on all civil rights, but mainly the right to free elections which is a fundamental value of democracy. (The New York Times,2016).

5.3.2.a) Social bots

Social bots are AI tools, which control social media accounts, that keep their nonhuman identity undisclosed and can automatically identify the activity of human users and produce content (ex. comments) or reactions (likes) and even engage in social interactions, in favor of a particular political candidate, a campaign, or a social issue protest. Bots can create and spread information to a very large scale of audience and hence they can artificially exaggerate the importance of an issue, or emphasize it, in a disproportionate way and by this practice, even mobilize citizens for dubious campaigns, protests and actions. The AI systems of sociopolitical bots can identify, frequently used keywords on posts, or words "trending" on an actual topic and produce their own posts (spam bots) or engage in conversations with users (chat bots). In this way, they can influence the public opinion by promoting or discrediting or even defame political candidates in a very convincing and truthful mode that can jeopardize and put at risk the election process, through the manipulation and the misinformation of the public opinion. (Binns & Johnson,2019)

The use of social bots in election campaigns appeared for the first time, as we will see later, in the 2016 US presidential elections and affected so significantly the political debate and discussion, that they probably produced a distorted and manipulated result. Since then, social bots along with political microtargeting, were used to a lesser extent, in many elections and referenda in Europe, for example the 2017 German Elections, the 2017 presidential elections in France, the Brexit referendum campaign and the Catalan independence referendum. Social bots are more frequently "working" in Platform X, due to its open application interface and its short, quick and direct messages and for this reasons, it is estimated that almost a quarter of all Twitter accounts are bots. Facebook has a great amount of bots fake accounts, too. (Bossi & Ferrara,2016)

5.3.2.b) Psychographic Political Micro Targeting-Annex III-Democratic processes

Freedom of elections and freedom of expression can be significantly impacted by the psychographic micro targeting of voters with targeted political advertisements, often coupled with the spread of disinformation. Apart from that, data collection for micro targeting political advertising AI tools, raises privacy concerns, since it is usually based on gender, age group, marital status, place of residence, online preference history and interests. (Griffin,2016)

Through political micro targeting, a political party can identify the individual voters which it is most likely to convince, through the development of large and sophisticated databases that contain personal data about as many voters as possible. The database essentially tracks voter habits and preferences, which are then data mined to identify issues of interest to each voter and whether that voter is more likely to identify with one party or another and vote for it. Individual voters are then put into groups on the basis of sophisticated AI system modeling. (The New York Times,2019)

Psychographic micro targeting AI systems, bring risks and violate the right to free elections. Online political micro targeting tools, involve creating messages targeted at narrow categories of voters, based on personal data analysis gathered from individuals' demographic characteristics and consumer and lifestyle habits' and can take the form of political direct marketing in where political actors address personalized messages (by mail, phone call, text messaging), to individual voters, by applying predictive modeling techniques to voters' big data. It is a type of **behavioral advertising, namely political behavioral advertising**, that involves tracking people's online behavior to use the collected information in order to display individually targeted advertisements and to identify voters who are likely to vote for a specific party and therefore can be targeted with mobilizing messages. Microtargeting also enables a political party to select policy proposals that match the interests of the targeted voter, for instance family aid for families, or student benefits for students. This means that, a political party can match its message to the specific interests and vulnerabilities of these voters and can address issues which are most important to these individuals, adapting the message and language to meet the individual's needs and interests, for maximum manipulative effect. (Smith,2024)

The threats for the citizens' rights, from psychographic political microtargeting, are that their privacy could be invaded, and of course they could be manipulated. Microtargeting threatens privacy because it involves gathering and combining personal data about people on a massive scale, in order to infer their political preferences. (Wagner,2020, Wagner,2021)

On the other hand, people who suspect that their behavior is monitored may adapt and appropriate their behavior, trying to escape political malicious targeting and this is a serious threat to their freedom of thought and expression (Article 10 and 11 of the Charter). If people know or suspect that their website visits are being tracked, they may feel uncomfortable visiting certain websites, or express openly their political opinion, especially in authoritative regimes. If, for example, people expect that an extreme rightwing party will win the elections, they might hesitate to visit websites about Islam or LGBT communities.

Another risk concerns the manipulation of the voters' opinion, by politicians. For instance, a party could target particular voters with tailored information that maximizes, or minimizes, voters' selfidentification. A party could use social media to expose xenophobic voters to information about the high crime rates amongst immigrants. Gorton (2016) warns that micro targeting 'turns citizens into objects of manipulation and undermines the public sphere by thwarting public deliberation, aggravating political polarization, and facilitating the spread of misinformation'. In sum, microtargeting could be used to manipulate voters and greatly distort the elections results. (Borgesius et al., 2018; Gorton, W., 2016)

5.3.2.c). AI Systemic Voting Advice AppsAnnex III-Democratic processes

Voting advice applications are AI systems that determine which political party or candidate, corresponds better to the political opinions of voters. They can classify them, through AI Systemic calculations to conservatives, liberals, leftists, centerleft, centerright, farright, anarchists etc. The use of AI voting recommender systems is being widely used and in some countries the percentage of their use are rather high (e. g. The Netherlands). It is natural that if the recommending system is not set or monitored by an independent body, the recommender AI system can be biased and could favor a specific political party, not to mention that many times these systems make inferences about someone's political beliefs from sensitive data, such as the religion beliefs and the sexual orientation of their users, which are being asked in these kind of surveys. As a result, these kinds of apps can violate the right to free elections and the freedom to vote. (Griffin,2016)

5.3.3. Freedom of expression, freedom of thought and information

A fundamental human right that is directly affected by the negative implementation of AI systems is the right to freedom of expression and information, which is enshrined in Article 10 and 11 of the Charter and 10 of the ECHR.

5.3.3.a) Filter bubbles and Echo chambers-Annex III-Democratic processes

Users who search information online have to do it through powerful search engines, like the Google search engine or through social networks, like Facebook. All of these platforms, use personalization AI systems to ensure that people see content similar to what they previously liked or commented on and this has created the phenomenon of the “filter bubble” where the “personal ecosystem of information has been catered by these AI systems". An internet user's past browsing and search history is built up gradually, when they indicate interest in topics either by clicking links, or viewing friends, or reading news stories, or listening to specific channels on You tube. Therefore, we most likely see content that reaffirms and strengthens our preexisting views and more unfortunately, that we don’t have the choice to criticize or think on anything new, that might challenge our thoughts and broaden our horizons. (Borgesius et al.,2016), and this is particularly chilling when it comes with our critical thinking and our role as independent and conscious citizens.

AI systems that tailor and personalize, especially the political information and agenda, that users can see, based on their assumed preferences and interests, can challenge the freedom of expression and the right to information, since not only do they narrow their information resources, but they also change the way people access and receive information, how they form their opinions and finally leads to an AI driven indirect oppression, if not radicalization. This phenomenon is called “echo chamber” or “filter bubble” and refers to the state of intellectual isolation and elimination of critical thinking, when for example, an AI System selectively guesses what information a voter would like to see based on information about the voter, such as location, past click behavior and search history. As a result, such a process of “sociopolitical sorting” that exposes them to similar content, can strengthen and amplify their existing views and choices and create a homogeneity of ideas and thoughts in groups of people who live in their own political or ideological “bubbles” and who only interact with likeminded people and don’t mix or share other conflicting points of view, that doubtfully promote the civic, democratic discourse. (United Nations,2018)

5.3.4. Freedom of assembly and Association-Annex III-Biometrics (1)(a)

The fundamental rights to freedom of association and peaceful assembly are protected in article 12 of the Charter and 11 of the ECHR.

The rights to freedom to assembly and association has been largely enhanced and promoted by social media platforms. Groups of people that were ostracized for centuries, gained in a decade more than they ever did. Social problems like sexual harassment, gender based violence, oppressive regimes and injustice, were confronted by users of the social media platforms. Social media around the globe serve both as a means to facilitate the exercise of the rights of association, when for instance they help organizers who seek to mobilize a large group of people in an effective manner and almost at no cost and also serve as virtual discussion and meeting places for groups of people, who may be marginalized by society and it's very difficult for them to assemble in public spaces. Moreover, social media and their AI tools, advanced dissemination, are considered as major factors for organizing major protests and huge activist movements, for example the #MeToo movement, the #BlackLivesMatter and the Euromaidan protests in Ukraine. So, as we can see the Internet is a very powerful tool of exercise of these two fundamental rights. (Bellovin et al,2018)

Yet, the use of AI systems in the AI biometric systems can and are being used for profiling, tracking and survey specific people, who can be targeted for their beliefs or rebellious actions against the regimes, namely civil society actors, human rights defenders, opposition political leaders and organizers of peaceful public assemblies. This fact has a deeply negative effect on the freedom of assembly, because it causes a form of crowd control not only online, but also offline.

Online, AI systems can be used on cyber surveillance of digital profiles of citizens' and their interactions, in order to predict their future actions and initiatives, and offline, AI systems with face recognition capacities, are being deployed at large cultural events, major sporting events, music festivals and political gathering in order to identify and survey all individuals who participate in a particular event or are present in a certain public space. These two types of surveillance, online and offline are being processed by AI systems, in order to match and identify people who may probably, try to mobilize others to defend and promote shared interests on the

Internet or in physical locations. As we understand, these types of AI systems facilitate state surveillance on every possible level and violate the right of assembly and association along with the right to privacy and data protection, especially in the cases, where specific civil actors are being targeted and almost spied upon. These forms of identification and data collection violate the individual's anonymity in public spaces and exert significant "chilling effects" on future decisions to participate in public assemblies and gatherings. The use of AI Systemic surveillance techniques for the indiscriminate and untargeted surveillance of those exercising their right to peaceful assembly and association, in both physical and digital spaces, should be prohibited.

For example, Amazon's facial recognition software, "Rekognition", has been marketed and sold to law enforcement agencies, enabling real-time identification of individuals in public spaces by integrating advanced image, video and facial analysis capabilities. It can identify people, activities, and even detect emotions or demographics from said images and videos. The major challenge with this AI tool, is that facial recognition for surveillance purposes, can lead to mass monitoring and the loss of individual privacy. If such AI tools are being deployed unchecked, they could easily contribute to a surveillance state, where the freedom to assembly, the freedom to protest, would be a relic of old times. Additionally, Rekognition has being accused of being biased. Studies have shown that facial recognition systems, including Rekognition, have higher error rates when identifying people of color, women, and younger individuals, leading to risks of discrimination. This can perpetuate systemic inequalities, especially in law enforcement contexts, where wrongful identifications could lead to unjust detentions or arrests. Another concern is the potential misuse of Rekognition, or any other "law enforcement" AI tool, that uses facial analysis among other features, may be already used, by authoritarian regimes for oppressive purposes and mass surveillance, such as tracking protesters or suppressing any opposing manifestations and rallies.

5.3.5. The right to non-discrimination, along with the right to work and the right to education

The right to nondiscrimination consists one of the most fundamental principles of human rights and is based on another great value, that of equality. Prohibition of discrimination is guaranteed in the normative framework of the EU and can be found by Article 14 of the European Convention for Human Rights, and Article 21 of the Charter of Fundamental Rights of the European Union and in articles 1825 of the Treaty on the Functioning of the European Union.

Discrimination could be defined as the unfair treatment of a person because he or she belongs to a particular group. Discrimination usually aims upon race, religion, sex, ethnicity, sexual orientation and other features. AI systems undermine the right to equality and nondiscrimination of people mainly due to the fact that they have embedded biases on their databases that cannot be easily detected. Big data AI systems use variables that serve as "proxies" for categories such as race and gender and thus cannot be recognized. The process of profiling is discriminatory almost by default, since individuals are being grouped in categories according to various variables, and automated decisions are made based on the profiles within these categories. (Sandvig et al.2016)

Discriminatory AI systems, can be found in almost every area, such as the justice system, the job sector, the loan sector, the insurance sector, as we will see with specific paradigms. (Noble,2018). Biased AI systems that deliberately and systematically discriminate against specific groups in society, based on their sexual orientation, race, gender or financial status pose a great risk not just for the enjoyment of fundamental freedoms from certain individuals but also for the society as a whole and for the values and principles of democracy.

Unfair profiling of AI systems, can create discrimination, for instance excluding or preventing people from employment opportunities, credit or insurance or offering them costly financial products. For example, people living in a certain neighborhood where an ethnic minority is demographically prevalent, are denied credit for this reason alone. This is an explicit form of indirect racial discrimination. (Council of Eyrope,2017,2018)

5.3.5.a) Discrimination in justice

The criminal justice system relies more and more on AI Systemic decision systems. A good example of discrimination in AI systems is the COMPAS AI Systemic tool that we have already mentioned above. COMPAS predicts individuals at high risk of repeating a crime, show future violent behavior and failing to appear, after being out on probation. COMPAS is also used to decide whether to release or detain a defendant or grant them a parole. NGO PROPUBLICA came to the conclusion that: "black defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism, while the white defendants were more likely to be incorrectly flagged as low risk". (Christin et al,2015)

Also, as we have already mentioned above, sometimes the police AI systems for predictive policing which make automated predictions about who will commit crime, or when and where crime will occur, can reproduce and even amplify existing discrimination, against blacks or Latinos in the US. It seems that these systems tend to generate self-fulfilling prophecies and reinforce traditional discriminatory effects. (European Commission, Article 29,2016)

5.3.5.b). Discrimination in the employment and educational sector-ANNEX III (3) Education and vocational training and (4) Employment, worker's management and access to self-employment

AI systems, apart from their discriminatory nature, have also greatly affected several social rights, such as the right to work (Article 15), the right to fair and just working conditions (Article 31), protection in the event of unjustified dismissal (Article 30) and the right to education (Article 14) of the Charter

Selection AI systems of future employees and prospective students mostly in the private sector can have discriminatory effects as well. Major employers, like Unilever, Hilton and Delta Air Lines use data driven hiring tools (Yang, J., 2020), mainly because the large numbers of online applications, cannot be handled efficiently and quickly. These AI systems hiring tools can be discriminatory and cause occupational segregation of employees based on their gender, race and ethnicity, and thus violate the right of an individual to work (Hunkenschroer& Kriebitz,2023)k. It seems that even today, a nonwhite woman is much less "AI System employable" and less likely to be interviewed for a job position in the private sector, compared to a white woman, with the same qualifications. People with disabilities or older workers face systemic employment discrimination by AI Systemic recruiters. AI systems which are used in automated hiring systems may also include online application processes that screen and evaluate completely irrelevant data for the job, or even analyze the applicant's facial expressions or voice, through video interviews. (Bertrand & Mullainathan,2004).

Amazon stopped using a hiring AI System, for screening job applicants because the system was biased against women (Dustin, 2018). According to Reuters," the company (Amazon) realized its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way." and that "Amazon's system taught itself that male candidates were preferable". Google, on the other side showed targeted advertisements for jobs to males, from certain career agencies that promised large salaries. On the contrary, women were shown much fewer ads for high paying jobs. It is also evident that if an AI system, targets high salary job ads only at men, women might not realize that they are excluded from the ad campaign and hence the discriminatory effect of the AI System would never be revealed, due to the opaqueness of the

AI systems. AI systems can theoretically remove and avoid these systemic barriers in employment practices, but in practice proxies used in AI systems can deepen and repeat existing bias and inequalities. Hiring AI systems, for example, that are trained on inaccurate, biased or unrepresentative datasets, can produce highly discriminatory decisions. (Sartor,2017).

Another fundamental human right, that seems to be violated in the workplace by AI systems is to fair and just working conditions. Employers are increasingly monitoring workers' location and activities, using GPS trackers in their smart phones or company vehicles, in order to monitor the time their staff works, the time they don't, or how much time does it take them to finish a work. Worker surveillance and tracking of their exact location, is supposedly aiming in productivity monitoring but, this practice is actually violating their right to privacy and their right fair working conditions. It would be also unacceptable if an AI system, decided to fire an employee because his location tracking was considered problematic. (The Guardian,2016, Lechter, C., 2019)

The right to education is mostly infringed be infringed by AI systems, mostly in the admission and the selection phase in a college. Some colleges are tracking students before they even apply and propose then an interview. They work with consulting companies and data brokers to collect and analyze data for AI systems predictions on prospective students, by tracking, for instance, their Web activity, their family income and the location where they live. This happens, as part of an effort of many public and private universities, mainly in the US, in order to identify students, fit for enrollment. In this way, colleges end up collecting a vast amount of personal data on prospective students, like their test scores, ZIP codes, academic interests, Web browsing histories, ethnic backgrounds and household incomes, for any clues in general that would make prospective students the best candidates for admission. (The New York Times,2019)

These practices, of course, raise a barrier to college education for unprivileged students and creates education discrimination, totally in contrast with the International treaties, that protect this fundamental human right and forbid any discrimination or inequality in this sector. If a university AI system tracks its prospective students by analyzing financial, social, demographical data to offer them an admission, this most probably means that lower income students, or memebersof the LGBT community, or Latinos, or residents of ghettos, trailer parks or public housing blocks, will probably receive a denial. (Selimgo, J.2019)

An example of AI driven discrimination and violation of the right to education ,is coming from a medical school in the UK,which showed that an AI system, it used could lead to discrimination against people of color. Another example of discrimination in the educational field is that of the University of Texas which deployed an algorithmic system to assess prospective PhD candidates in computer science. The algorithm (GRADE – GRaduate ADmissions Evaluator) was trained on data drawn from prior students that had been successful in applications and had been accepted previously. The system reduced the amount of time needed for processing applications but had been under representative in respect to female and black students. The system, trained on successful applicants, predominantly white and male, and consequently was prioritizing factors far more common amongst those students, less so amongst other groups. Even though the system was terminated in 2020 following a backlash amongst students, it continues to be a use case of violation of Article 14, the right to education, and Article 21, the right to nondiscrimination.

5.3.6. The right to health care and the right to social security and social assistance-ANNEX III (5)-Access to and enjoyment of essential private services and essential public services and benefits

Article 35 of the Charter states that "Everyone has the right “of access to preventive health care and the right to benefit from medical treatment». Furthermore, Article 34 of the Charter states that” The Union recognizes and respects the entitlements to social security benefits and social services providing protection in cases of maternity,illness,dependecy,or old age «It also states that “*In order to combat social exclusion and poverty, the Union recognizes and respects the right to social and housing assistance so as to ensure a decent existence for all those who lack sufficient resources, in accordance with the rules laid down by Community law and national laws and practice*s.”

Even though, AI systems are indisputably improving research in health and health care, their implementation can sometimes produce certain harms and violations to the right of health. Predictive AI models are today assessing a patient’s risk, but this method can be unreliable for a variety of reasons, if not even dangerous for the health of people. The main reason for that, is that these models cannot accept any deviation, any new feature from the learned and used datasets they have. If a patient’s medical record and current medical situation has different characteristics, different evidence and does not respond to the proposed model treatment, or responds badly, then the model has made an ineffective recommendation that could put at risk the patient, his health and his life. This fact is extremely crucial when these AI systems are being deployed in intensive care units, where a wrong decision can cost the life of a person, thus violating his right to health and life. Predictive AI systems in the health sector tend to generalize due to their stores datasets, from which they do not deviate. So any shift in characteristics or populations makes these AI systems insufficient, unsafe and can lead to harmful suggestions. (Angwil et al., Propublica,2019)

Furthermore, women, the elderly and minorities are much less well represented in control trials and research and the consequences may be that they are much less likely to receive the right treatment, since their symptoms will most probably not match those of a typical "white adult man», with whose data the AI systems are trained. For example, the underrepresentation of women (particularly women of color) in the research of heart health can be proved very problematic, since by tailoring health solutions of the majority groups to those of the minorities, doctors fail to help all patients.

AI systems can also jeopardize the right to health care, if they are trained and used for profit reasons of the health security system. Countries where huge amounts of money are being invested and moving among big health businesses, can use AI Systemic decision making systems, in order to minimize the costs or increase the profits from high-risk interventions or drugs, or devices or special therapies and end up "selecting' less expensive treatments for low-income patients with aggravated health and more expensive treatments for high-income patients with much fewer health problems. (Jason,2020)

An example of an AI system intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for essential public assistance benefits and services, including healthcare services, as well as to grant, reduce, revoke, or reclaim such benefits and services is the the UK's Department of Work and Pensions (DWP) has developed an automated system for distributing social security benefits, Universal Credit. However, it has been reported that there are significant inaccuracies within the system, has led to people's earnings to be miscalculated, and thus to benefit payouts to be significantly lower than they ought to have been. A report, produced by Human Rights Watch, "Automated Hardship: How the TechDriven Overhaul of the UK's Benefits System Worsens Poverty" details how a poorly designed algorithm is causing people to go hungry, fall into debt, and experience psychological distress.

Another very well-known example, of this category of high risk AI systems in the welfare administration, is that of The Netherlands childcare benefit scandal, also known as the "Toeslagenaffaire". It involved the wrongful accusation of thousands of families of taking by fraud, childcare benefits, leading to severe financial hardship and social stigma for many. The scandal has been widely cited as a case study about the misuse of automated systems in public administration, The Dutch Tax Authority implemented an automated fraud detection system, in order to detect fraud in childcare benefit claims. The system flagged thousands of families as potentially fraudulent based on minor administrative errors or incomplete documentation. A "zero tolerance" policy was applied, meaning that even trivial mistakes, such as a missing signature or incomplete paperwork were treated as intentional fraud. Families flagged by the system were required to repay the full amount of benefits received, often amounting to tens of thousands of euros. It was revealed that over 26,000 families were wrongfully accused of fraud and the impact on these families was enormous. Families were forced to repay massive sums of money, often with no warning. Many faced bankruptcy or severe debt. Victims reported high levels of stress, anxiety, depression and there were even suicide reports. Entire Families were labeled as fraudsters, damaging their reputations and creating further barriers to employment or community support. Children were affected since the financial instability caused some families not being able to afford basic necessities such as food, housing, and education.

As it was expected the system violated vastly the right to nondiscrimination, since it wrongfully and disproportionately families of migrant backgrounds. Indicators such as dual nationality or foreign sounding names were reportedly factors in the risk assessment, leading to systemic racial profiling, if not systemic racism.

<u>5.3.7. The right to asylum Article 18 of the Charter-Annex III AI ACT-Immigration, asylum, border control management</u>

Article 18 of the Charter declares the right to asylum which "*shall be guaranteed with due respect for the rules of the Geneva Convention of 28 July 1951 and the Protocol of 31 January 1967 relating to the status of refugees and in accordance with the Treaty establishing the European Community."*

Still, the increasing use of artificial intelligence (AI) in asylum and immigration processes has raised significant concerns about the potential violation of the fundamental right to asylum. While AI tools are often implemented to enhance efficiency, their application in this sensitive domain has shown critical risks related to fairness, accuracy and human rights compliance.

That AI systems that are being employed in various stages of asylum processing, include usually,1) Screening and Risk Assessment, where AI systems are used to screen asylum applications and assess the "risk" posed by applicants. For example, tools may analyze documents, language patterns, or behavioral data to flag potential illegal claims.,2) Facial Recognition and Biometric Analysis, where the system as we already discussed processes facial recognition, fingerprints, irises in order to identify asylum seekers or track their movements across borders.

There are many human rights concerns, regarding the use of these AI systems, as many of them have being proven to use unethical methods, methods that undermine human dignity and the autonomy of the individuals, along with the right to asylum and the right to free movement.

For instance, **the iBorderCtrl system,** which is a machine learning based video with a 'lie detector module,' and was deployed in Hungary, Greece and Latvia, 'used an avatar of a border guard to ask people 13 questions about their personal backgrounds and travel plans, assessed travelers' micro expressions, and notified human agents if it suspected them of being dishonest.'

The tool was criticized as 'pseudoscientific'. 'dystopian' and as leading to an 'Orwellian nightmare'. These criticisms are grounded on findings that evidence inaccuracies and bias, especially in its facial recognition systems. Of course the right to asylum was heavily violated here since, if a border guard was informed falsely by the AI lie detector module that a person's micro expressions are dishonest, then thetraveller would be refused entry.

Another example is that of Canada's asylum system, which was administered by Immigration, Refugees and Citizenship Canada (IRCC) and has increasingly incorporated artificial intelligence (AI) to manage immigration and asylum processes. While the government has framed this as a move toward efficiency and faster decision-making, the use of AI in such sensitive areas has sparked significant concerns regarding its impact on human rights, including the rights to asylum, due process, and nondiscrimination. The system used 'advanced data analytics' to process temporary visa applications, although it is primarily limited to simple applications, while more challenging cases are decided on by an officer.

IRCC did publish an impact assessment, which concluded that the impact level of the system was moderate, but this assessment was criticized since a report from the University of Toronto Citizen lab considered that the risks were higher, describing it as a 'high risk laboratory'.

5.3. 8. Case studies of human rights violations, under High Risk AI systems

5.3.8.a) The case of Cambridge Analytica. A paradigm of multi violation of several human rights. (right to privacy, freedom of expression, freedom of elections, right to data protection)

The 2018 Cambridge Analytica scandal is a highly discussed case that demonstrates the great impact of AI systems, especially profiling, on fundamental human rights, especially those related to privacy, data protection, freedom of elections and freedom of thought.

In 2013 Aleksandr Kogan and his company, Global Science Research, created an app named "thisisyourdigitallife" which was offered by the platform of Facebook. The app invited users to answer some questions in order receive some kind of a psychological profile. These questions could include the user's name, gender, birthday, relationship status, political views, religious beliefs, location and even in some cases their private messages. In order to install the app, in the first place, the users were required to give access not only to their data but also to their friends' personal data. It is estimated that 270.000 users installed it and almost the data of 50 million Facebook profiles were collected. In 2015, Kogan and Global Science Research sold the personal data they collected from this app, to Cambridge Analytica, a British political consulting firm, which used psychometric profiling of individuals in order to target them with personalized advertisements, which matched their preferences.

Cambridge Analytica was then hired for Donald Trump's presidential campaign and the Leave. EU campaign in the UK Brexit referendum. On 17th of March 2018, a whistleblower exposed Cambridge Analytica's illegal use of data collected from Facebook, which led to a huge debate about the effects of illegal acquisition of personal data and the impacts of profiling and targeted advertising to human rights. First of all, Facebook users were completely unaware that just by installing an app, such a large amount of personal data belonging to them and their friends, would possibly be collected and more precisely for the purposes of targeted advertising. It is more than obvious that their **right to privacy** was breached and that they had been misled and manipulated. It also seems that largescale profiling for psychographic targeted advertising from Cambridge Analytica, helped enormously in securing victories for both of these campaigns and in this way distorted the result of these elections, by **breaching the right to free elections.** If this conclusion is valid, that means that AI Systemic profiling used in targeted advertising can have a tremendously negative effect in democracy and its institutions and also on **individuals' free will and self-determination.**

Cambridge Analytica' s scandal demonstrated that political actors can use technology in order to manipulate people and moreover create "dark ads" that target small groups of their supporters and intent to polarize the public using fake news, negative messaging and falsified information of the opposing party. (The New York Times,2018)

5.3.8.b) The Schufa case: A social credit scoring AI tool with human rights multiviolations

Schufa Holding AG was a German credit scoring agency that assessed individuals' financial situation, through an AI scoring system. The automated credit scores, it generated were then shared with banks, in order to guide decisions about loans, mortgages, and other financial products. Schufa used opaque techniques, meaning black box algorithms, with data that were personal, (sometimes even demographic factors such as ones' family status were used, married/divorced), which were never disclosed to the concerned individuals. And even though, someone may claim that Schufa itself did not make final lending decisions, its scores had a determinative role in the decisions made, thus they heavily influenced the banks's choices.

As it is quite obvious, this AI tool and its services, violated a number of human rights under the Charter and the European Convention of Human Rights such as the right to privacy (Article 8 of ECHR)since it processed sensitive personal data without consent, the right to nondiscrimination under article 14 of the ECHR,due to the fact that the automated credit scoring system, was probably biased, by having incorporated biases, based on socioeconomic factors, a fact that leads to discriminatory results. For instance, individuals from low-income households or with limited would receive lower scores unjustly and thus the inequalities and the discrimination would be continued and fortified.

Another very important human right which was violated was that of the right to an Effective Remedy, under Article 47 of the Charter. The people affected by incorrect or unfair credit scores had in fact no right to an appeal, no option to contest these scores and this lack of a right to contest and challenge these decisions more than significantly impacted their lives. Last but not least, chufa's credit scores had vast consequences for different economic and social rights, such as the right to housing, the right to pursue a job, through opening a business, the right to education, through a student loan.

5.3.8.c)The SyRI (System Risk Indication) case

The SyRI (System Risk Indication) case, involved an AI system, designed to detect welfare fraud, which was used by the Dutch government and raised great concerns regarding human rights violations, regarding privacy, discrimination, and government accountability. SyRI combined and analyzed large datasets from various public institutions. The datasets were employment records, tax data, social benefits information and housing data. The scope was to track individuals or households which were considered considered high-risk for deceptive activities. But these large datasets included sensitive personal data, such as income and social security information. As it is well understood, the algorithmic logic behind SyRI was opaque, on transparent since there was no disclosure of the criteria used in the risk analysis and how were the individuals tracked and identified.(Wieringa,2023,Rachovitsa& Johann,2022).

It was then found that SyRI disproportionately targeted low-income neighborhoods, often with higher populations of immigrants or ethnic minorities, something that implied heavily that the AI tool was biased and in a very systematic way, since it used public agencies' data. It actually showed that the State was involved in marginalizing its own communities. As in the Schufa case, individuals were not informed about their inclusion in the system and they were deprived of their right to contest the decisions made about them.

Likewise, in the Schufa case, many human rights were violated by the AI tool deployed by the Dutch government and a subsequent scandal emerged. First and foremost, the right to privacy under Article 8 of the Charter was infringed since there was an excessive data processing of sensitive personal data. Meanwhile, the system was acting as a black box, in the dark, denying individuals information about which and how their personal data were used or analyzed. Then the right to nondiscrimination, under article 14 of the Charter, since there was a significant and more than obvious geographic and demographic targeting. The targeting stemmed from the algorithmic biases of the AI system which focused on low-income neighborhoods, where usually by default marginalized communities such as immigrants and ethnic minorities (like Romani's) live. The Syri case in particular, presented the Dutch government as an enhancer of systemic inequalities, as a reinforce of existing social and economic inequalities of vulnerable populations. In addition, the right to an effective Remedy, under Article 47 of the Charter was breached, since SyRI's lack of transparency not only denied individuals of any explanation or notification about the decisions affecting them but also they were not given the opportunity to challenge those decisions. There was also no human oversight, as in the Schufa case.

In the end, the Dutch District Court's decided to halt SyRI's operation, a historic decision that highlighted the importance of protection of the human rights human rights, in the design and deployment of such systems.

6.1.Protecting Human Rights Through the Fundamental Rights Impact Assessment (FRIA), Under the EU AI Act

Article 27 introduces the Fundamental Rights Impact Assessment (FRIA), a mandatory evaluation for certain deployers of high-risk AI systems. The primary objective is to ensure that the deployment of these AI systems does not infringe upon fundamental human rights, such as privacy, nondiscrimination, and human dignity. This requirement underscores the EU's commitment to a human centric approach to AI, emphasizing ethical considerations in technological advancements.

Article 27 from the EU Artificial Intelligence Act (Regulation (EU) 2024/1689), titled "Fundamental Rights Impact Assessment for High Risk AI systems" declares that"

"1. Prior to deploying a high-risk AI system referred to in Article 6(2), deployers that are bodies governed by public law, or are private entities providing public services, and deployers of high-risk AI systems referred to in points 5(b) and (c) of Annex III, shall perform an assessment of the impact on fundamental rights that the use of such system may produce. For that purpose, deployers shall perform an assessment consisting of: (a) a description of the deployer's processes in which the high risk AI system will be used in line with its intended purpose; (b) a description of the period of time within which, and the frequency with which, each high risk AI system is intended to be used; (c) the categories of natural persons and groups likely to be affected by its use in the specific context;(d) the specific risks of harm likely to have an impact on the categories of natural persons or groups of persons identified pursuant to point (c) of this paragraph, taking into account the information given by the provider pursuant to Article 13; (e) a description of the implementation of human oversight measures, according to the instructions for use; (f) the measures to be taken in the case of the materialization of those risks, including the arrangements for internal governance and complaint mechanisms”

This article sets out the methodology for conducting a fundamental rights impact assessment (FRIA) for certain high-risk AI systems, detailing the necessary components of the assessment and the obligations of deployers to ensure compliance with fundamental rights considerations. First if all, potential risks for human rights should be intedintfied, beginning with a comprehensive mapping of potential risks to fundamental rights, linking them to specific components of the AI system, such as datasets, algorithms, and decision-making processes. This granular approach ensures that no aspect of the system escapes control.

The other important element is that the the obligation to perform a FRIA primarily falls on deployers that are bodies governed by public law or private operators providing public services and operators providing high-risk systems referred to in Annex III of the AI Act, particularly those related to law enforcement, migration, asylum, border control management, and administration of justice.

Moreover, according to the Article, the FRIA must be conducted before the first use of the high-risk AI system and it should inform on how and when the system will be used, the context of its deployment, and the objectives it aims to achieve, analyze how the system might affect individuals' fundamental rights, considering factors such as discrimination, privacy infringements, and impacts on human dignity. Meanwhile it should propose measures to prevent or minimize identified risks, including human oversight mechanisms(Gatt et al.2024).

6.1.1. Legal Roots and Nature of FRIA

Mantelero (2024) highlights that FRIA’s roots lie, as introduced in the EU AI Act, builds upon established impact assessment methodologies, like the Data Protection Impact Assessment (DPIA) and Human Rights Impact Assessment (HRIA). These frameworks emphasize assessing risks to individuals’ rights and freedoms through structured evaluation processes in the broader risk based regulatory approach adopted by the EU, particularly in addressing the societal implications of emerging technologies. Unlike traditional assessments focused solely on compliance, FRIA emphasizes on an anticipatory model approach, proactively identifying and mitigating risks to fundamental rights, **before AI systems are deployed**. The AI Act's risk based approach categorizes AI systems based on their potential impact on human rights as we saw, so this framework aims to ensure that AI developers assess not only technical risks but also broader societal and ethical implications. Mantelero underlines, that FRIA serves multiple roles like preventing human rights violations, fostering accountability, and integrating ethical considerations into AI design. By bridging the gap between legal obligations and operational practices, FRIA provides a structured pathway for aligning AI innovation with fundamental rights.

In order to develop, an effective FRIA methodology it is required to integrate insights from existing impact assessments while addressing AI's unique challenges. Mantelero identifies key criteria for FRIA's design such as 1) a comprehensive risk analysis, where FRIA must evaluate a wide range of risks, from privacy infringements to discriminatory outcomes. This necessitates adopting a multidimensional approach that considers individual, collective, and societal impacts. For example, FRIA should identify risks arising from biased training data, algorithmic opacity for which we discussed, and surveillance applications, ensuring that all potential adverse outcomes are mapped and addressed,2) integration of other assessments, like the DPIA Data Protection Privacy Impact Assessment and the HRIA (Human Rights Impact Assessment). FRIA should complement these existing risk management processes, such as to provide a holistic evaluation of AI systems and 3) a scenario based testing where evaluators will simulate potential adverse outcomes under different conditions. This approach helps identify vulnerabilities and ensures that mitigation strategies are robust and adaptable to varied real world applications. Scenario analysis also aids in understanding the negative effects of AI related risks across different sectors and populations,4) a Real World Application and Testing, where the emphasis on empirical testing and real-world applications is one of the article's most significant contributions. Testing environments allow for the practical validation of theoretical principles, ensuring that assessments remain relevant and actionable.(Mantelero,2024)

The strengths of the Proposed Framework,acccording to Malgieri & Santos(2024) are the avoidance of Oversimplification where Traditional models often treat rights violations as absolute, failing to recognize less severe but still meaningful violations, the contextual Sensitivity, meaning that the framework is more sensitive to the specific contexts in which violations occur, by taking into account cultural, societal, and situational factors, the recognition of subjective the importance of societal and individual experiences, such as cultural norms and personal sensitivities that capture and demonstrate the lived realities of those affected by digital technologies. And last but not least, Real Life Consequences, by evaluating real and practical impacts, such as economic harm or psychological distress,

Epilogue

As this paper has shown, the EU Artificial Intelligence Act (AI Act) represents a groundbreaking milestone in addressing the risks and challenges posed by AI systems, while ensuring that technological innovation is consistent with fundamental human rights. The Act's comprehensive risk-based approach - distinguishing between unacceptable, high, limited and minimal risks - embodies a nuanced understanding of AI's dual capacity to enhance and threaten societal values.

The importance of this regulatory framework cannot be overstated. By explicitly prohibiting practices such as subliminal manipulation, exploitative profiling, and real-time biometric surveillance, the AI Act establishes robust safeguards against abuses that threaten human dignity, privacy, and equality. Moreover, its stringent requirements for high-risk systems, including compliance assessments and transparency measures, underscore the EU's commitment to mitigating potential harm without stifling innovation.

The analysis of case studies in this paper reveals the tangible impact of AI systems on democratic processes, civil liberties and social justice. From the dangers of social scoring and discriminatory profiling to the erosion of trust caused by opaque decision-making, these examples highlight the urgent need for legal and ethical guardrails. They also highlight the profound impact on vulnerable groups whose rights and freedoms are disproportionately affected by unregulated AI.

At its core, the AI Act promotes a human-centered vision of AI - one in which technology serves as a tool to enhance well-being, protect individual rights, and promote societal progress. This vision is underpinned by its alignment with the European Union's Charter of Fundamental Rights, which emphasizes values such as democracy, the rule of law and environmental protection. By enshrining these principles in a legal framework, the EU is setting a precedent for other nations to follow.

However, the way forward is not without with challenges. The rapid evolution of AI technologies requires an adaptive regulatory approach that can address emerging risks and complexities. Issues such as algorithmic bias, lack of transparency and ethical dilemmas will require ongoing dialogue, collaboration and vigilance. In addition, public awareness and education will be critical to empowering individuals to understand and navigate the implications of AI in their daily lives.

In conclusion, the AI Act is more than a regulatory instrument; it is a declaration of the EU's determination to prioritise human values in the face of technological disruption. As AI continues to shape our world, this thesis underscores the urgency of fostering a global consensus on ethical AI governance. Through the lessons of the AI Act, we are reminded that innovation and human rights are not mutually exclusive, but can coexist harmoniously in the pursuit of a more just and equitable future.

.

Bibliography

1. Abraham, K. S., & Rabin, R. L. (2019). Automated vehicles: Legal and policy implications. Stanford Law Review, 71(2), 107–146.

2. AI System Watch. (n.d.). OpenSCHUFA Shedding light on Germany’s opaque credit scoring. Retrieved September 28, 2018, from https://AI Systemwatch.org/en/openschufasheddinglightongermanysopaquecreditscoring/

3. AIHLEG High-level Expert Group on Artificial Intelligence. (2019). A definition of AI: Main capabilities and scientific disciplines. European Commission. https://ec.europa.eu/futurium/en/aiallianceconsultation

4. AIHLEG High-level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy AI. European Commission. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

5. Almada, M., & Petit, N. (2023). The EU AI Act: Between product safety and fundamental rights.

6. Angwin, J., Kirchner, L., Mattu, S., & Larson, J. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against Blacks. ProPublica. https://www.propublica.org/article/machinebiasriskassessmentsincriminalsentencing

7. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2019, October 24). Millions affected by racial bias in health care AI System. ProPublica. https://www.propublica.org/article/millionsaffectedbyracialbiasinhealthcareAI System

8. Article 19. (2016). AI systemsand automated decision-making in the context of crime prevention: A briefing paper. https://www.article19.org/data/files/medialibrary/38285/160715AI systemsandcrimepreventionA19.pdf

9. Article 29 Data Protection Working Party. (2017). Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/items/612049

10. Article 29 Data Protection Working Party. (2018). Guidelines on transparency under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/items/612048

11. Bekker, S. (2021). Fundamental Rights in Digital Welfare States: The Case of SyRI in the Netherlands. In Spijkers, O., Werner, W.G., & Wessel, R.A. (Eds.), Netherlands Yearbook of International Law (Vol. 50, pp. 2024). T.M.C. Asser Press, The Hague. https://doi.org/10.1007/9789462654037_24

12. Bellovin, S. M., Blaze, M., & Anderson, R. (2018). Security and privacy in the digital age. Communications of the ACM , 61(1), 4654. https://doi.org/10.1145/3173567

13. Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. The American Economic Review, 94(4), 991–1013. https://doi.org/10.1257/0002828042002561

14. Binns, R., & Johnson, N. (2019). Automating power: Social bot interference in global politics. International Journal of Communication , 13, 43254346. Retrieved from https://ijoc.org/index.php/ijoc/article/view/11088

15. Borgesius, F. J. Z. (2016). Personal data protection and electronic communications: The case of profiling and targeted advertising. In Privacy Due Process and the Computational Turn: The

Philosophy of Law Meets the Philosophy of Technology (pp. 119–132). Routledge. https://doi.org/10.4324/9781315564059

16. Borgesius, F. J. Z., Trilling, D., & Möller, J. (2016). Should we worry about filter bubbles?. In Privacy and data protection: A comparative analysis of the GDPR and its impact on the EU's digital economy (pp. 121–138). Routledge.

17. Bossi, A., & Ferrara, E. (2016). Social bots: How they manipulate and shape public opinion. Journal of Social Media Studies , 3(2), 45–62. https://doi.org/10.1016/j.jsms.2016.05.002

18. Bozdag, E. (2013). Bias in AI Systemic filtering and personalization. Ethics and Information Technology , 15(3), 209–227. https://doi.org/10.1007/s1067601393216

19. Bulgakova, D. (2023). The Prohibited Artificial Intelligence Practice. Theory and Practice of Forensic Science and Criminalistics, 32(3), 89-112.

20. Burell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning AI systems. Big Data & Society .

21. Butt, J. S. (2024). Analytical study of the world's first EU Artificial Intelligence (AI) Act, 2024. International Journal of Research Publication and Reviews

22. Butt, J. (2024). Analytical Study of the World's First EU Artificial Intelligence (AI) Act. International Journal of Research and Publications, 5(3).

23. Buttarelli, G. (2016). The European data protection reform: A turning point for the protection of fundamental rights. Retrieved from https://edps.europa.eu/sites/edp/files/publication/161209_edps_speech_giuseppe_buttarelli_en.pdf

24. CastetsRenard, C. (2019). Accountability of AI systems: A European perspective. European Data Protection Law Review , 5(3), 275285. https://doi.org/10.21552/edpl/2019/3/9

25. Christin, A., Rosenblat, A., & Boyd, D. (2015). A new era of judging: AI Systemic decision-making and its implications for justice. In Proceedings of the 2015 ACM Conference on Computer Supported Cooperative Work & Social Computing (pp. 70–81). Association for Computing Machinery. https://doi.org/10.1145/2675133.2675150

26. Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. Washington Law Review , 89(1), 1–33.

27. Confessore, N. (2018, April 4). Cambridge Analytica and Facebook: The scandal and the fallout so far. The New York Times .

28. Council of Europe. (1950). European Convention on Human Rights (ETS no. 5). Retrieved from https://www.echr.coe.int/Documents/Convention_ENG.pdf

29. Council of Europe. (2018). AI systems and human rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications. Council of Europe. https://rm.coe.int/AI systemsandhumanrightsstudyonthehumanrightsdimensions/16808f699c

30. Council of Europe. (2017, January 17). Guidelines on the protection of individuals with regard to the processing of personal data in a world of big data. Retrieved from https://rm.coe.int/16806f5d82

31. Data & Society. (2018). AI Systemic accountability: A primer. Retrieved from https://datasociety.net/pubs/ia/DataSociety_AI Systemic_Accountability_Primer_2018.pdf

32. Diakopoulos, N. (2016). Accountability in AI Systemic decision making. Communications of the ACM , 59(2), 56–64. https://doi.org/10.1145/2832347

33. Diakopoulos, N. (2014). AI Systemic accountability: The investigation of Black Boxes. Tow Center for Digital Journalism. http://www.nickdiakopoulos.com/wpcontent/uploads/2011/07/AI SystemicAccountabilityReporting_final.pdf

34. Eckhouse, L. (2017). Big data may be reinforcing racial bias in the criminal justice system. ProPublica . https://www.propublica.org/article/bigdatareinforcingracialbiasincriminaljusticesystem

35. Edwards, L. (2021). The EU AI Act: a summary of its significance and scope. Artificial Intelligence (the EU AI Act), 1.

36. Edwards, L. (2021). The EU AI Act: a summary of its significance and scope. Artificial Intelligence (the EU AI Act), 1.

37. European Parliament and Council. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending various regulations and directives (Artificial Intelligence Act). *Official Journal of the European Union, L 1689*, 12.7.2024. ELI: http://data.europa.eu/eli/reg/2024/1689/oj

38. European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union , L119, 188. https://eurlex.europa.eu/legalcontent/EN/TXT/?uri=CELEX%3A32016R0679

39. Floridi, L. (2013). On human dignity as a foundation for the ethics of information. In The ethics of information (pp. 41–60). Oxford University Press.

40. Friedman, M. (2019, September 11). For sale: Survey data on millions of high school students. The New York Times . https://www.nytimes.com/2019/09/11/technology/surveydatahighschoolstudents.html

41. G20 Insights. (2021). Governing data and artificial intelligence for all. Retrieved from https://www.g20insights.org/policy_briefs/governingdataandartificialintelligenceforall/

42. Gatt, L., Caggiano, I. A., Gaeta, M. C., Savella, R., Troisi, E., Pratesi, F., & Trasarti, R. (2024, October). FRIA Implementation Model According to the AI Act. In 2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE) (pp. 1224-1229). IEEE.

43. Goel, S., Perelman, J., Shroff, R., & Sklansky, D. (2017). Combatting police discrimination in the age of big data. New Criminal Law Review , 20(2), 181–232. https://doi.org/10.1525/nclr.2017.20.2.181

44. Grassegger, H., & Krogerus, M. (2019). The power of AI systems: How AI systems shape our lives. The New Press .

45. Grassegger, H., & Krogerus, M. (2017, February 17). The data that turned the world upside down. Motherboard . https://motherboard.vice.com/en_us/article/mg9vvn/howourlikeshelpedtrumpwin

46. Griffin, A. (2016, August 31). How Facebook is manipulating you to vote. The Independent . http://www.independent.co.uk/lifestyle/gadgetsandtech/news/ukelections2016howfacebookismanipulatingyoutovotea7015196.html

47. Hao, K. (2018, July 26). Amazon's face recognition falsely matched 27 lawmakers with mugshots. MIT Technology Review . Retrieved from https://www.technologyreview.com/2018/07/26/amazonfacerecognitionfalselymatched27lawmakers/

48. Hildebrandt, M., & Gutwirth, S. (2008). General introduction and overview. In Profiling the European citizen (pp. 1–13). Springer. https://link.springer.com/chapter/10.1007/9781402069147_1

49. Hildebrandt, M. (2009). Profiling and AML. In K. Rannenberg, D. Royer, & A. Deuker (Eds.), The future of identity in the information society: Challenges and opportunities (pp. 37–50). Springer. https://doi.org/10.1007/9783642024820_3

50. Hildebrandt, M. (2014). Location data purpose binding and contextual integrity: What's the message? In L. Floridi (Ed.), The protection of information and the right to privacy: A new perspective (pp. 31–62). Springer.

51. Hristozova, M. (2024). THE EU ARTIFICIAL INTELLIGENCE ACT: THE NECESSARY TOOL TO GUARANTEE THE FUNDAMENTAL RIGHTS OF CITIZENS. KNOWLEDGE PROCEEDINGS, 44(1), 91-96.

52. https://www.aclu.org/pressreleases/lawsuitchallengesgovernmentssuspiciousactivityreportprogram

53. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/

54. https://www.hrw.org/report/2020/09/29/automated-hardship/how-tech-driven-overhaul-uks-social-security-system-worsens

55. https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/dwp-disability-benefits-fraud-algorithm

56. https://www.article19.org/resources/eu-risky-biometric-technology-projects-must-be-transparent-from-the-start/

57. https://www.dpa.gr/en/enimerwtiko/press-releases/ministry-migration-and-asylum-receives-administrative-fine-and-gdpr

58. https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/ircc-immigration-and-visa-applications-automation

59. Hunkenschroer, A. L., & Kriebitz, A. (2023). Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring. AI and Ethics, 3(1), 199–213. https://doi.org/10.1007/s43681-022-00234-5

60. Håkonsen, A. (2024). The European Rights-Driven Regulatory Approach in the Technology Sector and its Extraterritorial Impact A Comparative Case Study of the GDPR and the AI Act (Master's thesis, NTNU).

61. Jason, J. (2020). Artificial intelligence for health and healthcare. MITRE Corporation . https://www.mitre.org/publications/technicalpapers/artificialintelligencehealthcare

62. Kaye, D. (2018). Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. United Nations General Assembly , A/HRC/38/35. Retrieved from https://www.ohchr.org/en/documents/reports/reportspecialrapporteurpromotionandprotectionright freedomopinionand

63. Kroll, J. A., et al. (2016, March). Accountable AI systems. Balkinization . http://balkin.blogspot.co.at/2016/03/accountableAI systems.html

64. Kusche, I. (2024). Possible harms of artificial intelligence and the EU AI act: fundamental rights and risk. Journal of Risk Research, 1–14. https://doi.org/10.1080/13669877.2024.2350720

65. Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016, May 23). How we analyzed the COMPAS recidivism AI System. ProPublica . https://www.propublica.org/article/howweanalyzedthecompasrecidivismAI System

66. Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., & Briggs, M. (2021). Artificial intelligence, human rights, democracy, and the rule of law: a primer. arXiv preprint arXiv:2104.04147.

67. López, A. L. (2024). The systematics of the European Artificial Intelligence Act in the context of the fundamental rights of the Union: the myth of the digital constitutionalism. Deusto Journal of Human Rights, (14), 73-100.

68. Malgieri, G., & Rebrean, M. L. (2024). Vulnerability in the EU AI Act: building an interpretation. Available at SSRN 5058591.

69. Malgieri, Gianclaudio and Santos, Cristiana, Assessing the (Severity of) Impacts on Fundamental Rights (June 25, 2024)

70. Mantelero, A. (2024). The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. Computer Law & Security Review, 54, 106020. https://doi.org/10.1016/j.clsr.2024.106020

71. Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of AI systems: Mapping the debate. Big Data & Society , 3(2), 121. https://doi.org/10.1177/2053951716679679

72. Musch, S., Borrelli, M., & Kerrigan, C. (2023). The EU AI Act: A Comprehensive Regulatory Framework for Ethical AI Development. Available at SSRN 4549248.

73. Neuwirth, R. J. (2023). Prohibited artificial intelligence practices in the proposed EU artificial intelligence act (AIA). Computer Law & Security Review, 48, 105798.

74. Noble, S. U. (2018). AI systems of oppression: How search engines reinforce racism. NYU Press.

75. Organisation for Economic Cooperation and Development. (2019). OECD principles on artificial intelligence. Retrieved from https://www.oecd.org/goingdigital/ai/principles/

76. Palmiotto, F. (2024). The AI Act Roller Coaster: How Fundamental Rights Protection Evolved in the EU Legislative Process.

77. Pasquale, F. (2015). The black box society: The secret AI systemsthat control money and information. Harvard University Press.

78. Pasquale, F. (2019). The second wave of AI Systemic accountability. Law and Political Economy . Retrieved from https://lawandpoliticaleconomy.org/thesecondwaveofAI Systemicaccountability/

79. Pehlivan, C. N. (2024). The EU Artificial Intelligence (AI) Act: An Introduction. Global Privacy Law Review, 5(1).

80. Perry, W. L. (2013). Predictive policing: The role of crime forecasting in law enforcement operations. RAND Corporation.

81. Pham, B. C., & Davies, S. R. (2024). What problems is the AI act solving? Technological solutionism, fundamental rights, and trustworthiness in European AI policy. Critical Policy Studies, 1-19.

82. Privacy International & Article 19. (2018). The human rights impact of government use of facial recognition technology. Retrieved from https://www.privacyinternational.org/report/1810/humanrightsimpactgovernmentusefacialrecognitiontechnology

83. ProPublica. (2016, May 28). Machine bias. ProPublica . https://www.propublica.org/article/machinebiasriskassessmentsincriminalsentencing

84. Rosenberg, M., Confessore, N., & Cadwalladr, C. (2018, March 17). How Trump consultants exploited the Facebook data of millions. The New York Times . https://www.nytimes.com/2018/03/17/us/politics/cambridgeanalyticatrumpcampaign.html

85. Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2016). When the AI System itself is a racist: Diagnosing ethical harm in the basic components of software. International Journal of Communication , 10, 19. https://ijoc.org/index.php/ijoc/article/view/5641

86. Sartor, G. (2017). Human rights and information technologies. In Brownsword, R., Scotford, E., & Yeung, K. (Eds.), The Oxford Handbook on the Law and Regulation of Technology (pp. 424450). Oxford University Press.

87. Smith, J. (2024). How Facebook knows which apps you use and why this matters. The Digital Privacy Journal . Retrieved from https://www.digitalprivacyjournal.com/howfacebooktracksapps

88. Stahl, B. C., Rodrigues, R., Santiago, N., & Macnish, K. (2022). A European Agency for Artificial Intelligence: Protecting fundamental rights and ethical values. Computer Law & Security Review, 45, 105661.

89. Stanford News. (2015, January 12). New Stanford research finds computers are better judges of personality than friends and family. Stanford News . http://news.stanford.edu/2015/01/12/personalitycomputerknows011215/

90. STOA Panel for the Future of Science and Technology. (2016). Understanding AI Systemic decisionmaking: Opportunities and challenges. European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/STUD/2016/573989/EPRS_STU(2016)573989_EN.pdf

91. STOA Panel for the Future of Science and Technology. (2018). The impact of General Data Protection Regulation on artificial intelligence. European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/STUD/2018/617975/EPRS_STU(2018)617975_EN.pdf

92. STOA Panel for the Future of Science and Technology. (2020). A governance framework for AI Systemic accountability. European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/658572/EPRS_STU(2020)658572_EN.pdf

93. Sullivan, M. (2020, November 17). Political bots spread misinformation: How automated accounts are influencing the discourse. The Atlantic . Retrieved from https://www.theatlantic.com/technology/archive/2020/11/politicalbotsspreadmisinformation/617035/

94. The Guardian. (2016, August 2). How AI systemsrule our working lives. The Guardian . https://www.theguardian.com/technology/2016/aug/02/howAI systemsruleourworkinglives

95. The Guardian. (2017, April 13). AI programs exhibit racial and gender biases research reveals. The Guardian . https://www.theguardian.com/technology/2017/apr/13/aiprogramsexhibitracistsexistbiasesresearchreveals

96. The New York Times. (2019, September 23). How game apps that captivate kids also captivate their data. The New York Times . Retrieved from https://www.nytimes.com/2019/09/23/technology/howgameappscaptivatekids.html

97. Tufekci, Z., York, J. C., Wagner, B., & Kaltheuner, F. (2015). The ethics of AI systems: From radical content to selfdriving cars . European University Viadrina. Retrieved from

https://www.euvfrankfurto.de/fileadmin/user_upload/Veranstaltungen/2015/Ethics_of_AI systems.pdf

98. United Nations. (1948). Universal Declaration of Human Rights . https://www.un.org/en/aboutus/universaldeclarationofhumanrights

99. United Nations. (2011). Guiding principles on business and human rights: Implementing the United Nations ‘Protect Respect and Remedy’ framework (pp. 2024). Office of the United Nations High Commissioner for Human Rights. http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

100. Wagner, B. (2016). AI Systemic regulation and the global default: Shifting norms in Internet governance. Etikk i praksis. Nordic Journal of Applied Ethics , 10(1), 5–19. https://doi.org/10.5324/eip.v10i1.1978

101. Wagner, K. (2020, February 6). Facebook uses artificial intelligence to predict your future actions. Vox . https://www.vox.com/2020/2/6/21123289/facebookaipredictionfutureactions

102. Wagner, K. (2020, October 27). The real bias built into Facebook’s ad system. Vox . https://www.vox.com/2020/10/27/21539156/facebookadbiasAI System

103. Wiek, K. (2023). The Artificial Intelligence Act-The Impact of AI on Human Rights Standards in European Law Enforcement (Bachelor's thesis, University of Twente).

104. Wörsdörfer, M. (2024). Mitigating the adverse effects of AI with the European Union's artificial intelligence act: Hype or hope?. Global Business and Organizational Excellence, 43(3), 106-126.

105. Yeung, K. & Harken, A. (2023), ‘How do ‘technical’ design-choices made when building algorithmic decision-making tools for criminal justice authorities create constitutional dangers? Part II’, Public Law, p.3-4, Sweet & Maxell

106. Zang, B., StallaBourdillon, S., & Gilbert, B. (2016). The transparency paradox: Building a sustainable legal framework for AI Systemic accountability in the EU. Computer Law & Security Review , 32(5), 756–768. https://doi.org/10.1016/j.clsr.2016.05.003