LUND UNIVERSITY

AI-based Automated Decision Making:

An investigative study on how it impacts the rule of law, and the case for regulatory safeguards

Author:
Ahmed Zaroff (Sean)

Lund University
Sociology of Law Department

Master's Programme:

Sociology of Law

Master Thesis
(SOLM02) - 2022

Supervisor: Jannice Käll
Examiner: Michael Molavi

Word Count: 21,982

Abstract

The development and expansion of artificial intelligence have significant potential to benefit humanity; however, the risks posed by AI-related tools have also become a growing concern over the past decade. From the standpoint of human rights violations AI-related bias, discriminatory practices, data protection practices and violations or potential infringements on fundamental rights are some of the core concerns revolving around this evolving technology.

This research inquiry primarily focuses on investigating ongoing discourse around AI-based digital surveillance, predictive policing and assessing the prospective contributions by automated decision-making. The study will critically review and discuss the impact AI-based technology has on policing, law enforcement and the rule of law in a democratic society, and how it could potentially influence the broader aspects of social justice. Moreover, this research inquiry investigates and critiques the 'biases' that allegedly exist within AI-based systems and deployment practices that have impacted certain communities more than others. The study focuses primarily on Europe and the U.S., with potential broader ramifications for other countries.

Accordingly, the research examines the need for enhanced legal safeguards, i.e., regulatory intervention, which has been a long-standing and ongoing public request. Consequently, this investigation was carried out through a discourse analysis of European and American cases on this topic, supplemented by content analysis of the various EU regulatory and legislative provisions, supported by a qualitative research mixed-method approach, including participant interviews with industry practitioners and impacted families. This research paper would complement the current research on the consequences of AI practices involving automated decision-making and contributes towards challenging the current AI-related industry policies and practices concerning transparency and accountability.

It is therefore of utmost importance to constantly question the EU's powerful position from an accountability standpoint. This includes the need for its attention and intervention, towards certain 'private actors' (which include large-scale multinational tech giants) and their relationship with state agencies. This is particularly important in the current context, where most public services and functions are increasingly being outsourced to and carried out by the very same 'private actors' using AI tools that are largely self-regulated.

Keywords:

Artificial intelligence, automating governance, automated decision making, rule of law, transparency, accountability, profiling, predictive policing, surveillance capitalism

Acknowledgement

Thank you very much to everyone who has helped me during this process of research inquiry – it has been both enlightening and challenging at the same time. A special thanks to my Faculty Director, Ida Nafstad, and Supervisor, Jannice Käll, who have been extremely helpful and cooperative during the entire period.

AI-based Automated Decision Making:
An investigative study on how it impacts the rule of law, and the case for regulatory safeguards

Table of Contents

Page

List of abbreviations

AI	– Artificial Intelligence
ADM	– Automated Decision Making
AI HLEG	– Artificial Intelligence High Level Expert Group
ECHR	– European Court of Human Rights
EU	– European Union
GDPR	– General Data Protection Regulation
HR	– Human Rights
ML	– Machine Learning
R&D	– Research and Development
RQ	– Research Question

01

Introduction

1.1 Background

Public power and governance are increasingly becoming subject to artificial intelligence (AI)-based automation. Consequently, the crucial question is fast shifting from "*how to regulate such evolving and self-regulating AI technologies*" towards "*how the technology we use, regulates us*" (Greenstein & Sannerholm, 2022). Automation of governance does not necessarily always complement conventional norms and ways of working, as Sheila Jasanoff (2016) has argued; on occasion, it challenges the established norms and democratic framework needed for good governance. Hence, it requires an ethical and moral intervention, supported by legal safeguards,[1] to be able to adequately uphold the established universal values and principles in this modern digital age (Jasanoff, 2016).

The existing body of research takes a macro perspective approach, although more has been focused on developed, democratic nations. Nevertheless, previous studies show how such AI-related technologies are fast evolving and related practices are changing at a rapid pace, including in other parts of the world (Leal Filho et al., 2022). Comparatively fewer insights and in-depth studies are available on holistic perspectives of understanding social injustice and on-the-ground real impact on society. As Minevich (2020) asserts, this has been affected by the unilateral and largely unchecked AI practices of private actors (including large multinational tech giants) and constant violations by the state agencies of various countries that include breaching the fundamental rights of their citizens and the increasing sway exercises on social justice across the continent and the world at large (Minevich, 2020).

[1] S. Jasanoff, The Ethics of Invention: Technology and the Human Future. Norton & Company. 2016.

1.2 Statement of Problem & Significance of the Investigation

A recent study has found that an overwhelming number of European citizens strongly believe that AI-related technologies should be instrumental for the protection of the fundamental rights of society (Ufert, 2020). Given the rapid development and expansion of AI-based technologies across the world especially in recent decades – from medical breakthroughs to enhanced and efficient travel and transport systems, to intelligent machine-based learning and predictions of future probabilities and eventualities etc – there is no doubt that AI can bring immense benefits to mankind. However, unchecked expansion and deployment of such AI-based systems paint a grim picture, which includes compromising on some of the core universal principles and values. This results in privacy violations and, more importantly, the breach of the fundamental rights of citizens in a non-transparent and unaccountable manner that is increasingly becoming problematic (Bartneck, 2021).

This goes against not just the basic established social norms, but also against the rule of law particularly in a democratic setting. This highlights and signifies the importance of this research inquiry. From a regulatory standpoint, however, balancing society's competing priorities has always been a challenging task. On the one hand, nurturing innovation and preserving the commercial interest of the technology industry, whilst on the other, defending the fundamental rights of the citizens. Legislators and regulators have grappled with finding the right balance for artificial intelligence for decades. This is particularly difficult given the strong global presence of AI across diverse industries and commercial interests, especially of influential tech giants, represented by political lobbyists in major capitals. This has become a far more challenging journey for policymakers seeking the right mix and balance. Therefore, as Fonseka (2017) argues, authorities should be held accountable for the manner in which AI has been able to operate and its consequences (Fonseka, 2017). For this reason, the specific focus area of study would be not only professionals and the academic community, but also the corporate sector across the globe and the general public.

1.3 Purpose

Aim & objectives

Protection of fundamental rights of the citizens whilst it's a state responsibility, it becomes even more imperative in this digital age spearheaded by technologies such as artificial intelligence. Therefore, it becomes a collective responsibility of the wider stakeholders including the legislators, policymakers, regulators, and the private sector, most importantly with the wider participation of the citizens. Consequently, it becomes imperative for such technology breakthroughs to be kept under check and balance, whilst proactively facilitating to nurture of technological innovations (Teich, 2020). This would help encourage new inventions to follow through in the right direction for the broader good of the global society and humanity by and large. As one could reasonably argue, this could be in the best interest of all stakeholders in concern including the most vulnerable in our society. In the recent past, there has been an increased focus on un-checked AI-based technology expansion and deployments (including automated governance) which has an impact on rule of law and its consequences that includes social injustice. Accordingly, this research paper critically examines the risk associated with the rapid expansion and largely unchecked deployment of AI-based tools such as digital surveillance and face-recognition techniques, and the broader influence it has on policing, law enforcement and the impact it has on rule of law in a democratic society. This provides the basis to investigate the need for regulatory intervention and legal safeguards in today's modern digital age.

According to the literature review on this field of study, it's an emerging field of study in rapidly changing dynamics for many reasons. Firstly, due to the reasons i.e. technologies are emerging and fast evolving. Hence many new and novel topics keep adding to the sphere of a reasonable person's understanding of what AI-driven technologies are, and are capable of. Secondly, it's actual or scoped out capacity in terms of various deployments prior to assessing its real impact and consequences to society. Due to these evolving and shifting challenges, we need to carefully consider and comprehend as we go! Thirdly, the social impact that arises as a consequence of AI-based tools and automated decision-making (ADM) widely remains either unaccounted, partially accounted for or unreported (Goodman, 2017). This is largely due to various factors including commercial reasons (from developers' point of view) and awareness factors (from participants' or potential victims' standpoint).

As we often see and hear how '*new technology breakthroughs*' are connected to innovations that aid and complement humanity's progress. However far less is explored on its drawbacks or those specific adverse effects in concern that would equally have a significant impact on society. Such emerging trends of AI-related modern practices and their wealth of benefits to mankind have been overshadowed in the recent past by modern controversies and deeper AI-related issues (The Economist, 2018). This has been largely due to unchecked AI-related industry practices and increasingly unaccounted expansion, including unreported glitches and anomalies. Therefore, should such overwhelming benefits overshadow the consequences in the interest of 'commercial progress'? If so, at what cost? What would be the social impact and the hidden consequences that do not reach the mainstream surface? As a result, what are the social costs that create unreported / and un-documented cases of social injustice?

Research questions

The following research questions connected to the aims and objectives of this research inquiry are formulated through the previous literature review of relevant and applicable research work done. Accordingly, to be able to investigate further on this, the following key research questions were formulated:

(1) What are the implications associated with the usage and expansion of AI-based tools and applications, and the impact it has on rule of law and the society in a democratic setting?

(2) Accordingly, is there a case for more regulatory safeguards?

The above research questions stem from global digital surveillance and AI-based transformation drive. Acknowledging the discourses and relevant interpretations around 'digital surveillance' and 'automated decision making' (ADM) reflecting on realities. Accordingly, the theorization of the concepts such as rule of law, privacy, transparency, accountability and surveillance capitalism in this modern digital context is key. In fact, each of these principles complements each other and feeds well into the need for complying with the core values of a democratic society. Therefore, this research inquiry not only contributes to the critical evaluation of the risks associated with digital surveillance, and automation of governance but also endeavours to address the necessity to continuously promote, nurture and comply with the core values of democratic societies. These practices have been largely outsourced in the modern context, both in the EU and in the American context. Accordingly, it would critically examine the impact of such factors on rule of law and social justice, and the inalienable responsibility and accountability towards the citizens and the general public at large.

Limitations of the study

The multifaceted nature of the platforms used by AI and the rapid pace of technological evolution add up to the complexities. This in fact blurs the lines to be able to fully appreciate and comprehend not only the risks associated with the same but also the true potential and capacity of AI-based tools and techniques. However, taking these factors into consideration, this research paper narrowly focuses on specific aspects of AI technology. Therefore, it is limited to certain confined aspects of digital surveillance and automation of governance. Whilst the focus is on social injustice, the above essentially limits the generalizability of the findings however achieves to provides a reasonable and meaningful perspective and insights. Accordingly, the focus group meeting and the interviews conducted were only limited to certain AI-related practitioners and academics, plus focus on a few participants through mutual contacts whom (victims) have been referred with the author's prior knowledge of their potential vulnerability within the society.

02

Literature Review

This research inquiry rests on the ontological understanding that AI-based automation and related concepts are a powerful advancement of science which continuously evolves and (re)produced by certain discourses. That brings about significant technological breakthroughs for the benefit of society and the broader good of mankind. The particular conception however enables destabilising 'taken for granted' nature and the characteristics of such AI systems itself and the general industry practices, and more importantly the perceptions they are grounded in. Such perceptions are deeply rooted in the following conceptual approach that forms the overall theoretical foundation of social constructivism that helps this research inquiry to enable the representations of key concepts in this research paper.

This is primarily in line with Tamanaha's theorization principle of the Rule of Law[2], Margetts's conceptual understanding of transparency (2011)[3], and the re-assessed view of the concept of accountability and how it's looked at in this digital era as per Koene (2019)[4] are some of the crucial factors critically evaluating this topic in this modern context (Tamanaha, 2004). Furthermore, this would also inspire and appeal for an overall analysis to be able to evaluate the merits and demerits of ADM and critically assess the impact of such AI-based systems on social justice and the consequences that it has on society by and large.

2.1 Previous Research

Accordingly, these concepts and theories primarily form the theoretical foundation, including the understanding of artificial intelligence (AI), and the increased need for better transparency and accountability whilst addressing privacy-related concerns around it. Furthermore, this would also enable us to better comprehend how such concepts can either complement or at times potentially conflict (which would require the balancing act of the competing interests) and to be able to fully appreciate how it impacts the rule of law, that potentially contributes towards social injustices as a consequence.

[2] Tamanaha, B. (2004). On the Rule of Law: History, Politics, Theory. Cambridge University Press.

[3] Margetts, H. (2011). The internet and transparency. The Political Quarterly, 82(4), 518–521.

[4] Koene, A., Clifton, C., Hatada, Y., Webb, H., & Richardson, R. (2019). *A governance framework for algorithmic accountability and transparency* (Study No. PE 624.262) Panel for the Future of Science and Technology, Scientific Foresight Unit (STOA), European Parliamentary Research Service.

2.1.1 Artificial Intelligence & Automated Decision Making

As per John McCarthy (2012) at Stanford University, Artificial Intelligence (AI) is the '*science and engineering of making intelligent machines, especially intelligent computer programs'* which is a much broader and generic interpretation in today's context (McCarthy, 2012). However, as per Advancement of Artificial Intelligence (AAA, 2022), it is '*the scientific understanding of the mechanisms underlying thought and intelligent behaviour and their embodiment in machines*'. As per EU (European Union) Commission Communication 2018[5] – which refers to AI and defines in a more specific manner as "*systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals*." (European Commission, 2018).

Although there is no single, unified or universal definition of AI, as an evolving and ever-rapidly changing subject matter, as per Article 3 of the proposed regulation by the European Commission, reflects in the European Strategy for Artificial Intelligence by the European Commission in April 2021, as '*a software that is developed with one or more techniques and approaches listed in Annex I, and can, for a set of human-defined objectives, generate outputs such as content and predictions, recommendations or decisions influencing the environments they interact with'* European Commission. (2021, April). As per the American *AI Now Institute* report by Fritsch & Thomas (2019), artificial intelligence essentially utilises a wide range of technology tools and techniques, including robotics and machine learning capabilities (Fritsch & Thomas, 2019). This also includes what has been extended to more advanced technologies i.e. facial and voice recognition and natural language processing techniques etc. For example, from a technical standpoint, operated through ICT (information and communication technology) based systems. Accordingly, ADM is defined as those with systems of mathematical logic (algorithms) that perform certain given tasks through such systems, for certain decision-making purposes. Accordingly, such systems make decisions autonomously with limited or no human intervention. Although digitalization and AI go hand-in-hand, in the modern context of algorithmic advancement, it is essential to distinguish each function and role to be able to better understand automation and ADM in particular.

[5] EU (European Union) Commission Communication 2018 Reference

Digitalization, on one hand, is a digitised form of well-defined formats with structured data used for predictable situations for general solutions whilst AI, on the other hand, addresses hard-to-define problems including addressing eventualities and unanticipated situations that includes un-structured / mass-collected data that is used for different solutions. Critically reviewing the following examples from different social-economic perspectives helps better comprehend the deep influence AI has on our daily lives. As per Sarah Pink, (2022) algorithms, ADM and AI, although they are defined by many in their contextual forms, and in its fields of practice, the algorithm for example is a descriptive and formalised set of codes which deliberates its expression within a set of practice based on a particular technology (Pink, 2022).

From a sociology standpoint, referring to the same is simply a set of formal instructions defined by professionals (i.e. IT engineers) to carry out certain tasks to resolve certain specific problems. However, the most imperative aspect of this as Pink argues is the nature of emerging applications. AI and automated governance are more likely to change their capabilities with the evolving technologies, hence a shift in outlook is needed to keep up with the time. The fact that automation is already part of our daily life, the very reasons Pink has argued why '*hype, hope and anxiety*' surrounding the consequences of automation in daily life often revolves around the future of AI (Pink, 2022). Whilst there's positive reports of great value and benefits from AI-based automation of governance, concerns are however often growing and marred by failures and outrages surrounding the dynamics behind the ADM process in particular. This is often perceived as '*inferior*' or '*untrustworthy*' compared to traditional human decision-making, such developments have triggered modern-day discourses around this topic on the *level of trustworthiness of AI* systems (Gardner, et al. 2022).

Grappled with this fast-evolving technology-based topic, even policymakers, regulators, researchers and scholars, and practitioners (data scientists) alike, at times mis-align with designers, developers and programmers when they debate on emerging AI-based digital technologies. For the same reasons, there are now established various global research centres, think tanks and interest groups for example paying special attention to the design process, and development stages and provide swift feedback and criticism. This includes the private actors, corporate sector, and state-funded entities, that show great interest (and perhaps concerns to some degree realising the consequences of 'unchecked' development) with regard to the present wave of a digital revolution involving AI and AI-based deployment. For these reasons, 'everyday' automation is now increasingly becoming at the centre of attention and subject to scrutiny (European Parliamentary, 2020).

Accordingly, the growing momentum supports the assertion that fully appreciating the present-day and future AI-based automation requires '*re-humanising automation'*[6] i.e. people-oriented approach (Pink, 2022). It is a fact that neither AI nor automated systems can subsist fully autonomously or completely independent from human involvement or intervention, which is 'entangled' as Pink (2022) argues, within human-made intuitions, cultural context and social relations. For example, corporate recruitment practices, decisions on compensations and social benefits, social media platform practices to even critical aspects of human life such as healthcare diagnostic or credit ratings on credit worthiness, all such processing are processed in some shape or form involved in AI-based ADM (Pink, 2022).

2.1.2 Concept of Profiling and Predictive Policing

As legal protection afforded to European citizens, non-discrimination is a well-defined aspect in the European legal framework that particularly prohibits any form of discrimination which corresponds to the principle of equality. A range of legal instruments affords this from the national level, international and supranational level, which has been given reference to a fundamental right and human right emphasising based on i.e. race, ethnicity, religion, age or sexual orientation as forbidden grounds of discrimination. As asserted by Hildebrandt (2016), fundamental rights and equal treatment does not necessarily mean people should always be treated as if they are equal, instead to the contrary, in fact accepting the reality which warrants different treatment to counterweight for 'unfair disadvantaged'. She argued the concept of equal respect that is grounded in both democratic values and the principles of rule of law (Hildebrandt, 2016).

Profiling is controversial yet a practice that's increasingly been used by authorities in recent times. A form of automated processing of personal information involving big data. As a result, people are being constantly subject to digital surveillance, i.e. with the usage of high-tech tools aided by big tech companies (and often state actors. including state agencies) who are responsible for the collection of such data, which is often carried out with or without the consent of the subject matter(s) which in turn, assess and infer things about us constantly (Walsh, 2021).

[6] *See, a term used by Sara Pink*. Pink, S. (2022). *Everyday Automation*. Routledge.

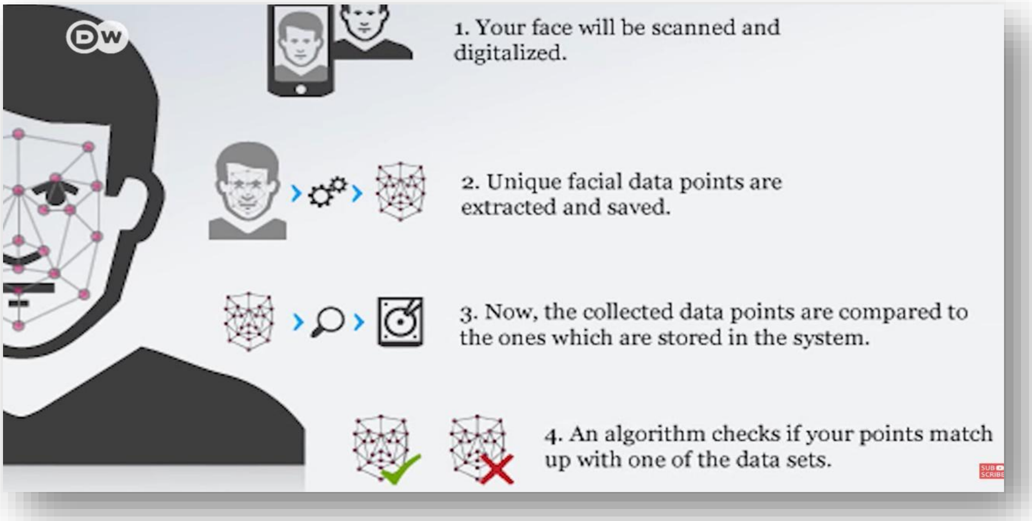
Such collected and assessed data is often shared with various other bodies (including public and private actors) for various reasons. The most common and disclosed purpose includes commercial reasons i.e. but not limited to 'advertising' purposes. This is often done to evaluate certain aspects based on target individuals' personalities, behaviour, specific interests and certain user habits in order to make predictions about them or their future prospective behaviour. State agencies and commercial organizations obtain individuals' personal data from various sources i.e. from national records, the internet, mobile phones or mobile operators, social networks or surveillance (video) systems or the internet of things by and large. Such data collection involves grouping or sectoring for the purposes of i.e. includes for finding something specific about individual preferences, whilst predicting their behaviour and accordingly making decisions about such target audience (Dave, 2018).

Such a process of profiling can involve AI-based techniques using algorithms and big data (a term used for mass-scale data processing) that often characterises our identities. This is done by inferring the information that it possesses based on the amount of data collected at a particular given time. It can be accurate, partially accurate or not entirely accurate at all. Hence not guaranteed as they are combinations of facts and predictions combined. This is rather evident, reviewing some of the recent cases[7] and leading legal battles (Waddell, 2016) involving some of the big tech companies including Uber Technologies Inc (Levin, 2017). This shows users' information that has been shared isn't entirely voluntary but also includes those shared passively, unavoidable and often unintentional. This includes user behaviours such as clicking, typing or simply browsing behaviour that includes eye/ or mouse tracking plus third-party data i.e, friends and family recommendations. This includes their perception of all that is been part of the inferring to characterise your own digital identity, which may or may not reasonably characterise one own self with or without their consent or awareness.

Profiling from a technical standpoint is a tool comprising a sequence of instructions provided through keywords, for specific search purposes. This uses algorithms to find matches and correlations between separate data sets. Profiling, also known as '*social sorting',* as Hildebrandt argues is forbidden by law that violates fundamental rights. Such technologies are used for numerous purposes including for a range of decision-making purposes i.e. predicting the future behaviour or purchasing trends of certain users in a given certain way. For this purpose, in the modern day, is increasingly using AI systems and machine learning technologies to create and generate such algorithms. Such profiling practices are widely used by state agencies aided by

[7] Waddell, K. (2016). *How Algorithms Can Bring Down Minorities' Credit Scores.* The Atlantic.

tech corporations in health care systems in particular for various purposes including maintaining medical records, for effective diagnostic purposes etc (Hildebrandt, 2016). Another such instance is from the commercial world, where for example, social media posts are utilised to analyse user's personal behavioural patterns including behaviour drivers. By deploying such algorithms to suggest possible predictions/ or judgements to arrive at certain calculated (automated) decisions about an individual i.e. 'safe' or 'unsafe' call. In order to independently assess the level of risks to such individuals to be able to assess for example i.e. insurance premiums accordingly. In essence, not just unlocking the iPhone or wondering how Facebook tagged us in a particular photo, face recognition technology goes way beyond that. From i.e. airport security screening to welfare benefits, to housing and employment decisions, and also for mass-scale law enforcement surveillance purposes as well. This is a significant and overarching AI-based tool that goes above and beyond the capacity to identify or verify persons through digital images or a video source.



1. Your face will be scanned and digitalized.

2. Unique facial data points are extracted and saved.

3. Now, the collected data points are compared to the ones which are stored in the system.

4. An algorithm checks if your points match up with one of the data sets.

***Figure 01** Courtesy of German public broadcaster DW Shift – impact on facial recognition practices*[8]

The obvious benefits that are evident, for example, to prevent and solve crimes. However, from a privacy and safety perspective, privacy advocates have long been raising their concerns against this technology deployment. However, the world has caught its attention increasingly with biases and racism concerns, the controversies around facial recognition systems deployed by state agencies and powerful private actors driven by large tech companies operating at the multinational level.

[8] DW Shift (2019) *impact on facial recognition practices*, German public broadcaster DW Shift. Reference

As per official figures from a Harvard Study in 2016[9], it is estimated that half of the American adults' images were on agency databases. Where police use facial recognition-powered automated systems to compare such images of 'suspects' to *mugshots* including with for example driver's licence images. This process transpires without awareness or consent of the subject matter, and above all with lack of legislative oversight (Najibi, 2020). More distressing is the fact that such technological deployments are marred with serious racial biases, particularly against, for example, black Americans and ethnic minorities. This can be a combination of reasons as argued by Najibi (2020) that fuels this situation. Such a situation includes a result of the lack of adequate amount of resources used (concerning images and related data) from such minority groups of people whilst have simply presumed such groups of people with ethnic features including black people and minorities are 'suspected' to be more 'prone' to violence and crime. For instance, the law enforcement systems have a well-established history of racial and anti-activist surveillance.

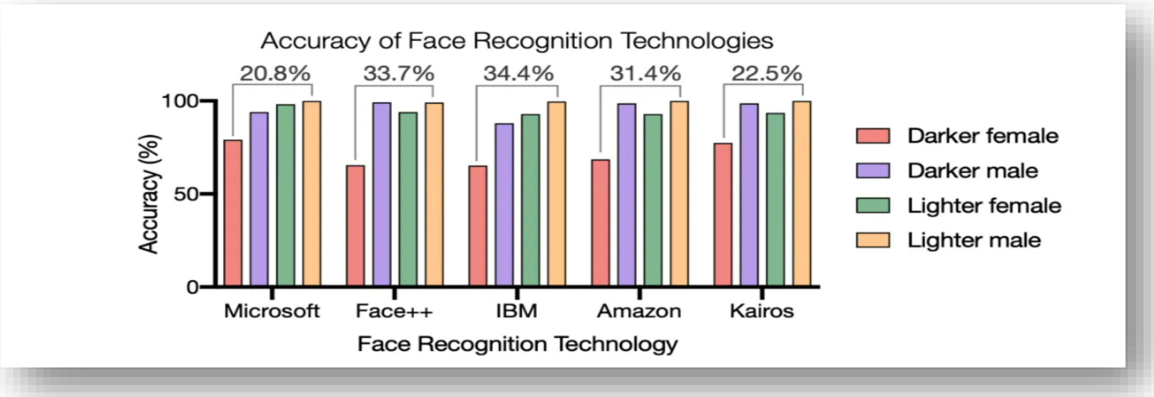


Figure 02: *Courtesy of NIST study findings published by Harvard University*

The databases are fed by images and videos that include but are not limited to those from mobile phone data, ATM machines, in-store cameras and from cameras installed around people's homes and from public surveillance cameras. Which is regarded as a '*jail-mugshot'* database by the police and law enforcement agencies. Accordingly, from the point of view of accuracy, the above illustrates the critical issue. The issue of inequality in facial recognition (programming) algorithms, although it boasts of precision of over 90%, such outcomes as argued by Najibi (2020) are neither universally consistent nor does it provide a complete picture.

[9] Najibi, A. (2020). Racial Discrimination in Face Recognition Technology. Harvard University press

In many areas of our lives, the technological advancements around us are changing our lives and the way we conduct our daily affairs. Government agencies including law enforcement agencies and social and criminal justice authorities use the vast amount of data they hold on to people they come into contact with. For example, stop and search data, crime reports, includes arrest data and other data held by public authorities and private (actors) contractors i.e. that includes information from welfare and benefits authorities and health services, financial and credit information etc., (Lee, 2021). These data have been utilised for artificial intelligence and ADM. This evolving tendency including a transformation for automation is often propelled by financial pressure for greater efficiency and the misguided perception about the impartiality of such systems and solutions. Hence as a consequence, such evolving systems have increasingly been deployed in social and criminal justice processes across Europe and largely in North America including in the United States and Canada. This includes the purpose of profiling and predicting their future by so-called ‘supposed behaviour’ and assessing their risk of offending from a criminality standpoint, through such automated predictions. Therefore, it warrants a critical review of how such AI-based systems are created (programmed) and operated (standards) from a governance standpoint.

Predictive policing

with the advancement of technology keeping pace with the new and novel ways of criminal activities in society, police departments and law enforcement agencies around the world, particularly in the developed world, have been testing predictive policing as a model of forecasting criminal activities (Lau, 2020). This predominantly deploys computer-based AI systems to analyse colossal amounts of data, including criminal record and crime data to aid in assessing and decide in identifying individuals or groups of people as Lau (2020) argues, who are supposedly more likely to commit a crime or to be a victim of crime, and to assess when and where to deploy police resources accordingly (Lau, 2020).

Whilst supporters of such AI-based systems claim that such advanced systems can aid predict crimes beforehand in advance more accurately and effectively compared to traditional means of entirely relying on human capacity. Big tech companies have claimed such processes can take out police ‘biases’ in their conduct and discharge their duties, whilst opponents have long argued, such AI-based technologies with algorithms that depend on historical data, in fact, risk reigniting the very biases that it was supposed to eliminate. This has been followed up and supplemented by surveillance technologies i.e. social media tracking and facial recognition systems. Predictive policing technology – is an evolving trend in North America in particular. Especially within law enforcement agencies, where the deployment of such technology is increasingly justified as much as criminal activities get sophisticated.

As per National Institute for Justice in the United States, predictive policing (that is opposed to traditional policing methods), attempts to control the power of information and evidence-based intervention models, with a view to improving public safety by reducing the crime rate. By applying advanced data analytics techniques for captured data sets through i.e. facial and voice recognition tools for example. This transforms from a simple reaction to crimes (when it happens) to more of an agile sphere of predicting what crimes (nature of acts) can take place and where (geographical location) are more likely to occur. To be able to deploy resources proactively to prevent such occurrences even before they could take place. Such community-led 'intelligence policing' and 'hot spot' policing as it's known, one could argue these are good progress benefiting mankind. As such breakthrough developments do bring benefits to society by strengthening law and order which, arguably manage to reduce crime rates in society although might not be able to eliminate (Alikhademi, 2021).

But at the same time, such predictive policing has also been heavily criticised for its controversial nature of the technology and practices more so in the manner in which it is used (Kent,2020). Particularly looking at the data sets that it consumes for its analytical purposes. Which entails controversy that is typically associated with surveillance of the population connected to certain socio-economic conditions or racially biased mechanisms. Such systems generally focus on predicting i.e. what time of crimes (nature) is more likely to occur (probability) in which part of the city (geographical location)[10] and who – '*nature of the person(s) is likely to commit a crime*' are made for profiling purposes. This is the foremost controversy out of all. According to the Law Commission of Ontario (LCO), Canada – such algorithms are reportedly being deployed in over 60 jurisdictions across North America (including the US and Canada). How science and technology are embedded in a social context describing conceptually, emerging technology such as AI tools – with reference to i.e. criminal justice system, education and healthcare systems for example are outsourcing human decision making, which has turned into AI-based 'risk assessment' tools which are allegedly discriminatory in nature (LCO-CDO, 2020).

[10] *The factors such as nature of the crime, nature of person(s), probability and the location its likely to occur – are the generally coordinated for predictive data analytics purposes*

Looking at history and society everyone is equal and has inalienable rights although the current usage of technology contradicts to such values. Benjamin, (2019) argues, it’s not to resist such advancement of technology but instead *‘how we utilise such advance tools and technologies that matter’*. Her great work and endeavours have largely helped a grand shift in narratives that would empower those who are marginalised in society by such advanced systems and technologies (Benjamin, 2019). The competency of the police department and law enforcement agency is similarly being reflected in the justice system via the implementation of so-called predictive prosecution technologies. Predictive prosecution is categorised into two areas: predictive bail and predictive sentencing. As asserted by Prof. Ferguson, this involves the identification process that targets suspects considered to be the most at risk for future serious criminal activities and utilises this information for bail determinations, and related decision-making, including sentencing (Fritsch and Thomas, 2019).

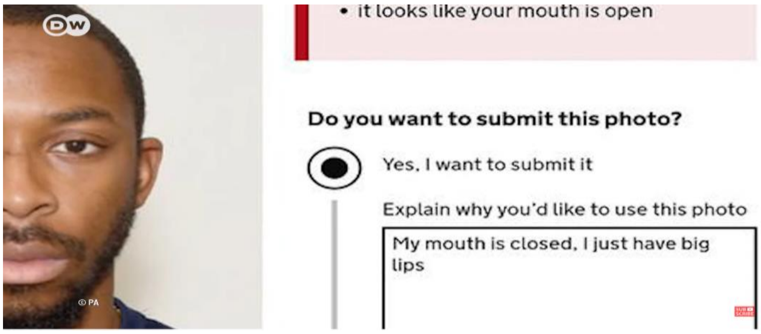
2.2 Artificial intelligence and the Impact on the Rule of Law

Fast-evolving technology includes digital surveillance and AI-based system automation, which typically heralds social and legal discourse. Compared to shifting society’s needs, legislative changes often take place in a reactive manner and are seldom done proactively. Understanding the broader concept of rule of law in this modern-day context where technology evolves at a rapid pace, is key to realising the broader challenges it poses by the mostly unchecked and largely unregulated automated governance including AI-based ADM. The age-old concept of rule of law and the principle of accountability in particular have been instrumental in constantly challenging and managing the relationship between individuals and the manner in which such evolving technologies have been regulated. This has also helped in limiting the arbitrary power of the state, which often tends to be abused in the absence of respecting such values if not kept under scrutiny at least in a democratic setting (Tamanaha, 2004).

2.3 Case Review:

2.3.1 UK Border Control: *The case of Joshua Bada, from the UK*

The British case (BBC 1, 2019) that involves UK Authorities. In September 2019, it was reported that the UK migration agency has rejected an application of this coloured person when he applied for a new passport under the UK immigration system. The reason for this as reported, where the system has taken his lips for 'an open mouth'! The formal application was rejected whilst claiming that he has to submit a 'neutral' photograph with his mouth closed, for something that he has simply not violated on migration authority rules! In essence, the algorithms (as programmed) couldn't in fact interpret his natural lips accurately. It was reported that the systems were not programmed or fed sufficiently enough with images from the black community.


- it looks like your mouth is open

Do you want to submit this photo?

● Yes, I want to submit it

Explain why you'd like to use this photo

My mouth is closed, I just have big lips

***Figure 03** – case of Joshua Bada - Courtesy of German public broadcaster DW Shift*[11]

Lorena Jaume-Palasi[12], an international expert and advisor on Ethics and Technology have claimed, for example, why the US Tech giant *Apple Inc*. had to change its face recognition software application many times within recent years. Simply because of the challenges dealing with the source images. When it comes to white persons that are opposed to, for example, people of black or coloured communities, which has created a vacuum, resulting in major flaws and problems. However, with the facial recognition industry is growing exponentially, and is estimated[13] to be EUR nearly 2a billion which has a growth potential of over EUR 6 billion by 2024 just in the US Market alone.

[11] DW Shift (2019) *Impact on facial recognition practices*, German Public broadcaster (DW Shift). Reference :

[12] Lorena Jaume-Palasi - Reference:

[13] Facial recognition market by components - Reference:

The British case of Joshua Bada may sound like a trivial matter. However, Alex Najibi claimed when it comes to the Criminal justice process such flaws (intended or otherwise) can have very serious consequences for its victims. This was the main reason some of the handful of US cities have seen the prohibition of face recognition by police and other government agencies including in Boston and San Francisco. Out of the different forms of biometrics (i.e. fingerprint, voice, face), facial recognition is regarded as the least accurate and reliable and is in conflict with serious privacy concerns (Najibi, 2020).

2.3.2 The case of Amazon *('Rekognition')*

Case study of Amazon's Facial Recognition software called the '*Rekognition'* used by law enforcement agencies in the United States for the purpose of monitoring and tracking down criminal suspects in many cities. This facial recognition development is undoubtedly one of the biggest technological ambitions of tech companies and the AI-based industry. That indeed includes the American tech giants i.e. Google, Facebook, Microsoft and Amazon, to name a few. The US multinational and Tech giant Amazon Inc. when realised the future potential of this tool from a corporate standpoint, the company hadn't wasted any time getting into it investing in its research and development of the face recognition software – *Rekognition*. The question concerns the connection between (for profit) private actors the likes of Amazon Inc – where the main motivation is to maximise corporate profits. This is opposed to non-for-profit organizations such as the state police departments and law enforcement agencies, whose primary intent and mandate is the service for the purpose of public safety. Companies such as Amazon are incredibly powerful in terms of financial strength and influencing power, not just in the US but also around the world in the territories they operate in.

The US government with its executive mandate to make decisions on behalf of the people, at the same time, companies like Amazon are incentivised to get this AI-based systems to as many places (private entities and public authorities regardless) as possible. This alliance (i.e. public-private partnership), as fundamental rights activists and many privacy advocates, argue, puts the general public at disadvantage against this 'powerful partnership' (Harwell,2019). For this reason, Amazon, for example, amongst other lucrative deals, have collaborated with Police departments in many cities across the U.S. and managed to have an agreement along with the US border control and migration agency as well for example, for the deployment of this AI-based tool regardless of the growing concern over its adverse impact on the society. Some of the adverse impacts and indicative challenges are i.e. – the growing trends and usage of big data and related data collection from users' routine and everyday lives plus those data connected to biometrics data. Such data is often disclosed to be used for advertising purposes. However, with the rise of mass surveillance, society often utilised '*insights of behavioural influence'* through 'data colonialism' as its argued to be the 'exploitation of people' through such data (Foresight, 2020). It is fact that such technologies and tools are used and are beneficial in fighting crimes with the usage of smart devices i.e. GPS functionality, face recognition technology, tracking cookies etc. However, such technologies are deployed and for example, for providing welfare benefits, evaluating job applications, assessing loan applications and credit risks and worthiness, customising social media feeds, or analysing 'good driving behaviour'. Such developments have created new norms and an economic and social new order that's been established globally beyond borders, which has a greater impact on citizens' autonomy (Foresight, 2020).

In the recent case of Google in the U.S., (Wakabayashi & Metz, 2022) where the tech giant ravelled with controversy for firing a couple of high-profile employees. This includes researchers who were connected to a publication criticising the production of AI-based computer chips and AI systems called *Google Brain* with in-built biases. Dr. Chatterjee (researcher) and Dr Gebru, a former Leader of the Google Ethics team, sought permission to publish certain research findings, explaining how google had developed AI-based tools including language systems, that possibly end up with bias and using hateful language (part of machined learned techniques). However, they eventually ended up being fired by Google whilst the parent company Alphabet Inc refused to publish such findings.

Evidently shows beyond a shadow of any doubt, that such private actors (tech corporations) neither care about admitting such flaws in its AI systems nor rectifying such practices as long as it keeps continuing to maximise profits at the expense of society. This revelation is the tip of the iceberg in the latest series of events connected to AI-related systems. Whilst the senior leadership of Google has compared AI technology to the "arrival of fire or electricity to mankind"! (Thomson & Bodoni, 2020).

The recent controversy overshadows a more customary pattern of dismissals connected to high-profile claims of misconduct among Google's AI-related researchers. This a rising concern for a tech giant like Google that has gambled its future on inspiring artificial intelligence into the everyday business aspects of its operations. This highlights the growing tensions between researchers on one hand (who take a conscience decision to fight against such corporate practices), and the tech companies on the other hand, with such unsustainable practices connected to social injustice. This completely disregards human rights, human dignity, and the struggle across the industry, that is faced by many such tech giants, dealing with the same social issues.

From a legislative and policy planning standpoint, Dutch Policy planning expert Maritia Shakaits explains, that it is hard to have the presumption that "*because your corporate intentions are good, that the outcomes would always be good*" – giving reference to the commonly provided '*good intention*' claims by the large tech firms. In fact, such big tech giants are more powerful, so much so that – even stronger than independent nations or their governments around the world. The recent Australian case of Facebook and the controversy around hosting news channels on its own platforms (Lovelace, 2021), is a classic example and a testament to tech giants' dominant position in the world. Their colossal financial strength and leverage in politics and among European politicians, (BBC News 2, 2018) often give them an edge over efforts to regulate their dominant position in the industry (Fox, 2010).

This was a case where Facebook blocked content from the Australian media and it effectively deprived Australian users of sharing news content, they chose to do on the Facebook platform. This was a clear evidential case of conflict between the interest of multinational corporations - social media giants, and the Australian government representing the general public. The political and legal consequences were much at focus at the expense of social injustice that takes place from the Australian society's standpoint. This unilateral move by the tech giant effectively deprived access to post links to news articles on Facebook, as all such postings have been taken down by the *Facebook Australia page* connecting to International Media organisations.

The timing was significant to the case where this occurred just days before the Australian covid vaccination rollout begins. Which raised serious concerns in terms of dealing with misinformation on the platform, whilst ironically the tech giant claimed to work on 'misinformation' on its platform, which exposed its alleged hypocrisy in dealing with the matter in the public interest. Nonetheless from a socio-legal standpoint, the most significant part was yet to unravel.

From a policy planning, advancing democratic values, and in the interest of fairness and equality standpoint, the European Union has taken significant initiatives in the recent past. Some of the legislative initiatives and regulatory steps (Chee, 2020) taken include two new directives i.e. **Digital Service Act** (DSA)[14], which envisioned and ensures entrust platforms like Facebook and Twitter to take on more serious responsibility over protecting fundamental rights, provide more transparency around advertising on its platform and managing and deleting illegal content on its platform.

Furthermore, from a fair market and best business practice standpoint, **Digital Markets Act** (DMA)[15], attempts to regulate fair competition policies allowing and encouraging more competition within European markets (given the allegation that tech giants monopolise this industry) Regulations in terms of data sharing, and companies like Amazon, Google, Facebook and Apple are intended to be the main targets in terms of compliance. However, could this be sustainable? The reality is that such large-scale, trans-atlantic multinational and powerful tech corporations fighting back to defend their non-transparent, anti-competitive and unaccountable practices in Europe (Riekeles, 2022) through various means including using powerful lobbyists, which predominantly represent not the people on the ground but in the interests of its own powerful tech companies.

Figure 05 *- Courtesy of German broadcaster DW (May, 2022)*

From a privacy standpoint, it is learned as per Washington Post published report (Harwell, 2019), that even a defence attorney of a suspect or a privacy activist reach out to investigate and seek to discover and unfold from a transparency standpoint. It cannot be questioned on how this process worked, as the corporates would argue with the defence of '*proprietary content'* or claim protection from copyright laws etc.

[14] Digital Service Act (DSA). Reference

[15] Digital Markets Act (DMA) Reference

This makes it even harder to challenge such unknown and undisclosed practices of the tech companies, as there is no ability to contest the accuracy of such matches and identifications. This raises serious concerns about its credibility and reliability from a transparency and accountability standpoint.

It is argued by activists that flaws in the system can lead to some serious misjudgements by the police and law enforcement agencies including the prosecution. When it comes to identifying and prosecuting so-called ‘suspects’ it would be based on such inaccurate data and information that it provides. which can potentially lead to wrongful convictions that are damaging to society and its social structures based on fairness and equality. This further underlines the constant surveillance of the general public, where at a certain occasion, ethnic groups of people's images could be fed into such AI-based systems particularly to identify certain groups of people in such ethnicities because they are suspected to be more ‘prone’ to be violent than others! Therefore it is argued that it can be weaponised and misused against marginalised communities across the continents, which is often accused of being used, mis-used and abused by law enforcement agencies around the world (Amnesty International, 2022).

This is a dire situation, where accountability is even more questionable in a non-democratic society that is opposed to in the democratic world. For an instance, the Chinese government uses AI tools and facial recognition in particular to track down from simple i.e. jaywalkers and other petty crimes such as trespassing and traffic offences to as much as large scale suppression of minority communities in the country (Taddonio, 2022). This is where the argument becomes even stronger where protection of the society with more priority for their security (securitization of the society) versus respecting civil and fundamental rights of citizens including privacy. Consequently, this shows that facial recognition software carries a serious risk of misjudging and misidentifying some people or groups of people over others by and large.

2.3.3 **Dutch & UK Cases -** exposes the challenges posed to the vulnerable

In terms of managing the impact of automation of decision supported by risk modelling, for a better overview on the impact, breaking down with some facts and data on some of such systems used. As per *FairTrial*, citing the example from the Netherlands, a reported case in 2012, where the Amsterdam municipality explored the ‘*automated risk modelling and profiling*’[16] in

[16] AUTOMATING INJUSTICE: by Fair Trails, *The use of AI & ADM systems in Criminal Justice in EU* (2021) Reference:

collaboration with the Amsterdam police department and its social services, involved profiling of top 600 young people above the age of 16. Who are regarded as most likely to commit or risk the possibility of committing a crime in the future. The criteria '*secret algorithms*' (Böhre, 2019) set for this were by the police in collaboration with the prosecution services and the local authority, including but limited to i.e. if someone has been arrested as a suspect for a certain crime, or a suspect of a crime within the last 2 years, as well. Other factors such as including i.e. if they have presented to a bankruptcy judge from a financial situation perspective etc part of the risk model. The use of such suspicion and arrest process that has criminal justice implications has deeply and obviously concerning from a reasonable person's perspective. Although the Dutch authorities claim its intention is to 'quickly punish' and consistently through a "combination of punishment and care".

This includes as per Dutch civil rights lawyers who claim that the prosecution office also pursues longer pre-trial detention for those top 600 on the list. The impact and the consequences of such risk profiling by such automated decision systems thus, demonstrates a very dangerous trend whilst obscures the limitations and boundaries between so-called 'care' and the punishment. This fuels discriminatory outcomes of the system which is rather evident. In fact, more than a quarter of such profiled suspects were of foreign descending (i.e. for example from Morocco).

Another instance is that which is regarded as the 'Top400' that targets those under 16 years of age, profiled them on the basis if they have been arrested as a suspect for certain crimes, or suspected association of any gang activities. In addition to other criteria such as based on police intelligence / or police reports i.e. if there are absent from school or changed schools regularly or if they have been subject to any shape or form of police surveillance. In a different example from the UK, in England and Wales, the metropolitan police department had developed its own AI-based (machine learning) algorithms that profiles suspect to be able to forecast and predict and assess whether to be prosecuted based on the chances or risk of 're-offending' in the future providing them with so-called a 'risk-score' (Marsh, 2019). This AI-generated 'risk-score' is used for the purpose of advising whether or not to charge a suspect or even release them on a for example a rehabilitation program etc. The impacts and outcomes caused by such automated systems therefore have rather substantial consequences from a criminal justice process standpoint, where historical data is used for the predictions about their future actions and behaviour challenges those established legal principles and undermines the presumption of innocence.

At face value, this may sound like it potentially substitutes exercising of own subjective judgement by the police officers, which may sound progressive. That potentially eliminates the subjectivity element that can be marred by individual law enforcement officers' prejudices etc (if any). The real and grave concern, however, is much larger that overweighs the benefit that it brings to society. Critically reviewing the following hypothetical scenario as Benjamin (2019) argues, can put things into contextual perspective, consider the following example. For instance, if and when one individual makes ‘hateful’ comments but has in fact used beautiful and rosy words (linguistically) but of course with sarcasm. In such cases, AI might not have detected true to its spirit of the context as it is simply not geared for the same. On the other hand, one could be completely innocent and have used the English language but with a different dialect, and the chances of an algorithm absorbing that are rather high for its deliberations and conclusions (Benjamin, 2019), which is the practical issue in a nutshell.

03

Theoretical Framework

3.1 Concept of Rule of Law

In this day and age, where social needs are ever-changing, and fast shifting is supplemented by the rapid pace of technological advancements that impacts societies beyond traditional borders. In this context, it is imperative to understand and appreciate the concept and conceptions of law. As Hildebrandt (2016) claims philosophers of law and legal theorists would align that the law is fundamentally a contested concept, which tends to attract disagreement and opposition over the connotation and the role it plays in constructing human society. This is even more relevant from time to time with societal change and social construction which shape up the norms which are the shared standard of acceptable behaviour by society. In essence social normative influences social norms in a society which are instrumental in society’s behaviour. With reference to the normative position of what conception of law best suits the rule of law in a democratic society in this modern context, it becomes imperative to fully appreciate the concepts of law historically. In this inquiry, which has normative implications unavoidably, it is not a matter of individual taste as Hildebrandt argues but it means that it should be acknowledged and stand the ground for the normative position one you choose and ready to explain (Hildebrandt, 2016).

In essence, as Hildebrandt argues, legal positivism weighs that law is a system of general rules, that largely depends on the authority of the state and are separated from morality, although as H. L.A. Hart asserts they are closely associated but not necessarily related. Although this notion has been disputed by American legal theorist Ronald Dworkin, especially in the context of dealing with cases that he coined as resolving ‘*hard cases*’ by the judiciary, where the law may be silent. In this context, from the standpoint of justice, law and technology, as Hildebrandt claims, the notion of *pragmatic conceptions of law*, which is directly relevant to today’s global context, and the challenges posed by AI-related tools and ADM to the society, which warrants the constant questioning if the law continually be subject to moral scrutiny in this digital day and age (Hildebrandt, 2016).

As Hildebrandt (2016) asserts, the protection of privacy for example, which is an ongoing negotiation of boundaries between the private sphere, the social and the public space. Necessitating the notion of privacy, which doesn't take these boundaries for granted. At the same time, she argues (p.85) whilst in a constitutional democracy, it's the state's responsibility to protect the citizens and serve the public interests. Hence from a normative aspect of rule of law, where the citizens are provided with an opportunity to challenge and contest the state's claims over its actions in the public interest.

Furthermore, the Dutch lawyer argues that within such constitutional democracy (that opposed to a repressive regime in a communist country for example), privacy is perceived as a 'public good' so long as it protects the civil society. Which highlights the primacy of individual rights over the collective, which as she asserts '*thrown into an existing web of normative'* aspect i.e. possible constraint that forms the base of the collective. This point of aspect would be critically assessed below in line with AI-related 'digital sorting*'* of citizens or commonly known as *profiling* practices by the law enforcement agencies including in modern democracies. As she asserts, such practices go against the principles of rule of law in a democratic society. For example, she critically argues, around the biasness, citing the following stereotyping of human profiling examples, i.e. *"CEOs of large corporations are essentially shady or corrupt", "black people match with low income or criminal intent", "those who buy diapers are less likely to be alcoholics", "children with divorced parents needs special attention for obtaining better grading"* and such algorithms are seldom made visible for various reasons that raise serious transparency related concerns (Hildebrandt, 2016).

The rule of law however traditionally has served as a crucial component for legal compliance, by respecting the core principles of accountability, open government, just laws, and more importantly access to justice. This has played a significant role from an enforcement standpoint between the individuals, decision-makers and the social construction in a democratic society. The *Venice Commission* of the Council of Europe has constructed two specifications supporting the rule of law principle, and the most recent publication covers five core principles that include, legality and legal certainty, non-discrimination and equality before the law, and access to justice. According to the World Justice Project (WJP), which has its own methodology and approach to measuring of rule of law in a given state, which includes factors such as accountability, open government, impartial laws, unbiased regulations, and most importantly access to justice for dispute resolution[17]. Each of these principles would be unpacked and critically examined on how the rule of law plays a significant role in a democratic society as the normative aspect, in line with emerging AI-based technology deployments in the context of digital surveillance and automated decision-making (World Justice Project, 2021).

A pragmatic approach established by the EU as a measuring mechanism for rule of law from a normative aspect in modern democratic societies is illustrated in its published ***Rule of Law Report***[18] by the European Commission. This indicates the rule of law is fading in many EU member states in recent years. Although automation and digitalization is not one of the main causes for this trend, it is more likely that AI-based automation will be on the European Commission's focus in the future for the following reasons. In relation to large volumes of daily litigation that takes place in the areas of i.e. road traffic fines, insurance, taxation etc, automation has become far more important and a necessity for reasons such as legal certainty, which particularly mitigates the inherent risks associated with manual decision-making. As Greenstein & Sannerholm (2022) has argued, in their latest publication *'Responsibility and Accountability: AI, Governance, and the Rule of Law'* [19] that human intelligence is not only superfluous, in certain situation, it's not even required or desired.

Citing an example from the Swedish Government Agencies Ordinance**,** where it specifically indicates as a key rule in public administration, i.e. the need for decisions to be reported prior to arriving at a final decision. However, ADM in particular, typically impedes this requirement. Consequently, the legal adaptation that has resulted as a consequence of the Administrative Procedure Act (2018)**,** which effectively provides a lawful ground for ADM to operate satisfying the legal requirement. This move puts the legal profession in limbo, raising many in-depth questions on how the age-old concept of rule of law is understood in this modern context of the digital age. Therefore, the concern is not, should or should not the governance be automated instead, does this process of AI-based automation deserves a rule of law perspective? The above legal limbo involves questions not only strictly legal but also from a socio-legal standpoint. This warrants a critical review from a sociological standpoint, which is largely behavioural in nature. Whilst the legal questions focus more on compliance

[17] World Justice Project, Rule of Law Index 2020.

[18] Communication from the Commission to the European Parliament, the Council (2020), the European Economic and Social Committee and the Committee of the Regions, Rule of Law Report. The rule of law situation in the European Union. COM (2020) 580 final. 2020.

[19] Greenstein, L. C. A. S., & Sannerholm, R. (2022). *Responsibility and Accountability: AI, Governance, and the Rule of Law*. Law in the Era of Artificial Intelligence. eddy.se ab.

in terms of the identified risks associated with digitized or automated governance, (which largely depend on coding and AI-based algorithms), the socio-legal aspect is largely looked at from the point of view of the impact to society.

Rule of law generally requires more substance from an enforceability standpoint, which generally takes a more conventional approach. Therefore, from an individual safeguard perspective, not only that warrant legal backing but also should be complemented by the principles of transparency and accountability. Consequently, it could be argued that the concept of the Rule of Law is an empirical construction with a complex system beyond mere binary conditions. In simple terms, i.e. in its normative aspect, the concept of the Rule of Law in a democratic setting, needs to be respected to minimize the arbitrary power of the state, which can be cascaded down to many other scenarios which is critically looked at further in this report. Nonetheless, the theoretical foundation by Tamanaha (2004), suggests that the Rule of Law from a substantive and universal standpoint, incorporates individual rights that opposed to the state arbitrary power. In other words, the core values and general rights i.e. right to privacy, equality and non-discrimination, freedom and liberty i.e. freedom of speech and freedom from cruel punishments or torture etc are embedded to have a check and balance system in governance in a democratic setting.

However, as Tamanaha has argued[20] there can be conflicts of rights and interests. A living example, on one hand, is the right to privacy (which the citizens cherish so dearly), whilst on the other, the practice of securitization by the state, which is in line with the state's responsibility towards keeping its citizens safe. It has been asserted however that such conflicts of interest or competing priorities cannot be resolved completely merely through consultation of the rights alone. AI-based systems particularly those regarded as high risks, as per the European strategy for artificial intelligence, a report[21] published by the European Commission in April 2021. This highlights the controversies impacting society at different levels (European Commission, 2021). In a setting where decisions are made by automation aided by AI-based systems, with less or no human involvement or intervention. Prima-facie there are reasonable and justifiable concerns i.e. can society have faith and trust on such systems? Are such systems transparent enough to build that trust? Have the authorities or private actors involved taken reasonable steps to restore and re-establish the diminished faith in such systems?

[20] Tamanaha, B. (2004). *On the Rule of Law: History, Politics, Theory*. Cambridge University Press.

[21] European Commission. (2021). A European Strategy for Artificial Intelligence. Reference

Although such systems are supposedly embedded with transparency, fairness, accountability and ethical considerations on paper, there aren't universally accepted standards applied when it comes to AI-based development and deployments. It is generally believed as such, can have an impact on global society's fundamental human rights thereby affecting social justice. Therefore, it is critically examined the social and policy-related challenges in this regard that may potentially amount to fundamental rights violations whilst potentially impacting criminal justice, which potentially feeds into social injustices. Understanding what exactly AI-based ADM is, and its core relations and interconnection between algorithms and AI-based systems is key to understanding and fully appreciating automation of governance and its impact on policing, law enforcement and the concept of rule of law (Foresight, 2021).

For all these reasons, individual rights from the citizens' standpoint, inevitably have anti-democratic implications as argued by Tamanaha (2004). Where every Western liberal democracy in the modern day has wrestled to find the right balance. On one hand, respecting individual privacy, and on the other hand, increasingly taking steps towards securitization of the society in the name of 'keeping the citizens safer'. Which obviously imposes enhanced limits on democracy, freedom and liberty, even impacting the judiciary i.e. the power afforded to judges on issues such as national security. This hasn't changed much in the context of the virtual world (Tamanaha, 2004). Consequently, it is generally perceived that the Rule of Law is more involved and committed to individual liberty than democratic governance. But is it really the case?

AI-based technology in particular has come under severe scrutiny in recent years for many reasons including for reasons of privacy violations whilst alleged to have infringed on the fundamental rights of citizens. Automation of governance from a public service standpoint, can span from (a) aiding a simple binary task i.e. operating street traffic lights to issuing road traffic fines from speed cameras etc right up to (b) deploying sophisticated technology for collection, processing and interpreting data (including people information) rather independently, i.e. for social security related or benefits and income support related where the law has allowed automation for evaluation and appraisal. Furthermore, in a modern context, AI-tools has also been deployed (c) for more controversial tasks such as for human profiling and predictive policing purposes by the law enforcement agencies (Tamanaha, 2004). Out of the above, ADM is most commonly seen for decades now. However, by all reports and estimates, it’s the latter two where AI is largely directed, out of which the most controversial has been the latter category, which would be critically reviewed by this research inquiry. Therefore, it could be argued that the rule of law at its finest when it's formed not just with the inclusion of various forms of safeguards that can deal with the threat of arbitrary power, but also has the capacity to effectively limit and mitigate the exercise of arbitrary power and authority by the public officials.

The EU, the UN, and other international institutions often exert the rule of law concepts in a captivating manner although not specifically referred to in its entire context (Greenstein & Sannerholm, 2022). The Council of Europe has emphasized the need to establish a comprehensive regulatory framework for AI, with the sustenance based on principles including the protection of rule of law, democracy and human rights. Individual safeguards afforded by the principles of the rule of law however have found their way into European legal scripts. For example, the General Data Protection Regulation (GDPR) and the proposed regulations on AI by the European Council in recent times. Greenstein & Sannerholm (2022)[22] however have argued in their publication ‘*Responsibility and accountability: AI, governance, and the rule of law*’ that such a shared assumption that the rule of law is largely an inner feature of the legal system.

However, when they are inferred into a regulatory framework, it is typically detached from the theological meaning of the rule of law due to its broader cultural aspect, social context (norms), and political stimulation that it requires for the principle to be respected and upheld. For example, GDPR is not a piece of regulatory instrument that lays out the significance of the rule of law; instead, it deals with data protection. Such regulations however have to be harmonized with the broader regulatory frameworks for it to be more effective in terms of respecting the core spirit and the principles of the rule of law.

[22] Greenstein, L. C. A. S., & Sannerholm, R. (2022). *Responsibility and Accountability: AI, Governance, and the Rule of Law.* Law in the Era of Artificial Intelligence. eddy.se ab.

3.2 Privacy Vs Security Argument

In this modern digital age, one could reasonably argue, is privacy dead? What if Bias becomes a feature rather than a flaw? In case, if people don't really care about privacy anymore, simply because of the attitude that they don't have anything to hide / or not breaking the law? The fact that, when it comes to communication for example, we all use some form of protection at different levels in different shapes and forms, i.e. encryption in our communication, password protection in our emails or simply putting curtains in our own homes! Why do people do these things? It can be argued that it is primarily because we all still care about privacy. Critically looking at privacy for individuals and the need for data protection, Solove's (2006) approach of elucidating the concept of privacy as 'Family resemblance'. Which in essence refers to the various dimensions of privacy. Commencing from the '*right to be left alone*', to own right to decide the extent to which we want to be subject to public observation, and something that closely entails i.e. secrecy and control over our own personal information i.e. right to be 'forgotten' (Solove, 2006).

From an European standpoint, this has been largely instrumental by 3-tier privacy laws in the European Union (EU). That can be further illustrated by **Article 8** of the **European Convention on Human Rights (ECHR)** 1950 which confers rights to citizens of the EU on this regard. **Article 7 & 8** of the **Charter of Fundamental Rights of the European Union (CFREU)** 2009 in respect of private and family life and protection of personal data, and constitutions of the National Laws of each member state that aims to protect citizens' rights for privacy and against state encroachment or prospective infringements. For example, national intelligence services etc, which has been strengthened and largely influenced by the legal instrument General Data Protection Regulations[23] (GDPR) from an enforcement standpoint. This significant and recent piece of legislation, which is applicable across the EU-wide, has strengthened the existing provisions with a strong and rigorous enforcement mechanism. GDPR primarily applies to personal data, which is any data or information with reference to an individual who could be distinctively identified with such information i.e. name, age (date of birth), gender, telephone number, location data (IP address or via GPS), or factors that characterise physical or psychological, mental, economic, culture or social identity can be classed as personal data. This applies to the '*processing'* of personal data as defined under Article 4 of GDPR, which is interpreted broadly to include any operations performed on personal data regardless of its

[23] *See* provisions of GDPR (General Data Protection Regulations) on Reference

purpose or magnitude. Which collects, records, organise or structure, store or adopt, use or transmit or disseminate etc (with the exception of processing within the context of household needs or related activities or prevention or prosecution of crimes or threats to public security). This includes for example for purposes of intelligence services which is beyond EU competency or its jurisdiction.

Critically reviewing the individual rights under the GDPR[24], they include but are not limited to i.e. right to be informed. Where individuals as EU citizens have the right to know who is *processing* their personal data. Under the right to access, individuals have the right to access personal data that has been collected on them. Under rights of rectification, have the right to require companies to rectify inaccurate information. This also empowers EU citizens with the right to object and restrict processing by requiring private actors and 3rd party companies to restrict the processing of their personal data or a specific category of same. Furthermore, under the right to be forgotten, individuals in the EU have the right to have their personal information deleted or thwart further collection. More importantly, one that directly relates to ADM, citizens have the rights in relation to this aspect of 'outsourced' system-based decision making and the process of profiling where the option of opting out of the use of their personal data by automated systems including AI-based tools.

Having said that, in the modern context, policy planners and law enforcement agencies backed by regulatory authorities all are championing the cause for the securitization of our modern societies (Schuilenburg & Hall, 2015). One drastic evidence of this includes but is not limited to putting 'big-brother'[25] on every street corner and in every single pocket of our cities in the name of national security. With the evolution in modern technology, smart cities in particular and in the case of traditional cities included, are collecting and processing colossal amounts of personal data on a daily basis for various reasons including for public service improvements and for national security purposes. Swire & Woo (2018) argues, by its very nature, data collection techniques create data privacy issues and related social problems since it involves information that includes identifiable data related to individual citizens' personal data (Swire & Woo, 2018).

More risks are associated when law enforcement agencies including the police force and other state agencies, or civil litigants need broader access to personal information. In this regard, Swire & Woo (2018) argue that there is more material risk that could potentially occur when

[24] GDPR Reference - Art. 4 GDPR – Definitions - General Data Protection Regulation (GDPR) (gdpr-info.eu)

[25] *See, Big brother is a term widely used that refers to digital surveillance in some shape or form*

law enforcement agencies have widespread access to digital surveillance and related personal data. For example, licence plate readers of vehicles by police patrolling cars that collect and process large-scale databases on road users and ordinary citizens on a day-to-day basis. Furthermore, to illustrate some examples taken from the digitally evolving applications *NoiceTube* & *SmartBay*. The former turns smartphones held by ordinary citizens into a distribution of network of noise-pollution sensors whilst the latter attempts to turn smartphones into traffic monitors. Although both these apps officially provide the option for users to control their data sharing and the latter also provides a data anonymization option with regard to the traffic monitoring process.

However, Swire & Woo (2018) argue, such systems yet lack clear and transparent information about what really happens once the users decide to share their information. Furthermore, the latter doesn't really explain its (internal) anonymization practices or its process of de-identifying personal information which, is a huge gap in the process and policies with regard to the protection of data privacy (Swire & Woo, 2018).

From a social injustice standpoint, some of the most pressing challenges facing AI-based systems include, i.e. increased usage of public biometrics, the question of ethical considerations & potential human rights violations. The question of whether or not AI-based systems (including ADM) make better decisions than humans. One could argue that AI tools and technology increasingly enable humans to make better and more informed decisions. For example, in the medical profession, it helps doctors to make better diagnoses and assess patients' medical conditions efficiently. That includes i.e. AI medical imaging and interpretations through machine learning in conjunction and collaboration with humans competency) to be able to arrive at a more accurate diagnosis and informed judgements, which is fast and effective. This is done via closely working with huge AI-based data sets that humans could possibly not imagine its scale, magnitude and capacity where humans cannot simply oversea themselves. Therefore, as Sannerholm argues, to meet its intended objectives, the key-word here is that, if and when such systems are 'sufficiently programmed' (Greenstein & Sannerholm, 2022)

However, the downside of AI is that it can have a biased or limited capacity (depends on the capacity/ way it's programmed) in thinking beyond the box (i.e. beyond what's trained/ programmed). The pre-determined data sets for example, which can be rather selective, based on certain criteria. This can be argued as a 'limitation' in this process automation. Therefore, it becomes imperative to be open and transparent by the involved actors. Be it private actors or state agencies, to be able to know what type of data sets such AI-based systems are deployed with. This in turn trains such algorithms. For example, datasets collected of public images used by captured face-recognition surveillance cameras in public or private spheres, to be able to predict, for instance, '*who might more likely to commit crimes*' in certain areas. Which might be certainly biased against certain given minority groups in certain areas. For example, due to social injustice or inequality, where certain groups may have been seen to be 'more likely' to commit crimes than others, where we could potentially be feeding AI (data source) which is already biased. In other words, although it's a reflection of the society, how representative (adequacy) is when it comes to making decisions based on such very limited perspective of data sets that could potentially drive the systems to arrive at its narrow conclusions in these real-world scenarios.

Some of the possible arguments for and against to be considered are, in medical science, for example, medical doctors often practice - what's called a '*difference of opinion'*. Especially on a complicated case, i.e. how to proceed based on different given factors. As one specialist can potentially be better than others in a given medical context and circumstances. Such variety and variables are truly valuable. Therefore, one could possibly be in danger by simply relying on AI that we simply and prematurely deprive and dismiss such available choices and consider the various possibilities in its given context and deliberations. Another example is tackling the problem of hate-speech online using tools based on AI. The biggest challenge perceived, is defining what is 'hate speech'? Which does not have an universally accepted norm or legal framework as it's deliberated in its given context (Laub, 2019).

However, if it simply fed into such AI-based systems, before it could meaningfully look at (rather than superficial patch-up work), before it could look into solutions which are far from confronting the core issues. In such situations, it could be argued that AI simply isn't helpful for the cause, as AI may not necessarily understand and fully appreciate the context in which humans would do (i.e. breach the peace). Accordingly, it could be argued that systems are not built or equipped to deal with such societal complexities which have many different dimensions and often need to be assessed from a contextual perspective. For example, AI cannot fully understand human feelings such as empathy, sorrow, misery, excitement, and social nuisance[26]– for that matter shown by research that racial biases can be entwined with language and related use of particular algorithms on such platforms.

[26] Wu, J. (2019). Empathy in Artificial Intelligence. Forbes. Reference

3.3 Presumption of Innocence

Furthermore, from a criminal justice and law enforcement standpoint, in the interest of intelligence and public security, certain people may be targeted as potential suspects on certain specific grounds who are more likely to commit a crime or develop unlawful behaviour breaching the peace. This is largely considered to be a 'pre-emptive profiling' part of predictive policing without a legally valid warrant by a competent court. Strictly from a legal sense, unless and until a lawful arrest is made, interfering with personal liberty purely based on such predictions, can amount to an unlawful action by the authorities. However as argued by Hildebrandt (2016), the consequences of such monitoring are based on so-called data-driven analysis for such 'potential suspects'. This goes against the general and legal understanding of the principle of presumption of innocence. Whereas established legal norms, a person is presumed to be innocent until proven guilty by a competent judiciary (Hildebrandt, 2016). Apart from the legal aspect, there could be an enormous amount of social impact on continuous monitoring of the so-called 'potential suspects. This would be critically discussed further below in the empirical analysis stage by reviewing case studies (supported by the conducted interview results).

3.4 Transparency & Accountability

Why does Transparency in AI matter in this digital era? It could be argued that transparency is not just a principle but also is a tool that enables accountability (Fox, 2007). In other words, if you aren't aware of what an organisation or for that matter government is doing, in such circumstances, it neither can reasonably be held accountable for its actions nor to be regulated. As transparency may relate to many aspects of AI, from data, and personal information, to goals and objectives, algorithms, coding, and compliance, which in turn could influence the usage of ADM systems. This depends on various levels of information from the authorities, general public, regulators and third-party researchers and forensic analysts.

From an AI-related tools deployment and transparency standpoint, the requisite components from an operational perspective, are gathered through specific disclosure practices. That is deemed for public awareness and for predictive algorithmic governance that closely entails, for example, digital surveillance to processes and systems concerning automating governance. Although in general, explaining the challenges from a socio-legal perspective, transparency is a multifaceted concept which can be looked at in a multidisciplinary perspective (Margetts, 2011).

However, in recent decades, this concept of transparency has gone through a revival from a present-day discourse, particularly around the topic of artificial intelligence This demands a critical revisit of defining this concept of ‘trust-worthy-AI’. This comes in, particularly in the wake of different and multiple ways in which AI-based algorithms are used for different purposes by various actors, which makes it even more complex and linked to the internet of things in this evolving digital world. Although terminologies such as algorithmic transparency and AI-based automating governance (including ADM) have become largely accepted in the research world. Larsson & Heintz (2020) argues that a broader and expanded conceptual framework is required from a contemporary standpoint (Larsson & Heintz, 2020)

The general industry practice of obscurity around the algorithmic concept as Hill (2021) argues, has been largely criticised that is often intertwined and interlinked with i.e. AI-based technologies and ADM in particular (Hill, 2021), which triggers governance issues much often associated with accountability concerns (Koene et al., 2019).

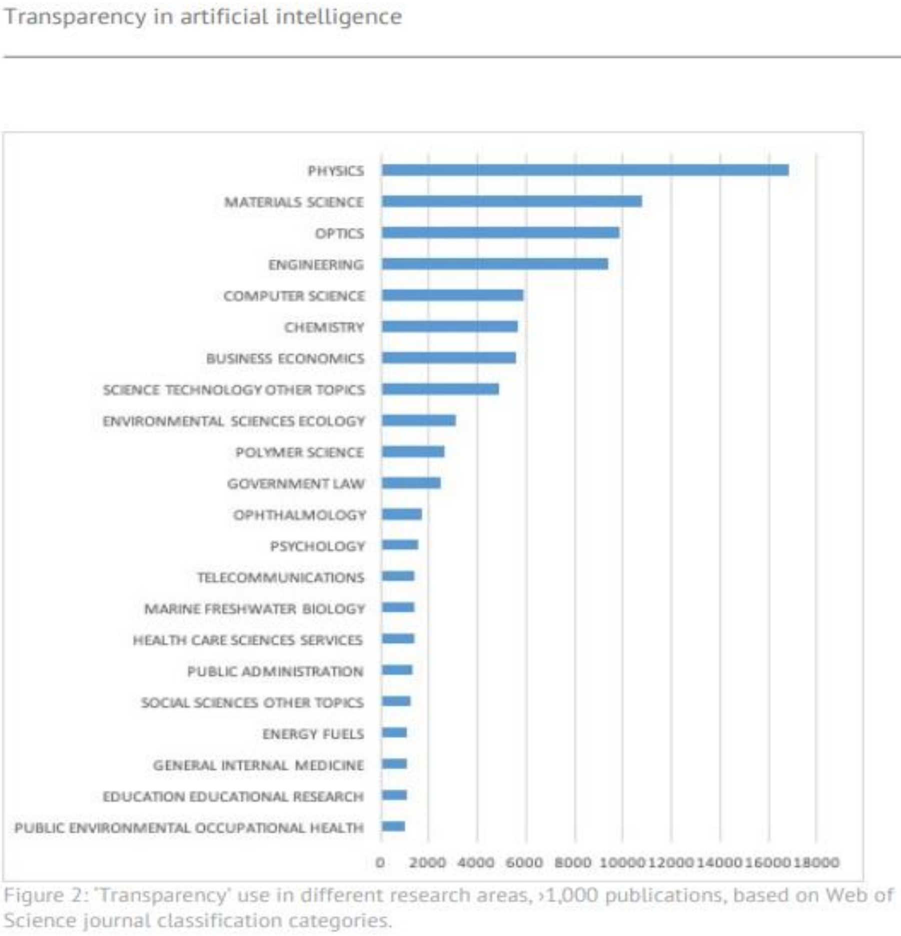

Transparency in artificial intelligence

Figure 2: 'Transparency' use in different research areas, >1,000 publications, based on Web of Science journal classification categories.

Figure 04 – *Transparency in AI by sector (Larsson & Heintz, 2020)*[27]

[27] Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. Internet Policy Review

One of the core characteristics of the rule of law is accountability, which effectively has the ability to mitigate arbitrariness, abuse of power and excess authority. This refers to the implied or explicit expectation that one may be asked to justify and defend one's belief or actions (Greenstein & Sannerholm, 2022). In the context of automated governance, critically reviewing ADM deployment using artificial intelligence or machine learning technologies, which produces direct and indirect challenges to the principle of accountability. This is because of challenges in the identification of the accountability, which is more complex due to many factors.

This includes but is not limited to ever-changing and fast-evolving technology and the dynamics around the same, i.e. the knowledge and competencies to keep up with the change. Furthermore, the context in which the deployment has taken place, the timing and the process outcome etc. besides how the responsibility factor is framed also matters (Towers-Clark, 2018) i.e. for someone (or authority) to be perceived they are accountable. For this reason, comparing the following two scenarios can elucidate the complexity, where automated governance is deployed for handling simple (quantifiable) and straight forward decisions such as issuing road traffic fine tickets for speeding, where there is less or no room for discretion. This contrasts with a more complex situation such as finding a suspect for a specific crime committed from a given central database and following up prosecution based on such AI-based analysis. This in fact leaves bigger scope for evaluation and discretion by AI-systems in the decision-making process can be challenging to map out accountability for the same complex reasons.

As per Tetlock's theoretical framework argued by Greenstein & Sannerholm (2022), AI has several hindrances in relation to accountability, (a) as it shifts when evolving technology is deployed and disputably becomes even more challenging to identify accountability especially dealing with more complex issues. Consequently, accountability typically gets transferred from public authority (where its generally perceived to be found), to private entities. This can include from developers, programmers, and those procuring such technological services (typically private actors) of such systems for the usage purposes in the public sphere (Greenstein and Sannerholm, 2022).

Accordingly, sphere of control and oversight similarly shifts as well. Therefore, in essence private actors, not only gain control of such AI-base systems but also retain oversight as well. In this backdrop that ultimate accountability has to be assessed and determined in line with the core principles of rule of law. The process of regulation by the European Commission[28] is a testament, (reflected in the accountability principles of AI), which attempts to encompass the above mentioned complexities by re-focussing its attention to incorporate AI supply chain that includes holding responsibility the suppliers and the users of AI.

Furthermore, from a transparency standpoint concerning AI and algorithmic-based systems, it is imperative to explore how the law and the supporting regulatory framework ensure such applications are not only designed, built, and programmed but also operated in a transparent and accountable manner complying with the legal framework. This includes respecting norms and principles on human rights, free from biases, and discrimination whilst promoting, openness, and transparency through constant testing and auditing process in place. The ***White paper for trustworthy AI***[29] by the European Commission, where transparency and accountability are placed as a key foundation part of its ethics guidelines (European Commission., 2019).

Citing an example, automated governance as per current trends in technology, progressing well into the future, provided technologies such as machine learning is more commonly deployed, it would mean grounds for such decision to be challenged or decoded would be much harder. For example, i.e. when admin tasks are complex, and value-based output, it becomes rather challenging for AI systems to fully understand, comply and align with the spirit and intention of the legislators and policymakers.

[28] European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain European Union Acts, Brussels, Reference

[29] High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI. European Commission. 2019; European Commission, *White Paper. On Artificial intelligence* – An European approach to excellence and trust. COM (2020) 65 final. 2020.

3.5 Surveillance Capitalism

In this day and age where public authority is increasingly being outsourced to private companies, and in return, such private companies enjoys a cordial relationship with the state agencies, who are profiting from such lucrative undertakings. Given the circumstances, which brings us to a very important and interesting point in which, where some might even argue that this alliance is a monopoly power! '*Surveillance capitalism*', coined by Shoshana Zuboff, an award-winning author and professor at Harvard University (DW, 2022). Within this process where tech companies collect oceans of data. *Meta* (parent company of Facebook) alone for example, holds and processes a total of nearly 3 Billion users account details - thanks to its recent social media platform acquisitions of Instagram and WhatsApp (Zuboff, 2019). Zuboff argues that private actors (primarily big tech firms) through such means of AI-based surveillance extract our behavioural data by 'invading' our private lives and claims such rendered (behavioural) data as their 'private property'.

This is coined as '*surveillance capitalism'* that operates and eventually became the predominant socio-economic paradigm, which has set new and largely unregulated standards. This she argues, is a complete 'illegitimate' operation (Jackson, 2021). Such rendered data (behavioural) is supposedly utilised for '*advertising purposes'*, is now also increasingly used for other objectives such as political campaigns and many other undisclosed purposes around the world. This has been a trend as Zuboff claims, started by Google and Facebook, and now from Facebook set the standard for the tech industry as a '*default option*' (Zuboff, 2019). This makes a higher benchmark for the investors with what's called a 'surveillance dividend' that produces more and quick revenues compared to 'old-fashion' capitalism, that focusses on real needs of traditional products and services.

04

Methodology

This chapter presents the methodological approach that is applied in this research inquiry. Firstly, briefly explained with the rationale of the research strategy, and the justification of the research design and method. Followed by a brief theoretical discussion of the qualitative aspect of data collection of this research paper, which also includes ethical considerations.

The fact that AI-based technology is a fast-evolving and ever-changing subject matter, which means, what's relevant and applicable today for instance, can be totally redundant and outdated tomorrow! In this context in which appreciating all factors around these dynamics, this research inquiry is conducted. This research inquiry aims to investigate the impact AI-related tools concerning automated decision-making (ADM) have on the rule of law and society. Hence the need for better and adequate regulatory safeguards. Accordingly, this research study employs a qualitative research approach with a positivist and pragmatic worldview approach to be able to fully appreciate and address the research problem. As for the approach taken, one of the core variations between a qualitative (as opposed to i.e., a quantitative research inquiry), is the linkage between the theoretical framework and the applied methodological approach (Bryman & Bell, 2011).

4.1 Research Design

The research methodology that is referred primarily to the research technique used for the data collection purposes during this research inquiry (Bryman & Bell, 2011). Where an orderly literature review has been conducted as part of the preliminary survey of the field on this novel and technical topic of artificial intelligence and related tools backed up with a solid linkage to the foundation of the theoretical framework explained above concerning this topic. To be able to meaningfully argue the selection of the methodological approach whilst defending its feasibility, and to be able to meet the above stated research aim and objective. The investigation primarily focussed on (a) secondary data collection through a well-structured critical discourse analysis as well as (b) qualitative research mixed method, that includes semi-structured interviews and focus group sessions for primary data collection.

4.2 Data Collection & Approach to the framework for analysis

4.2.1 Analysis of Empirical Materials

The empirical material for this thesis research work comprises content analysis whilst employing critical discourse analysis (CDA). Where I have primarily reviewed as explained below, which includes reports, journal articles, and other related published materials on this topic area including publications from government archives. This includes the EU, the U.K., the U.S. and Canada - to explore and establish key related facts and patterns as part of the secondary findings and analysis. This would of course be supplemented by the primary data collected from semi-structured interviews and focus group meetings as explained below.

After formulating the findings, derived from the empirical data, the next chapter presented with the analysis of the key findings where the literature review has been linked to the empirical findings and results. In this regard, firstly an overview of the analytical framework is presented, which would be followed by the respective in-depth discussions summarising and condensing the empirical findings and results of this research inquiry whilst highlighting the key findings and the implications of the same.

Critical discourse analysis as a research tool

Framework for analysing as Fairclough explains using Critical Discourse Analysis (CDA) to enhance my knowledge as a researcher whilst leading to several matters concerning as a social scientist. Fairclough's framework of discourse analysis helps to better understand and analyse the existing literature on ADM and its impact on rule of law suggesting the framework to be productively used to critically examine and address a range of issues in social science and related fields. Critical discourse analysis as Fairclough (2003) argues, can be drawn upon a wide variety of approaches to critically examine content governance of new capitalist societies, which could have blurring lines of social boundaries, and digital boundaries that shift in space and time related to globalisation for example. Where a particular discourse represents, for example, a social change or a change in communication technologies as in this research inquiry i.e. enables review of the legitimacy of such social action / or social order, and contemporary social issues amongst many other factors (Fairclough, 2003). Accordingly, CDA can be regarded as a creative and disciplined method which is grounded on the text and various forms of content which is utilised not only for describing and interpreting the digital transformation and the shifting needs of the society in its given context (Brown and Yule, 1985)[30].

This would be more so relevant to justify the qualitative research methodological approach that's been selected for this research inquiry not just from a sociology of law perspective but also broadly from social science in particular. Which is inspired by the objective of providing a scientific basis for critically questioning the evolving social order and inquiring into current digital and social life questions from an ontological inquiry perspective. Critically reviewing enables to arrive at epistemological conclusions not just strictly legal but also from a moral and ethical standpoint in terms of social justice and power (Chouliaraki and Fairclough 1999).

Applying and justifying the selected method firstly, by engaging with the narratives and social discourses around AI-based digital surveillance and automating governance specifically, ADM. Secondly, examining the legislative provisions and the social discourses around it that impact the rule of law, and the broader society, within the lens of social justice. The findings from this research approach are understood and discussed within the theoretical framework as explained above. This includes comprehending the concept of rule of law and its normativity in this digital era, whilst enabling us to fully appreciate the concept of artificial intelligence, and ADM related AI-related tools such as profiling, predictive policing and surveillance capitalism, from the standpoint of transparency and accountability.

In this manner the theoretical discussion is extended to fully understand and appreciate the multi-dimensional discourses around this controversial subject matter. This is done primarily focusing on the European Union, which has been compared with certain other specific jurisdictions including North America. Analysing examples, comparing with related cases and reviewing the legislative provisions from within the EU (i.e. Sweden, and the UK) and outside the European Union i.e. United States, Canada and China, for example, helps better understand from a wider global perspective.

[30] Brown, G. and Yule, G. (1985) Discourse Analysis. Cambridge, Cambridge University Press.

4.2.2 Primary Data Collection

In relation to collection of the primary data, several semi-structured interviews and focus group meetings have been conducted with different stakeholders, from different geographic locations, representing the general public, academic and legal scholars, and the industry practitioners.

The nature of the interviews was such that whilst it provides some degree of focus as an interviewer, that was navigated with the help of the interview-guide[31] used. However, at the same time, using a semi-structured interview guide, also provided a great deal of flexibility and autonomy for the interviewees to respond as they wish depending on the context and circumstances (Bryman & Bell, 2011). The intent was to exhibit and facilitate a more flexible approach during the interviews conducted. Creating room for more open and honest discussion was regarded as a paramount factor that allowed participants to feel more comfortable with sharing their open and honest input on and off the digital sphere that enables them to capture useful insights and findings. Accordingly, to be able to arrive at above stated research aims and objectives, identifying the most fruitful way of this qualitative research mixed-method approach, as opposed to a single-method model may simply not be adequate to enable exploring the answers to the multiple research questions. This is in line with Keating & Donatella Porta's (2008) assertion that "*social science knowledge is a collective enterprise, built using various techniques, methodologies and methods*" (Keating & Donatella Porta, 2008)[32].

This has been carried out by a combination of discourse analysis and content analysis, to critically review the various discourses around the rule of law implications and legislative provisions in the EU concerning AI-based ADM. Furthermore, as a strategy for the literature review – used keywords, based on preliminary research done including the survey of the field in my previous research inquiry. Which has been beneficial to further this cause for an enhanced research inquiry. Accordingly, this would focus on expanding on the same including the keywords that suitably sum up the subject matter including *LUB search* and *google scholar* focussing on AI-based technology and its consequences of unchecked expansion of AI-based automation and related development initiatives to be able to search and find relevant and suitable material in this regard.

[31] Refer to the sample interview guide deployed in the appendix section

[32] Keating, M. & della Porta, D. (2008) *Comparing Approaches, Methodologies and Methods. Some Concluding Remarks*, in D. della Porta and M. Keating (eds) Approaches and Methodologies in the Social Sciences: A Pluralist Perspective. Cambridge: University Press, pp. 316-322

4.3 Ethical Consideration

This research project has been given due ethical consideration given the fact that dealing with ethical dilemmas is an integral part of each stage of the research inquiry. This includes preserving and processing sensitive information including personal data of those in concern who are considered to be a part of this research work. Thus, have strictly adhered to the general academic standards[33] and guidelines[34] whilst acknowledging the importance of recognising that ethical considerations encountered throughout this development of the research process. As per Guillemin and Gillam (2004)[35], have looked at and given due consideration to both forms of procedural ethics i.e. regulations and ethics in reality and practise i.e. dealing with day-to-day ethical dilemmas (Guillemin & Gillam, 2004).

This includes but is not limited to the preparation stage, and the stages within the process of collecting, and processing sensitive information obtained from the participants (including through i.e. interviews and focus groups) and all other forms of details for this research purpose as a matter of confidentiality. For all intent and purposes, all necessary personal data and sensitive information collected have been anonymised and kept confidential to protect and respect participants' privacy. This is in line with institutions' general academic standards and guidelines. This includes maintaining the privacy and confidentiality of all participants concerned, whilst respecting and adhering to guidelines on obtaining informed consent.

In this regard, I have strived to meet the highest academic standards and procedural requirements at every possible stage of this research inquiry. This is primarily to be able to comply with the requirement of informed consent obtaining both orally and in writing where it's needed, with participants' fullest knowledge. This has been compiled whilst having remained flexible and served the needs of the research participants including the interviewees. Moreover, conducting a constant ethical review, whilst avoiding any form of research misconduct and adhering to general academic guidelines. Furthermore, as a researcher from a Scandinavian institute - Lund University, dealing with this research inquiry involves a critical evaluation of the social impact concerning minorities and largely black and coloured communities. As one could appreciate, it is a considerable factor that has its fair share of

[33] Research Data, Rules and regulations, Lund University (2020) Reference

[34] Guidelines for processing of good research practice at Lund, Lund University (2021) Reference

[35] Guillemin, M. and Gillam, L. (2004) *Ethics, Reflexivity, and "Ethically Important Moments*" in Research, Qualitative Inquiry, 10 (2), pp-261-280.

sensitivities surrounding same. Hence making sure my own identity does not or perceived to be influencing any potential biases, concerning to my understanding, opinions or outlook on the world and related topics researched, in relation to positionality and the production of knowledge (Gary& Holmes, 2020) in this regard.

05

Empirical Findings & Analysis of Results

5.1 Empirical Data Analysis & Discussion

From a simple algorithm-based recommendation to watch a movie by streaming giant - Netflix to an instantaneous automated result - provided by an aptitude test used by many recruiters by many corporations. A decision awarded to a banking customer on his/ her loan status, and all the way up to a life-long / life-impacting decision i.e. whether they get a job (get hired / or fired!). All this could get further complicated for complex fields such as immigration decisions (that determine an individual's immigration status), or even more sensitive topics such as decisions involving prisoners and their release from prison or for that matter how children are removed from their families etc.

With the growing influence of big data and AI-based system tools, the following non-exhaustive and broad-range of scope, as per Fritsch & Thomas (2019), where ADM is used even within the justice system. The following areas that include but are not limited to i.e., education (exam results related to predictions) pupils' behaviour, to child-welfare related assessment decisions. Access to state benefits (including health care or job-seeking compensation related). Moreover, access to housing (i.e., eligibility and queuing system), immigration-related case processing, and surveillance systems (including computer-based camera systems that are deployed for fighting crimes by different state affiliated agencies including those sub-contracted to private actors) in the interest of national security by the law enforcement agencies. Furthermore, even at prison services (i.e., involving bail and sentencing and parole-related decisions based on system reviewed records) etc., which are few out of many other examples where AI-based systems are deployed as increasingly part of automated governance and decision-making (Fritsch & Thomas, 2019).

Computer-based calculations which are 'sufficiently' programmed are good at 'number crunching' that has been in use for years or decades now for various calculations and automation purposes at different decision-making levels. However, as per Deloitte's Report, the evolution of new technological breakthroughs in the fields of algorithm and AI-based tools including computer program-based data capture and analysis has taken the ‘number-crunching’ process to a whole new level in terms of automated predictions. Be it weather-related forecasting, sporting and scoring probabilities, ground and air traffic management, to even more complex and sensitive activities such as medical diagnostics and predictions that are beneficial to society (Kudumala et al., 2022). Such forecasting and prescribing advance actions including but not limited to those connected to medical science, or autonomous (self-driving) vehicles are spearheading in this area of technological breakthrough. This is primarily by making advanced calculations and predictions particularly on prospective human behaviour etc. All this has been facilitated through collecting and processing of complex and large amounts of data from various sources that include those from ordinary citizens’ personal data on a day-to-day basis.

Hence these developments in technology and more importantly, its data collection, processing and analytical practices, beg the following questions of (a) whether adequate safety and security measures are in place for the protection of citizens’ privacy and fundamental rights, and (b) if not, at what (social) cost? In this context AI-based algorithms and data analytics increasingly substitutes and even to a larger extent replaces the traditional human decision-making. Consequently, the great concerns and ethical questions are far more prevalent and pressing. Although corporates are happy with its cutting-edge breakthrough progress, without sufficient countermeasures for transparency and accountability are in place. The civil society increasingly questions such rapid development without adequate safeguards in terms of privacy and human rights (Goodman, 2018).

Legislative Developments

There has been very little consideration provided in terms of the right to due process, as argued by Hildebrandt, which is a vital component of the principle of the rule of law. Plus, the fact that it also relates to privacy, non-discrimination and presumption of innocence in a direct manner. For example, in America, due process is provided as a legal right which is enshrined in the Fourteenth Amendment of the United States Constitution. Which specifies amongst other things, that *‘no one shall be deprived of the citizens’ right to life, liberty, or property without due process’*. This can also be equated with procedural fairness, which has a similar connotation in Article 6 of the **European Convention of Human Rights (ECHR)**, which is based on the concept of equality and respect for individual liberty. This also focuses primarily on fair trial, concerning the impartial judiciary, trial before a competent judge, affording public hearing, and most importantly, presumption of innocence in the case of criminal justice (Hildebrandt, 2016).

Some of the recent EU legislative approaches particularly the **Administrative Procedure Act** show some progress towards incorporating principles of rule of law. This can be seen to be drawing attention to technical details although to a certain extent in terms of individual safeguards. Whereby a simple paragraph, refers to considering ADM directed at public authorities. Which in fact, has a high threshold to comply with the legality. This however does not fully include sufficient safeguards as specified in the European data protection regulations (**GDPR**) and in compliance with **Article 29 of the Working Party** (HLEG-AI) on trustworthy AI. Furthermore, the **Administrative Procedures Act** also entirely disregards the issue of accountability of those designs or codes the algorithms that facilitate the automation of governance or the responsible public authority who decides for such deployment (Greenstein & Sannerholm, 2022).

Thus, the above legal lacuna opens up a huge gap in terms of complying with the principles of rule of law that effectively denies the right to be informed, the right to challenge a decision, right to fair trial whilst potentially breaching many other fundamental rights of the citizens with the automation of governance including the ADM process. Legislative provisions such as Article 22 of the Working Party Act, generally forbid decision-making entirely based on automation, which includes profiling if it creates legal effects on individuals. However, there are certain exceptions for the public bodies, i.e. if and when such automation is endorsed by labour unions or by the Member State law for example, which provides suitable safeguards. This could be further illustrated by Article 29 of the Working Party Act, which stipulates that safeguards could refer to that include right to human intervention, or specific information to the data subject. For which to be heard or obtain an explanation of the submitted decision or right to challenge such decisions. The legal basis in Sweden however, poses certain challenges including legal barriers with regard to ADM particularly at the municipal level. This could be regarded as beyond public authority's capacity to make decisions i.e. ultra vires (beyond authority). Such legal grounds should comply with fairness principles, which should also be clear, and easy to understand (not complex technical jargon), and more importantly transparent, which is regarded as a cornerstone in-line with HLEG-AI's ethics guidelines by the European Commission.

Moreover, following the GDPR[36], when such decisions are solely made by an automated process, European citizens and residents have the right to be informed. Although the extent of this right has been fiercely debated amongst academics and legal practitioners. This

[36] GDPR Reference

exponential growth and complexities of AI-based machine learning however can be seen as posing a challenge for comprehension on how the process including algorithms of ADM or profiling for example, may turn out with its rapid change and evolution in the AI technology.

Controversies around Access to justice

AI-based tools used in the process of profiling, predictive policing and prosecution confront criticism mainly for the following reasons. This includes such interpretations involving complex and non-transparent algorithms. Which fails to explain or reasoning behind such a mechanism. Secondly, such data analysis and reports are ambiguous and how certain quarters of the population are unjustifiably and disproportionately targeted. Furthermore, most critically, the big data largely relied on and utilised for such predictions purposes are often a result of racialized marginalised minority communities (Benjamin, 2019). This could be argued that such practices violate basic principles of fairness and the rights of due process. Therefore, on one hand, which can be argued potentially triggers human rights violations. On the other hand, this begs the question of being able to fully appreciate the consequences and impact of ADM. Initial step towards mitigating this challenge would be the awareness of its existence which is connected to access to justice.

The risk of data breaches and responsibility towards the public

When such colossal amounts of people's data is retained by 3rd parties, another growing risk that has emerged of late is the risk of data breaches. In such circumstances, what are the responsibilities possessors and processors owe to the general public? From a European standpoint however, for example, under GDPR, there are 9 core governing principles for the processing of personal data. They include having a lawful and legitimate purpose, whilst taking into account the context and specific circumstances in which it is processed, which is to be done in fair, just and reasonable manner. All reasonable steps taken to maintain accuracy and the integrity of the personal data held in a secure manner complying with principles of transparency and accountability.

GDPR regulation which is regarded as a strong and far-reaching legal instrument concerning data protection within Europe. The data privacy and security law include fresh requirements that potentially impact organisations around the world dealing with European consumers. This compliance impacts not only the companies within Europe but also entities outside the EU due to its extraterritorial reach of the law. Organisations who have a market reach in the EU or deal with the EU consumers, will have to comply with these strong regulations. In other words, any business that has a digital presence in the EU would require strict compliance with GDPR.

The above EU regulation which came into effect on May 25th 2018, is considered to be the toughest security and privacy law in the entire wide universe. Although it is a law passed in the European Union (EU), whilst retaining enforceability jurisdiction extraterritorially. This means that it imposes obligations on organisations anywhere in the world, provided they have targeted or collected data concerning people in the European Union. Any possible breach of violating the GDPR is swiftly dealt with, in terms of monetary penalties. Rather larger fines are imposed compared to previous laws. This can amount to a maximum amount of €20 million or 4% of global revenue, whichever is higher. Furthermore, the subject matter has the right to pursue compensation for damages. As per GDPR, personal data is defined broadly, which amounts to any information, concerning individuals (citizen or resident) who can be directly or indirectly identified. Personal details such as names, email addresses are some of the obvious personal data, whilst location-related information, and other i.e. biometric data, gender, religious beliefs, ethnicity, and other technical-related subject matters such as accessed web cookies, or political opinions could also be regarded as personal data under this legal provision. Under this regulation, data processing can mean any action performed on data, regardless of whether automated or manual, for instance, cited in the text that includes, collection, recording, organising, using or storing (including erasing) and any other related activity.

In essence, it’s a far-reaching legal instrument that provides safeguards to EU citizens and residents in relation to data protection, concerning the privacy and security of their data. This can be regarded as a huge step in the direction towards individual safeguards at least technically against any unchecked and unregulated AI tools directed at big tech corporations (Browne, 2022).

Nonetheless, in this modern day and age, cyber risk is far too real in any domain, beyond borders and across the world! In essence, no cyberspace can be regarded as a safe haven from cyber-attacks or widely known as ‘cyber-hacks’! Such cyber-hacks can be detrimental in many ways, i.e. on critical infrastructure that can be devastating from a national security standpoint.

Similarly, at the same time, cyber threats on companies can compromise corporate credibility and loss of reputation in addition to financial loss and penalties for prospective data breaches including the risks of identity theft. As Anderson (2013) argues, in this day and age every corporate entity is a victim of cyber threat, which is nature elusive, and the precarious part of this is the fact that victims don't even know how many times they have been victimised nor do they have a quantifying or measuring mechanism or identify the level of impact it really had (Anderson, 2013). Unless voluntarily undertaken a cyber forensic investigation or more popularly known as *'penetration test'* part of an IS audit. Regardless, as per the above legal provisions has become a primary responsibility of the respective companies who possess and process the personal data of the public from a data protection standpoint.

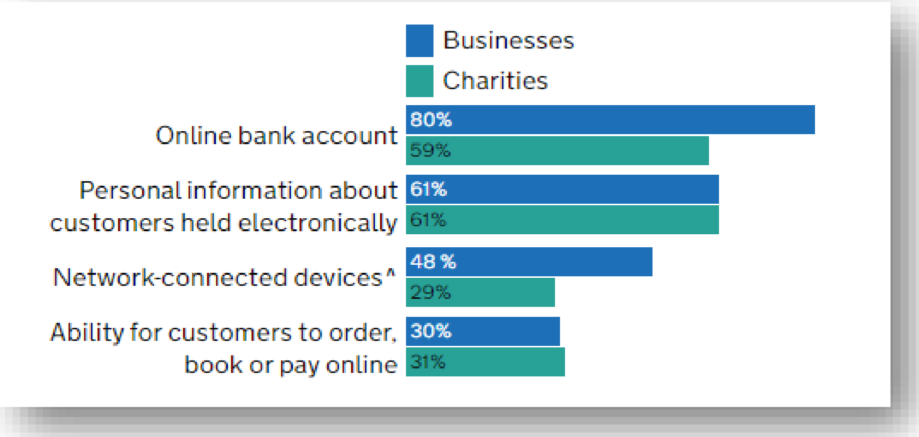


Figure 05.

illustration by GOV.UK - data shows entities currently possess or utilise digital services[37]

[37] *See*, Statistical illustration shows the digital penetration and its distribution channels and various modes across UK entities. Reference

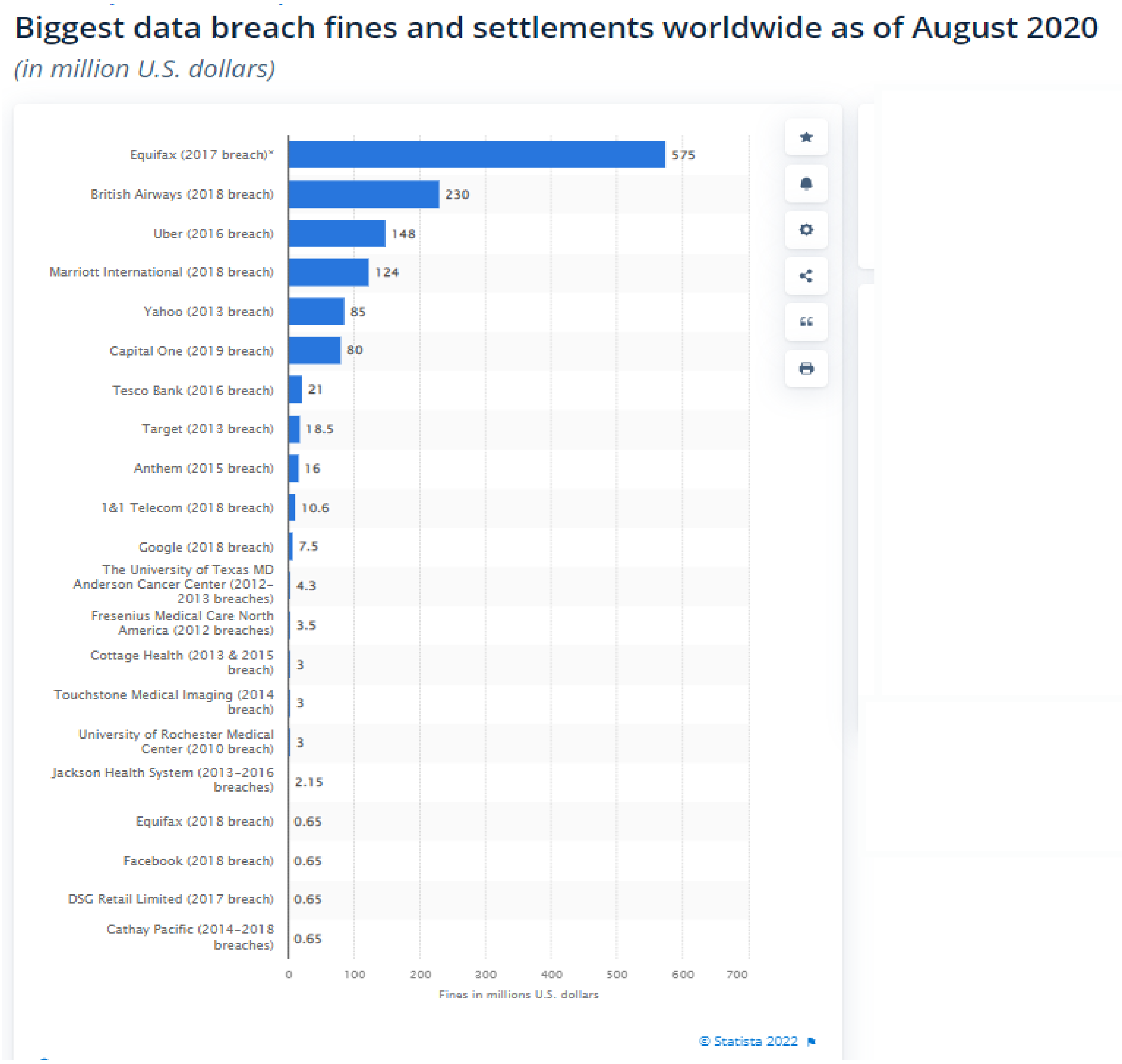


Figure 06*: - Global corporate cases involving data breach fines as of 2020*[38]

The above illustration (Statista, 2020) reveals those entities that are entrusted including those corporate brands that have a fairly good reputation. This includes the American hospitality chain Marriot International, UK's British Airways, and the world's largest online search engine Google are among the victims cum culprits (depending on the manner we interpret the circumstances). It is connected to facing the threat of cyber-attacks directed indiscriminately which results in data breaches involving private citizens' information.

[38] *See*, Landmark cases that involves some of the biggest data breach fines globally, courtesy of Statista. Reference

For the same reason, its a high priority on national security grounds and on the corporate agenda (Anderson, 2013). Privacy Laws in the United States, unlike in the EU, where there is no comprehensive set of legal frameworks for data protection laws, nor do they have the additional layer in terms of privacy protection. Instead, the American legal system relies on certain provisions and takes a sectoral approach (e.g., healthcare, finance, consumer protection etc) over many hundreds of privacy laws at the state level. For example, the new **California Consumer Privacy Act (CCPA) 2018** – emphasises i.e., right to access and notice, the right to opt-out, the right to be forgotten, right to request for deletion and the right to equal service etc. The **US Constitution of 1787 (Bill of Rights)**, where the American Supreme Court has broadly interpreted this with certain specific articles of the constitution and the Bill of Rights. For example, the fourth Amendment safeguards individual privacy rights.

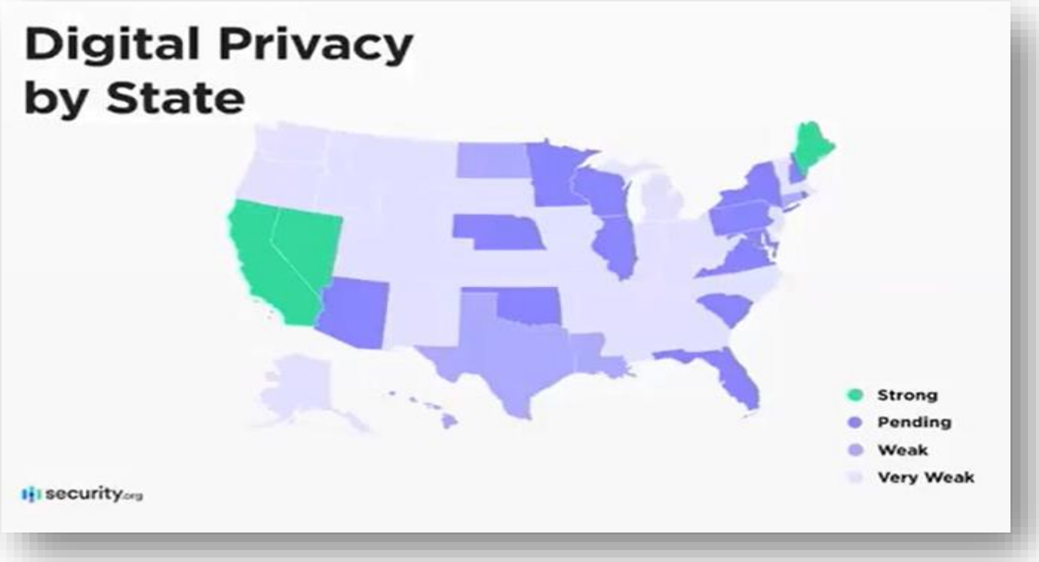
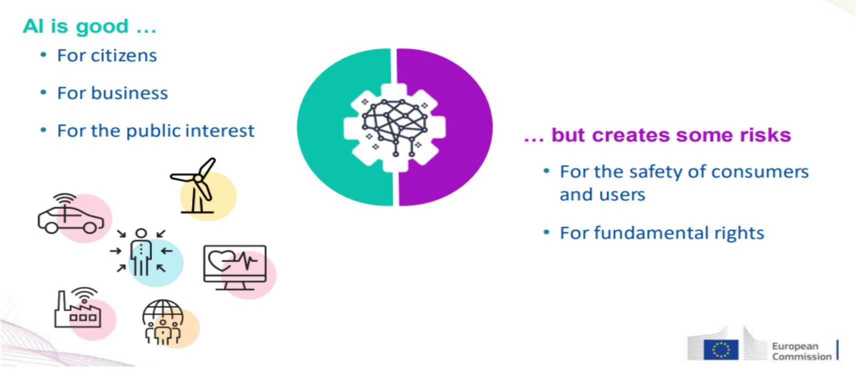


Figure 07*: an illustration sourced from security.org*[39]

[39] *See, for more information –* Reference

The case for effective regulation

Technological barriers potentially undermine transparency and accountability. Technological hurdles and technical capacities thwart effective and profound scrutiny of AI-related tools. For the same reason, it's a challenge concerning transparency and therefore, accountability for its practices and actions. To elaborate on this, AI-based automated decisions, or technologies based for predictions or profiling and awareness around such subject matters to many people including certain industry experts, as it's a fast-evolving - that makes it even harder or next to impossible to scrutinise or challenge the processes, decisions and its outcomes[40].



AI is good ...

- For citizens
- For business
- For the public interest

... but creates some risks

- For the safety of consumers and users
- For fundamental rights

Figure 08 - *Illustration from the EU's proposal on AI*

European Strategy and proposal for legal framework on artificial Intelligence[41]

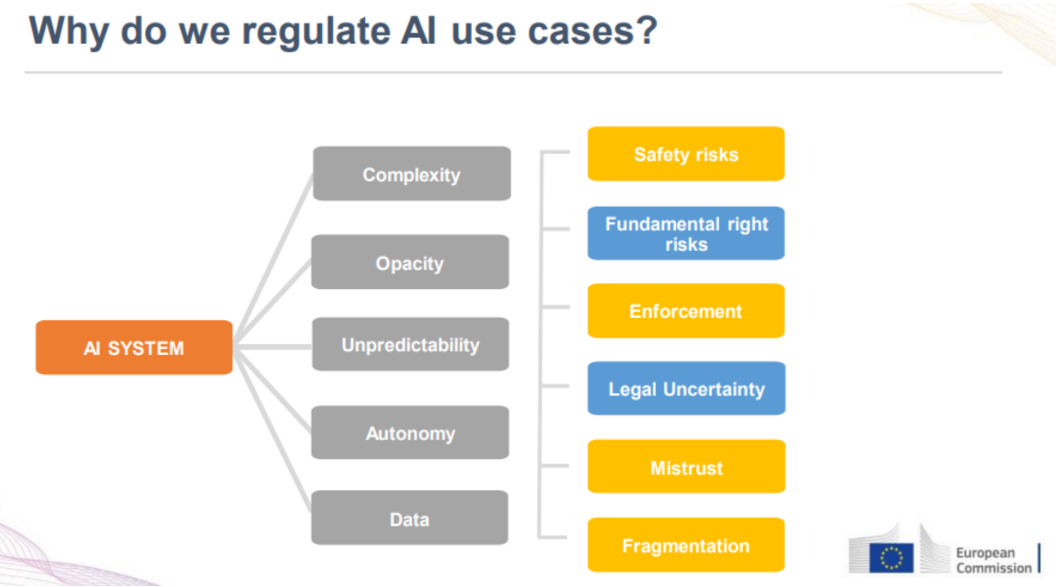
A positive step in the right direction

As discussed above in detail, although there is no single, unified and universal definition of artificial intelligence, nor has a single, universal or holistic and far-reaching legal framework on its role, responsibilities, and accountability from a regulatory standpoint. Nonetheless, looking at the recent EU legislative developments, including for example, Article 3 of the recently proposed regulation by the European Commission[42] indicates that policymakers and regulators are progressively looking in the direction of a risk-based assessment and have proposed an enhanced regulatory intervention. This would not only benefit the broader society but also at the same time would enable reasonably bridge the prevailing legislative-gap keeping pace with the evolving technology challenges and the needs of the society.

[40] Shaw, J. (2018, December 6). Artificial Intelligence and Ethics. Harvard Magazine

[41] Sioli, L. (2021). A European Strategy for Artificial Intelligence.

[42] *See, the above definition subject to AI being as neutral as possible in order to cover the techniques which are not yet discovered / developed, which covers all forms of AI (i.e. machine learning, symbolic AI, and hybrid systems), and Annex 1 – refers to AI approaches to provide some form of legal certainty.*



Why do we regulate AI use cases?

AI SYSTEM

Complexity

Opacity

Unpredictability

Autonomy

Data

Safety risks

Fundamental right risks

Enforcement

Legal Uncertainty

Mistrust

Fragmentation

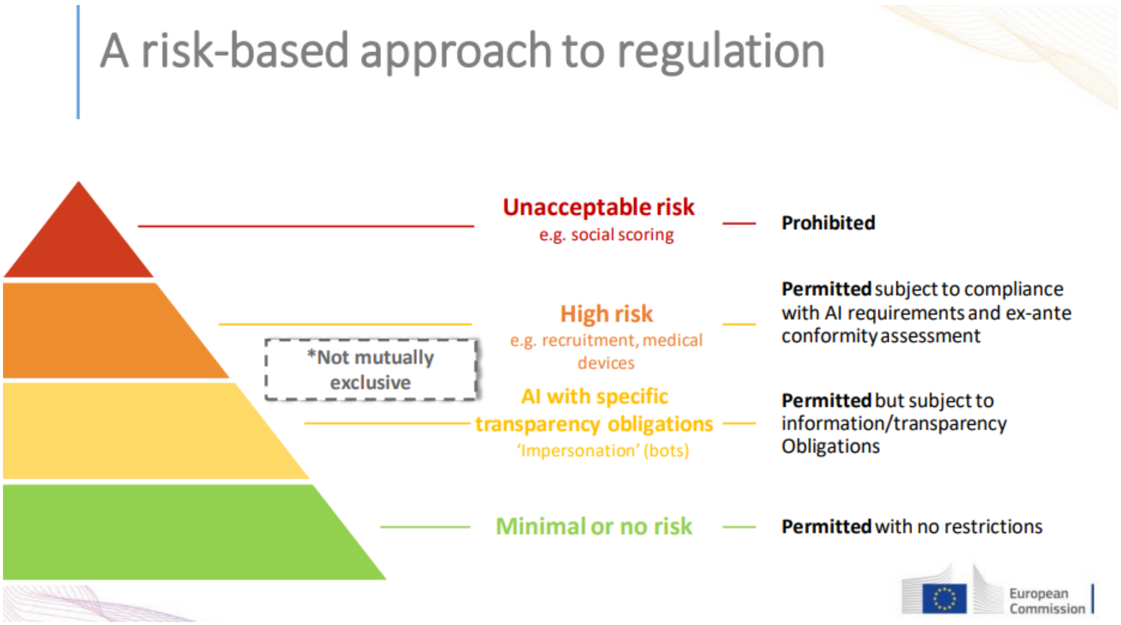
European Commission

Figure 9 –

Above illustrates the complexities of AI and the direction in which future regulatory interventions are proposed to strengthen the vulnerabilities

This could be further illustrated as follows. One such proposed approach for the regulation of artificial intelligence could take a *4-tier risk-based* approach as the European Council looks at recommending. Where it is proposed starts with ***Level 1***. minimal, low or no risk, which can be allowed with less or no restrictions. ***Level 2***, AI-specific, which could be permitted but of course subject to certain transparency obligations. ***Level 3***, which is regarded as '*high risk'* i.e. medical services, autonomous vehicles on streets etc., can be regulated and permitted, subject to strict compliance within a regulatory framework with stringent and regular conformity evaluation. ***Level 4***, which is classed as unacceptable risk category, that includes for example, the Chinese Government-led social scoring system, which involves AI-based human profiling, [43] which should be prohibited undoubtedly.

[43] Sioli, L. (2021). A European Strategy for Artificial Intelligence.



A risk-based approach to regulation

Unacceptable risk
e.g. social scoring
— Prohibited

High risk
e.g. recruitment, medical devices
— **Permitted** subject to compliance with AI requirements and ex-ante conformity assessment

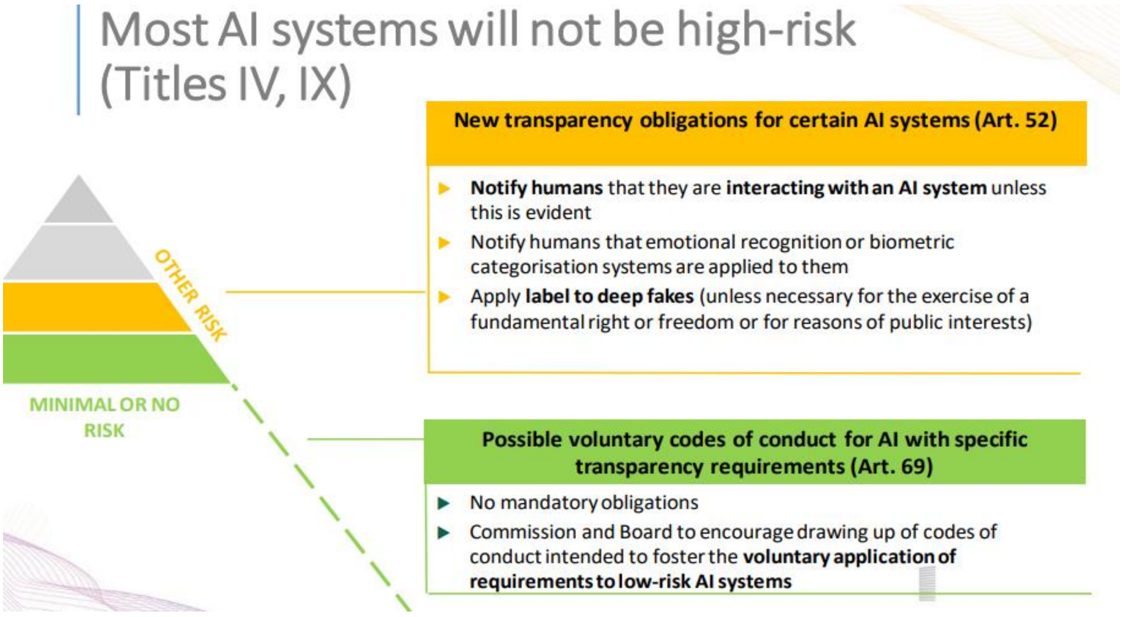
*Not mutually exclusive

AI with specific transparency obligations
'Impersonation' (bots)
— **Permitted** but subject to information/transparency Obligations

Minimal or no risk
— **Permitted** with no restrictions

***Figure 10**:*

With the above in mind, whilst most AI-based systems aren't regarded as high-risk, and in fact the potential high-risk AI systems are recommend in the above manner (figure 10, and other categorization provided below) in the very recent EU proposal / AI framework by the European Commission recommends as follows



Most AI systems will not be high-risk (Titles IV, IX)

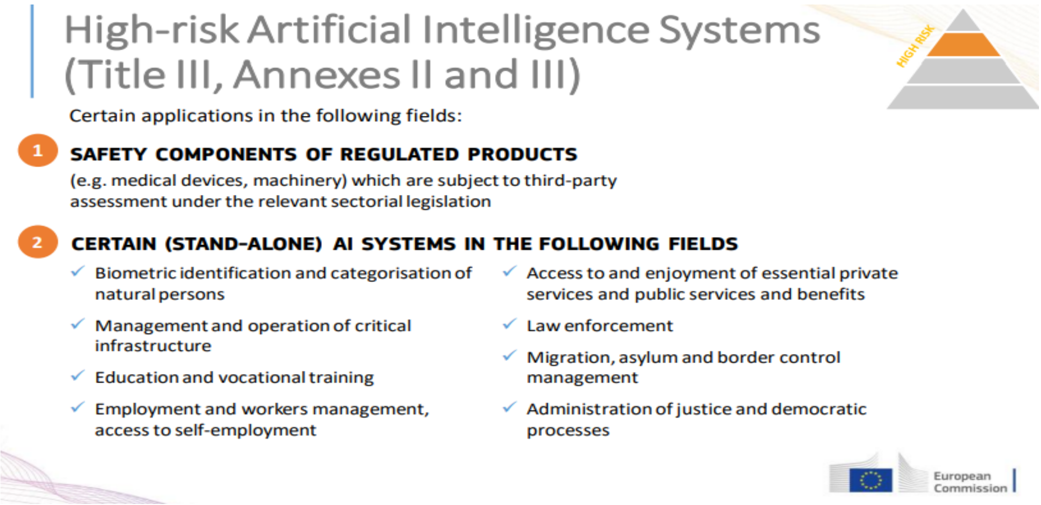
New transparency obligations for certain AI systems (Art. 52)

- **Notify humans** that they are **interacting with an AI system** unless this is evident
- Notify humans that emotional recognition or biometric categorisation systems are applied to them
- Apply **label to deep fakes** (unless necessary for the exercise of a fundamental right or freedom or for reasons of public interests)

Possible voluntary codes of conduct for AI with specific transparency requirements (Art. 69)

- No mandatory obligations
- Commission and Board to encourage drawing up of codes of conduct intended to foster the **voluntary application of requirements to low-risk AI systems**

***Figure 11**:*



High-risk Artificial Intelligence Systems (Title III, Annexes II and III)

Certain applications in the following fields:

1 SAFETY COMPONENTS OF REGULATED PRODUCTS

(e.g. medical devices, machinery) which are subject to third-party assessment under the relevant sectorial legislation

2 CERTAIN (STAND-ALONE) AI SYSTEMS IN THE FOLLOWING FIELDS

- ✓ Biometric identification and categorisation of natural persons
- ✓ Management and operation of critical infrastructure
- ✓ Education and vocational training
- ✓ Employment and workers management, access to self-employment
- ✓ Access to and enjoyment of essential private services and public services and benefits
- ✓ Law enforcement
- ✓ Migration, asylum and border control management
- ✓ Administration of justice and democratic processes

European Commission

Figure 12*:*

5.2 Primary Data Analysis & Discussion

Overview of the Interviews & Focus Group

Below ***Table 01*** *provides an overview of the interviews and focus group (FG) sessions held with selected participants, within the categories of (a) individuals/ perceived victims, (b) Academic scholars with legal background and (c) Industry practitioners with AI/ digital transformation expertise, are from different geographic locations. Interviews/ FG were conducted from February 16th to April 18th, 2022.*

Participants for the interviews and focus group were identified through mutual contacts, joining virtually from Manchester, Birmingham (UK), & Malmö, (Sweden), this was carried out during the partial lockdowns experienced in the UK due to the covid pandemic restrictions. As for the interview with the academic scholar, a well-established and renowned professor from Stockholm University, having a background in Law and in digital transformation, met in person at Stockholm University, Stockholm for a face-to-face interview. As for the perspective of AI and transformation specialists, I have interviewed three (03) industry practitioners as part of this research. Two (02) practitioners met at *Boise State University* (Idaho), in the USA, and one (01) at the *Ingka Group (part of IKEA)* in Malmö, Sweden between the period of March 09th – April 18th, 2022.

Category of Interviews/ FG	**Date**	**Interview type/ (code)**	**Role of the Interviewee**	**Location**
		FG		
Focus Group (FG) (virtual)	16/ 02/ 2022	**A1**	**Parent** of a child victim	Manchester, UK
	16/ 02/ 2022	**A2**	**Parent** of a child victim	Birmingham, UK
	16/ 02/ 2022	**A3**	**Family member** of the victim	Malmö, Sweden
		Interviews (Face-to-face)		
Interview (1-to-1)	02/03/2022	Face-to-face **B1**	**Academic Scholar** / Legal Counsel	Stockholm, Sweden
Interview (1-to-1)	09/03/2022	1-to-1 **C1**	**Industry practitioner** / AI expertise	Boise, ID (USA)
Interview (1-to-1)	09/03/2022	1-to-1 **C2**	**Industry practitioner** / AI & Transformation expertise	Boise, ID (USA)
Interview (1-to-1)	18/04/2022	Face-to-face **C3**	**Industry practitioner** / Transformation expertise	Malmö, Sweden

Table 01: *Overview of the interviewed respondents (part of primary data collection)*

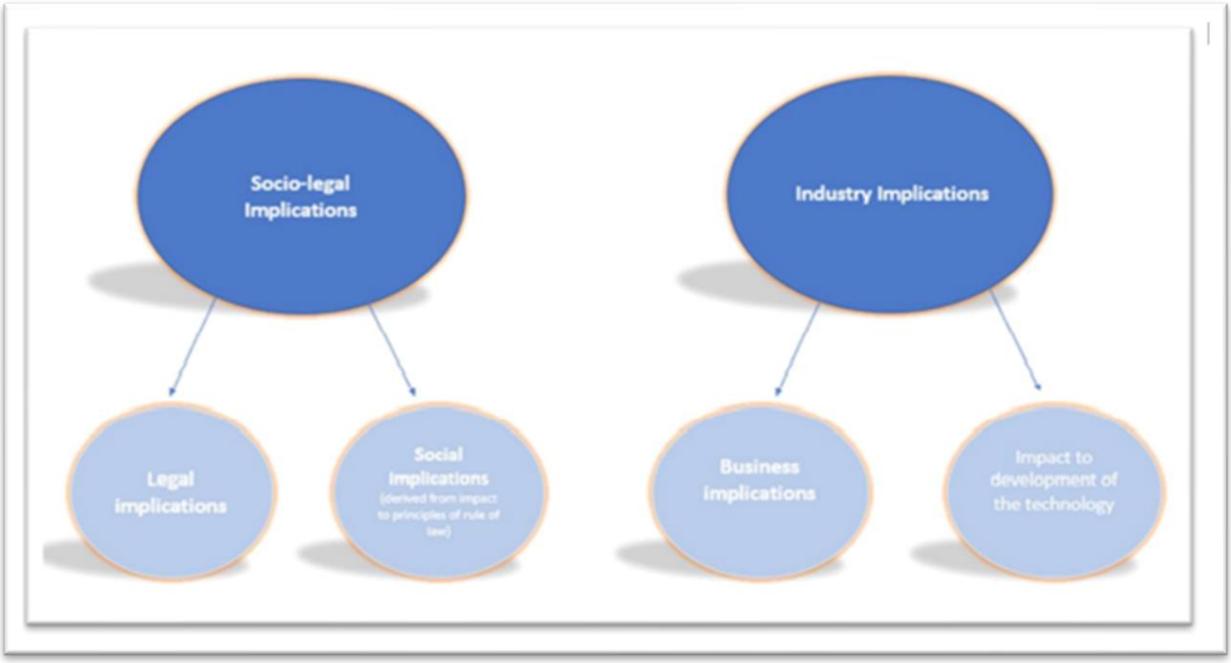


Figure 13: *– Help visualise the framework of the analysis*

5.2.1 From a Citizen’s Standpoint

General aspects of awareness of the evolving technology and its implications

The interview questions aimed at investigating participant knowledge and awareness particularly with regard to AI systems (i.e., digital surveillance, and ADM), including the impact and ramifications on society. Participants were selected through mutual contacts, who were perceived to be casualties observed at different degrees in various shapes and forms of victimisation. The interview guide[44] that has largely aided me to conduct the focus group interviews, where the participants generally showed some degree of awareness of the existence of such (AI) systems although had very little awareness of its long-term ramifications or prospective impact on their daily lives except for one interviewee (A2), who had somewhat a better knowledge and awareness although it was very limited in nature, especially on their knowledge concerning legal rights that were afforded to them.

[44] *Please refer to sample attached in the list of appendices below*

This was rather evident when dealing with further specific questions around AI tools concerning *profiling* and *predictive policing* in their neighbourhood. Participants have largely shown very little awareness of such tools nor had specific knowledge of the law enforcement usage on their fellow citizens except for one participant (A2) - parent of a victim, who acknowledged her limited knowledge of their usage on marginalised neighbour in particular but admitted rather clueless on how they actually work or why they have been targeted at all. She further goes on to explain that they are unable to understand *"especially when their children's parents in particular had clean records of conduct and also the fact that not all children particularly other children from the same neighbourhood mostly of white backgrounds were not targeted"*. However, they admitted their limited awareness of such AI tools have been used against their children in particular, within the marginalised communities, mostly in minority neighbourhoods.

With regard to AI-related tools such as *profiling* and *predictive policing,* although participants had very little knowledge of the technical aspect, they appear to show some degree of awareness and their potential social impact to their lives although some showed more knowledge than others. One of the participants (A2) explained the situation concerning their loved ones with reference to her two sons. A single parent, primarily from a decent neighbourhood, and most of her sons' associating friends are from a well-educated and largely elite white community. Other participants are largely from an ethnic minority neighbourhood and according to them a generally perceived to be an 'under privilege' neighbourhood.

On the question of how it could impact theirs/ and their family members' future in terms of potential legal implications, most of them have shown very little awareness of the same. At the same time, one of the participants (A2) whilst identifying herself as a *parent of the victim* (referring to her 15-year-old & 17-year-old sons), was happy to get into a deep discussion due to their circumstances and her relative knowledge of such AI-related tools and its implications (with reference to profiling & predictive policing practices). When inquired if they ever felt marginalised or unduly targeted, some of them did admit they had strongly felt the same. Although other participants had a similar marginalised background, they haven't specifically felt the same way as strongly as some, from an artificial intelligence driven tools perspective.

'Digging deep into – for no avail!'

Further into specifics, one of the respondents opened up explaining in detail. On the question around having personally experienced any social dilemmas or discrimination as a result, one of the participants explained, the fact with the peaceful nature and the decent neighbour they live in, she had never felt that way (marginalised or discriminatory treatment) until she had encountered her children's episode and state of affairs. Admitting that she was shocked and traumatised when she first figured it out.

She admits, when she first got to know both her children (aged 15 & 17) were on a local police surveillance database, she was shocked and devastated, and at the same time more so curiously furious as she explained her dilemma and thirst to unravel the many mysteries behind such rationale and reasoning. She explains when none of their (children's' close friends or associates - "*largely from the white community, both at school and in the neighbourhood, were not part of this 'so-called' police database"* she added. And also admitted she hadn't got a clue how her children ended up getting into that police database in the first place! When she got to know this database is connected to AI in some shape or form, at which point, she admits she had started digging deep to find out more on this (AI) system and on how it works (or not works!). Confessed to figuring out the reasoning behind the dilemma, which she had no answers to-date from the law enforcement agencies nor the responsible authorities, unfortunately.

When inquired about her knowledge and awareness (and her family members') of prospective violations on their fundamental rights and potential legal remedial measures, all participants showed very little awareness of the same. Although they wish there was a transparent process to such systems and procedures and more importantly, a clear procedure to question and scrutinise the process (system) that they admitted having severe social impact although showed very little awareness on either legal implication nor long term ramification as a result.

5.2.2 From a legal standpoint

Academic perspective

The questions aimed at investigating the knowledge and awareness of the participant in AI-related systems, including the impacts it has on society and the rule of law and its ramifications. The respondent whilst admitting to having little knowledge of artificial intelligence and its evolving technical tools, but is rather aware of ADM, and its socio-legal implications to society and the broader global community.

Biasness, impact on rule of law and its broader implications to the society and case for regulatory intervention

The question around AI-biasness, and the impact the AI-systems have on rule of law and its broader implications to the society, the Stockholm-based academic scholar (B1), admitted his acute awareness of this evolving topic especially around the system biasness. Referring to the process and the ways of working concerning AI-based technology tools and such providers (largely private actors); he admits having a great concern around such systems' conformity to core values and the principles of the rule of law, and the compliance aspect due to its transparency concerns.

Citing the question around system biasness (due to its algorithms), and its related practices (data feeds), which is 'inherently' discriminatory in nature and the question of accountability remains a high priority across the industry he claims. He further describes the fact that such systems "aren't entirely transparent, be it the process, the ways of working or their transnational operating models" that cut across different countries and territories, which he believes can potentially have profound socio-legal consequences from a fundamental rights perspective concerning the social impact it has on EU citizens and beyond! He further affirms, the above is despite for example *"privacy laws and laws that protect fundamental rights of the citizens are much stronger in EU compared to many other countries and jurisdictions, for example, the U.S., Russia or China"*.

Furthermore, he explains with this fast-evolving topic concerning AI-related technology, "*the law and the legal framework that governs and regulate is far too sluggish and slow in keeping up pace and catching up with the fast-evolving developments of AI, to be able to reflect the society's needs"*. In this context, citing many examples concerning AI-related tools and practices i.e. that goes against many core values and principles that we all cherish, such as equality amongst citizens, non-discriminatory practices and most importantly the 'presumption of innocence' until proven guilty. Furthermore, citing some of the global cases such as i.e., Apple Inc.'s global tax avoidance litigation cases[45], and Facebook's social media dominance[46], and Tesla's AI-powered autonomous (self-driving) initiatives, he claims, "*are just the tip of the*

[45] The case against the EU "state aid" case against Apple: 13 billion euros out of thin air – Reference

[46] See, Facebook Australian case.

Flynn, K., (2021) *Facebook bans news in Australia as fight with government escalates*. CNN. Reference *iceberg from the real social challenges the society currently faces that he claims "would not just have an impact on today's society, but would have lasting ramifications for many decades to come"!*

Responding to the question around the sufficiency of the regulatory safeguards afforded to the general public against the actions of the state or the private actors, who are responsible for deploying such AI-based systems that can potentially have an impact on rule of law and the broader society. Citing some of the recent EU legislative enactments, for example i.e. GDPR (General Data Protection Regulations), which he asserts focuses on strict compliance and accountability from a data protection standpoint. He argues for example, this piece of legislation "*has re-affirmed the importance of compliance i.e. on obtaining prior consent concerning data collection, obligation for such collected data to be protected, and the right for individuals to be informed about automated decision making etc"*. Another example that was cited, goes further in providing the "*right to request for their own personal data to be deleted off the system"* within a reasonable time period, which he claims are significant developments from an European Union standpoint. Which he argues are some of the regulatory provisions that intend to tackle transparency and accountability to a certain degree. Having said that he asserts, however there is a long way to go when it comes to the question of sufficient regulatory oversight in automating governance, and automated decision making in particular addressing the question of upholding rule of law in this modern and digital era.

He further argues, at present, not fully appreciating the true and full potential of AI-based technologies and their fast-evolving capacities (that cut-across and powers multitude of streams and industries) from i.e. medical industry to ground transport and aviation to agricultural and even space technology. Which he claims, currently and largely left for companies (private actors) to self-regulate with relative and limited regulatory interventions by the state governments across the globe. From a socio-legal standpoint, he affirms there are and potentially could be further non-quantified ramifications beyond the comprehension of the ordinary person. Therefore, he strongly contends *"the governors (i.e. the state, policymakers and perhaps at times, the powerful 'private actors' etc), and the society (the governed) may have to come together aligning themselves and strike a balance in terms of transparency and accountability concerning such fast evolving technological development and expansions"*.

5.2.3 From Industry Practitioner Perspective

Investigating the internal industry (corporate) practices concerning the process, procedures and the ways of working and their (practitioners) level of awareness of social implications. The questions aimed at investigating the internal aspect of industry practices and possibly, how they (practitioners) view the outside world from an industry practitioners' standpoint with regard to AI-related systems (ADM in particular), and the impacts and ramifications to society.

Participants in general have largely responded to having a solid knowledge and background on artificial intelligence and its related tools technically (including digitalization and ADM). One of the common threads observed across the board was the fact that, whilst all participants are well aware of the benefits of such technological breakthrough, in fact showed very limited knowledge of the impact to rule of law and the prospective complications to the society at large. Whilst admitting certain industry practices '*maybe*' in violation of the general principles of rule of law whilst admitting limited knowledge on that sphere affirms that they are acutely aware of social impact to the societies which he claimed against their conscience although the industry isn't doing much about it adding that "*as long as they (companies) hit the revenue targets year-on-year"!*

At the same time, one of the factors however that was admitted, they themselves as industry practitioners, are not too convinced with either the general standards of industry practices (in many areas including the transparency aspect) or its ethical practices concerning industry ways of working. This was confessed by citing some of the recent global cases including high-tech global giants such as Google Inc (firing some of its leading researchers)[47] – where such corporate actions were heavily criticised that lack transparency or accountability by and large. One of the participants acknowledges on the condition of anonymity, that he's not too happy about neither the company (he works for) nor the industry practices in relation to their ways of working (what he pronounced as 'WOW'). The fact that "*they (corporates) act with 'complete impunity' when it comes to self-regulation*", which he claims, "the only interest that matters to corporate agenda is their own (not the public interest)". Hence as long as companies and the big-tech giants in particular keep making colossal amounts of wealth, such practices not going to transform through 'self-regulation'! Hence, he claims "*it's a job for the legislators and regulator to put their thinking caps on before it's too late!"*

[47] Referred to the case mentioned above concerning the outspoken researchers of Google Inc*

06

Conclusion

In conclusion, it could be stated that the above EU legislative initiatives and AI-related strategy proposals are rather encouraging in terms of bridging the long due policy gaps whilst addressing some of the key social challenges. Compared to many other jurisdictions (countries) grappled largely with similar and related social challenges, such strides appear to be made in the right direction. However, in line with the emerging trends of AI, and related tools and its deployment practices, to be able to address not just from today's existing controversies but also envision tomorrow's broader AI-related growing concerns - would undoubtedly require an abundance of complex planning. This needs to be critically looked at more importantly from a socio-legal standpoint particularly on the challenges facing the wider society. Hence one could reasonably argue that the academic community, industry practitioners, and the wider stakeholders including the corporates, state sector and government agencies, policy planners, regulators and legislators should come together in a much more open and collaborative manner for a broader and effective regulatory framework.

Consequently, this research inquiry would conclude by calling for the following 3-point recommendation from a policy perspective, enabling us to pay close attention to detail on addressing some of the most challenging issues based on research results of this inquiry. In line with the current and emerging challenges profoundly discussed in the report, this would enable not only to bridge and narrow the policy gaps gradually, but also address some of the key challenging issues discussed at the social front. The following specific areas could be given due consideration for a broader regulatory framework from a policy planning standpoint, in terms of strengthening transparency and accountability.

(01) **From a responsible industry perspective** - decision makers and controllers (including subcontractors of the private and state sectors), should be held responsible and ultimately held solely or jointly accountable for their actions. This includes but not limited to inaccurate/ un-justifiable data-sets used or AI-based systems employed (including defective algorithms) that has a profound impact on the society that is either biased or discriminatory in nature.

(02) **The scope of the regulator's responsibility** - which should essentially encourage regulators for more proactive intervention from an accountability standpoint.

Dealing with biases: if such AI-based algorithms rely on actual or perceived 'biased data' (in certain shape or form), the degree of accountability that lies upon states or private actors (corporates or individuals) including the business owners, controllers/ processors and sub-contractors, regardless of their size and scale of operation, who should be held responsible for the development and usage of such flawed systems.

Dealing with data breaches: Responsibility to take swift action by the data handlers, on such flaws identified. Firstly, by acknowledging the reported problem. Secondly, to provide reasonable consideration to investigate and report such issues to the industry and the regulators. Finally, to take reasonable steps to rectify such reported problems within a reasonable time.

Accordingly, strengthen the degree of the responsibility to disclose such developments/ decisions made within the organisation(s), that could potentially have an impact to the society and broader ramification to the wider stakeholders from a transparency standpoint.

(03) **From an automated decision-making (ADM) standpoint** - the following could be given due consideration. Primarily a reasonably defined process for the purpose of contesting by way of a robust disputing mechanism concerning ADM. That is potentially afforded to the general public including the most vulnerable / and the potential victims against decisions made by an automated system (i.e.,

(04) with or without human intervention). Which could be followed by an effective and robust remedial or redress process (preferably out of the courts resolution / by way of an alternative dispute resolution) afforded to victims, from a social justice and fairness standpoint.

The above could enable us to focus more on various risk-based assessment models in the interest of wider stakeholders to facilitate an effective and robust regulatory intervention. This above to be made whilst encouraging and nurture innovation and creativity within the industry and preserving business interests around developing and expanding AI-based technologies for the wider good of the society and the humankind.

References

Amnesty International. (2022). Ban dangerous facial recognition technology that amplifies racist policing. Available at: https://www.amnesty.org/en/latest/press-release/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/
[Accessed 12 Aug. 2022].

Alikhademi, K. (2021). A review of predictive policing from the perspective of fairness. SpringerLink. Available at: https://link.springer.com/article/10.1007/s10506-021-09286-4?error=cookies_not_supported&code=7dcc2f74-8f5d-4a77-8be7-7a364053afc6
[Accessed 01 Aug. 2022].

Association for the Advancement of Artificial Intelligence. (2022). AAA. Available at: https://www.aaai.org/ [Accessed 29 Sep. 2022].

Anderson, T. (2013). *How to safeguard your data in cyberspace.* The Guardian. Available at: https://www.theguardian.com/media-network/media-network-blog/2013/feb/11/cyber-attack-data-mobile-security [Accessed 23 May. 2022].

Bartneck, C. (2021). Privacy Issues of AI. Springer Publication.

BBC News 2 (2018). Facebook hires former deputy PM Sir Nick Clegg. Available at: https://www.bbc.com/news/technology-45913587 [Accessed 11 Aug. 2022].

Benjamin, R. (2019). *Race After Technology (1st ed.)*. Wiley. Available at: https://www.perlego.com/book/1536396/race-after-technology-abolitionist-tools-for-the-new-jim-code-pdf [Accessed 11 Aug. 2022].

Brown, G. & Yule, G. (1985) *Discourse Analysis*. Cambridge, Cambridge University Press.

Browne, R. (2022). Fines for breaches of EU privacy law spike sevenfold to $1.2 billion, as Big Tech bears the brunt. CNBC. Available at: https://www.cnbc.com/2022/01/18/fines-for-breaches-of-eu-gdpr-privacy-law-spike-sevenfold.html [Accessed 19 Sep. 2022].

Bryman, A., & Bell, E. (2011) *Business research methods*. Oxford University Press

Chouliaraki, L., & Fairclough, N. (2021). *Discourse in Late Modernity*. Discourse in Late Modernity, Rethinking Critical Discourse Analysis. Oxford Brookes University.

Chee, F. Y. (2020). Tech giants face fines or even break-up if they breach new rules: EU's Breton. U.S. Available at: https://www.reuters.com/article/us-eu-tech-rules-idCAKBN2852NI [Accessed 29 Aug. 2022].

DW is a German public broadcast service. (2022, May 19). *Google, Facebook, Amazon - The rise of the mega-corporations | DW Documentary* [Video]. DW Documentary. Available at: https://www.youtube.com/watch?v=Dy8ogOaKk4Y [Accessed 01 Oct. 2022].

European Commission. (2021, April). *A European Strategy for Artificial Intelligence*. Available at: https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf? [Accessed 01 Oct. 2022].

European Commission (2018) *Coordinated Plan on Artificial Intelligence*. EU Lex Europa. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52018DC0795&rid=3#:~:text=Artificial%20Intelligence%20refers%20to%20systems,or%20speak%20with%20digital%20assistants. [Accessed 02 Sep. 2022].

European Commission. (2019, February). *White Paper. On Artificial intelligence* – An European approach to excellence and trust. Available at: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf [Accessed 23 Aug. 2022].

European Parliamentary. (2020, June). *Artificial intelligence: How does it work, why does it matter, and what can we do about it?* European Parliamentary Research Service. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU(2020)641547_EN.pdf [Accessed 19 July. 2022].

Fairclough, N. (2003) *Analysing Discourse:* Textual Analysis for Social Research, Routledge, 2003.

Flynn, K., (2021) Facebook bans news in Australia as fight with government escalates. CNN. Available at: https://edition.cnn.com/2021/02/17/media/facebook-australia-news-ban/index.html [Accessed 23 Oct. 2022].

Fritsch, R., & Thomas, N. (2019). AI and Automated Decision-Making: Impact on Access to Justice and Legal Aid. Available at: https://www.lco-cdo.org/wp-content/uploads/2019/06/LCO-ILAG-Paper-AI-Legal-Aid-and-Access-to-Justice-June-3-2019.pdf [Accessed 02 Sep. 2022].

Fonseka, C. (2017). Hold Artificial Intelligence Accountable. Science in the News Available at: https://sitn.hms.harvard.edu/flash/2017/hold-artificial-intelligence-accountable/ [Accessed 02 Oct. 2022].

Foresight. (2021, April). Automated decision-making impacting society | Knowledge for policy. European Commission. Available at: https://knowledge4policy.ec.europa.eu/foresight/automated-decision-making-impacting-society_en [Accessed 12 Oct. 2022].

Fox, J. (2007). The uncertain relationship between transparency and accountability. Development in Practice, 17(4-5), 663–671. Available at: https://doi.org/10.1080/09614520701469955 [Accessed 19 Aug. 2022].

Gary, A., & Holmes, D. (2020). Researcher Positionality -A Consideration of Its Influence and Place in Qualitative Research -A New Researcher Guide. Available at: https://files.eric.ed.gov/fulltext/EJ1268044.pdf [Accessed 22 Sep. 2022].

Gardner, A., Smith, A.L., Steventon, A. et al. (2022). *Ethical funding for trustworthy AI:* proposals to address the responsibilities of funders to ensure that projects adhere to trustworthy AI practice. AI Ethics 2, 277–291. Available at: https://doi.org/10.1007/s43681-021-00069-w [Accessed 12 Oct. 2022].

Goodman, E. P. (2018). *Algorithmic Transparency for the Smart City*. Yale.

Guillemin, M., & Gillam, L. (2004). *Ethics, Reflexivity, and "Ethically Important Moments*" in Research. Qualitative Inquiry, 10(2), 261–280. Available at: https://doi.org/10.1177/1077800403262360 [Accessed 22 Oct. 2022].

Greenstein, L. C. A. S., & Sannerholm, R. (2022). *Responsibility and Accountability: AI, Governance, and the Rule of Law*. Law in the Era of Artificial Intelligence. eddy.se ab.

Goodman, B. (2017, October 2). European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation" | AI Magazine. AI Magazine Publication. Available at: https://ojs.aaai.org/index.php/aimagazine/article/view/2741 [Accessed 11 Aug. 2022].

Harwell, D. (2019). *Oregon became a testing ground for Amazon's facial-recognition policing. But what if Rekognition gets it wrong?* Washington Post. Available at: https://www.washingtonpost.com/technology/2019/04/30/amazons-facial-recognition-technology-is-supercharging-local-police/
[Accessed 05 Aug. 2022].

Hildebrandt, M. (2016). *Smart Technologies and the End(s) of Law*: Novel Entanglements of Law and Technology. Edward Elgar Publishing.

Hill, K. (2021, November 2). The Secretive Company That Might End Privacy as We Know It. The New York Times. Available at: https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html [Accessed 01 Oct. 2022].

Jasanoff, S. (2016). The Ethics of Invention: Technology and the Human Future. W. W. Norton Company.

Jackson, L. (2021, May 25). *Shoshana Zuboff Explains Why You Should Care About Privacy*. The New York Times. Available at: https://www.nytimes.com/2021/05/21/technology/shoshana-zuboff-apple-google-privacy.html [Accessed 01 Sep. 2022].

Keating, M., & Donatella Porta, D. (2008). *Approaches and Methodologies in the Social Sciences*: A Pluralist Perspective. Cambridge University Press.

Kent, R. (2020). Interview: How Policing in One US City Hurts Black and Poor Communities. Human Rights Watch. Available at: https://www.hrw.org/news/2019/09/12/interview-how-policing-one-us-city-hurts-black-and-poor-communities [Accessed 30 Aug. 2022].

Koene, A., Clifton, C., Hatada, Y., Webb, H., & Richardson, R. (2019). *A governance framework for algorithmic accountability and transparency* (Study No. PE 624.262) Panel for the Future of Science and Technology, Scientific Foresight Unit (STOA), European Parliamentary Research Service.

Kudumala, A., Ressler, D., & Miranda, W. (2022, November*). Scaling up AI across the life sciences value chain*. Deloitte Insights. Available at: https://www2.deloitte.com/us/en/insights/industry/life-sciences/ai-and-pharma.html [Accessed 29 Aug. 2022].

Larsson, S. & Heintz, F. (2020). Transparency in artificial intelligence. Internet Policy Review, 9(2). Available at: https://policyreview.info/concepts/transparency-artificial-intelligence [Accessed 03 Oct. 2022].

Lau, T. (2020). *Predictive Policing Explained*. Brennan Centre for Justice. Available at: https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained [Accessed 19 Sep. 2022].

Laub, Z. (2019, June 7). Hate Speech on Social Media: Global Comparisons. Council on Foreign Relations. Available at: https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons [Accessed 19 Aug. 2022].

LCO-CDO (2020). *AI, ADM and the Justice System*. LCO-CDO. Available at: https://www.lco-cdo.org/en/our-current-projects/ai-adm-and-the-justice-system/ [Accessed 21 Sep. 2022].

Lee, D., (2018). *Facebook's data-sharing deals exposed*. BBC World. Available at: https://www.bbc.com/news/technology-46618582 [Accessed 11 Aug. 2022].

Lee, Y. (2021). Using Big Data to Prevent Crime: Legitimacy Matters. SpringerLink. Available at: https://link.springer.com/article/10.1007/s11417-021-09353-4?error=cookies_not_supported&code=f73b5708-08e7-406b-8133-07c14e140ef4 [Accessed 11 May. 2022].

Leal Filho, W., Yang, P., Eustachio, J. H. P. P., Azul, A. M., Gellers, J. C., Gielczyk, A., Dinis, M. A. P., & Kozlova, V. (2022). *Deploying digitalisation and artificial intelligence in sustainable development research*. Environment, Development and Sustainability. Available at: https://doi.org/10.1007/s10668-022-02252-3 [Accessed 11 Oct. 2022].

Levin, S. (2017). *Uber's scandals, blunders and PR disasters.* The Guardian. Available at: https://www.theguardian.com/technology/2017/jun/18/uber-travis-kalanick-scandal-pr-disaster-timeline [Accessed 11 May. 2022].

McCarthy, J. (2012). What is AI? / Basic Questions. Stanford. Available at:: http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html#:%7E:text=A.,methods%20that%20are%20biologically%20observable. [Accessed 23 May. 2022].

Margetts, H. (2011). *The internet and transparency*. The Political Quarterly, 82(4), 518–521. Available at: https://doi.org/10.1111/j.1467-923X.2011.02253.x [Accessed 11 Aug. 2022].

Marsh, S. (2019). UK police use of computer programs to predict crime sparks discrimination warning. The Guardian. Available at: https://www.theguardian.com/uk-news/2019/feb/03/police-risk-racial-profiling-by-using-data-to-predict-reoffenders-report-warns [Accessed 17 June. 2022].

Merton, R. K. (1975). *Thematic Analysis in Science: Notes on Holton's Concept.* Science, 188(4186), 335–338. Available at: http://www.jstor.org/stable/1739319 [Accessed 17 Aug. 2022].

Minevich, M. (2020). *How To Combat The Dark Side Of AI.* Forbes. Available at: https://www.forbes.com/sites/markminevich/2020/02/28/how-to-combat-the-dark-side-of-ai/?sh=56c3f89d174b [Accessed 11 Oct. 2022].

Najibi, A. (2020). Racial Discrimination in Face Recognition Technology. Science in the News. Available at: https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/ [Accessed 12 July. 2022].

Nowell L.S, Norris J.M, White D.E, Moules N.J. (2017) *Thematic Analysis*: Striving to Meet the Trustworthiness Criteria. International Journal of Qualitative Methods. December. Sage Publication

Pink, S. (2022). *Everyday Automation*. Routledge.

Riekeles, G. (2022). *I saw first-hand how US tech giants seduced the EU – and undermined democracy*. The Guardian. Available at: https://www.theguardian.com/commentisfree/2022/jun/28/i-saw-first-hand-tech-giants-seduced-eu-google-meta [Accessed 29 Sep. 2022].

Schuilenburg, M., & Hall, G. (2015). The Securitization of Society: Crime, Risk, and Social Order. NYU Press. Available at: http://www.jstor.org/stable/j.ctt15r4044 [Accessed 22 Aug. 2022].

Shapiro, S. J. (2007). The "Hart-Dworkin" Debate: A Short Guide for the Perplexed. *SSRN Electronic Journal*. Available at: https://doi.org/10.2139/ssrn.968657 [Accessed 04 Oct. 2022].

Shaw, J. (2018, December 6). Artificial Intelligence and Ethics. Harvard Magazine. Available at: https://www.harvardmagazine.com/2019/01/artificial-intelligence-limitations [Accessed 03 Aug. 2022].

Sioli, L. (2021). A European Strategy for Artificial Intelligence. Available at: https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf [Accessed 17 Sep. 2022].

Solove, D. J. (2006). A Model Regime of Privacy Protection (Version 2.0). SSRN Electronic Journal. Available at: https://doi.org/10.2139/ssrn.699701 [Accessed 05 May. 2022].

Statista. (2020). Biggest data breach fines and settlements worldwide (n.d.). Available at:https://www.statista.com/statistics/1170520/worldwide-data-breach-fines-settlements/ [Accessed 09 Aug. 2022].

Tamanaha, B. (2004). *On the Rule of Law: History, Politics, Theory*. Cambridge University Press.

Taddonio, P. (2022, August 24). How China's Government Is Using AI on Its Uighur Muslim Population. FRONTLINE. Available at: https://www.pbs.org/wgbh/frontline/article/how-chinas-government-is-using-ai-on-its-uighur-muslim-population/ [Accessed 03 June. 2022].

Team, A. N. P. (2018, August 13). *The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems*. Access Now. Available at: https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/ [Accessed 03 Sep. 2022].

Teich, D. A. (2020). Artificial Intelligence And Data Privacy – Turning A Risk Into A Benefit. Forbes. Available at: https://www.forbes.com/sites/davidteich/2020/08/10/artificial-intelligence-and-data-privacy--turning-a-ri sk-into-a-benefit/?sh=758b05986a95 [Accessed 02 Aug. 2022].

The Economist. (2018). *The sunny and the dark side of AI*. Available at: https://www.economist.com/special-report/2018/03/28/the-sunny-and-the-dark-side-of-ai [Accessed 20 Aug. 2022].

Thomson, A & Bodoni, S (2020) Google CEO Thinks AI Will Be More Profound Change Than Fire. (2020, January 22). Bloomberg.com. Available at: https://www.bloomberg.com/news/articles/2020-01-22/google-ceo-thinks-ai-is-more-profound-than-fire#xj4y7vzkg [Accessed 07 June. 2022].

Towers-Clark, C. (2018). *Can We Make Artificial Intelligence Accountable?* Forbes. Available at: https://www.forbes.com/sites/charlestowersclark/2018/09/19/can-we-make-artificial-intelligence-accountable/?sh=68f5b660364e [Accessed 17 May. 2022].

Ufert, F. (2020, September 20). *AI Regulation Through the Lens of Fundamental Rights*: How Well Does. European Papers. Available at: https://www.europeanpapers.eu/en/europeanforum/ai-regulation-through-the-lens-of-fundamental-rights [Accessed 12 May. 2022].

Waddell, K. (2016). *How Algorithms Can Bring Down Minorities' Credit Scores.* The Atlantic. Available at: https://www.theatlantic.com/technology/archive/2016/12/how-algorithms-can-bring-down-minorities-credit-scores/509333/ [Accessed 29 July. 2022].

Walsh, C. (2021). Solving racial disparities in policing. Harvard Gazette. Available at: https://news.harvard.edu/gazette/story/2021/02/solving-racial-disparities-in-policing/ [Accessed 13 Aug. 2022].

World Justice Project. (2021, April). The World Justice Project Rule of Law Index 2021. Available at: https://worldjusticeproject.org/sites/default/files/documents/WJP-INDEX-21.pdf [Accessed 12 Aug. 2022].

Wu, J. (2019). Empathy in Artificial Intelligence. Forbes. Retrieved December 22, 2022, Available at: https://www.forbes.com/sites/cognitiveworld/2019/12/17/empathy-in-artificial-intelligence/?sh=38cd9cd63270 [Accessed 07 Apr. 2022].

Swire, P., & Woo, J. (2018). Privacy and Cybersecurity Lessons at the Intersection of the Internet of Things and Police Body-Worn Cameras. Papers.ssrn.com. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3168089 [Accessed 17 May. 2022].

Zuboff, S. (2019). *Age of Surveillance Capitalism*. Shoshana Zuboff (Ebok). Bokus.com. Available at: https://www.bokus.com/bok/9781782832744/age-of-surveillance-capitalism/?msclkid=aeeffcf548d9112c228f9e003b453f40&utm_source=bing&utm_medium=cpc&utm_campaign=Search%20%7C%20DSA%20%7C%20All&utm_term=bokus&utm_content=DSA%20-%20All [Accessed 12 June. 2022].

Appendix

List of Figures

Appendix 01

Appendix 02

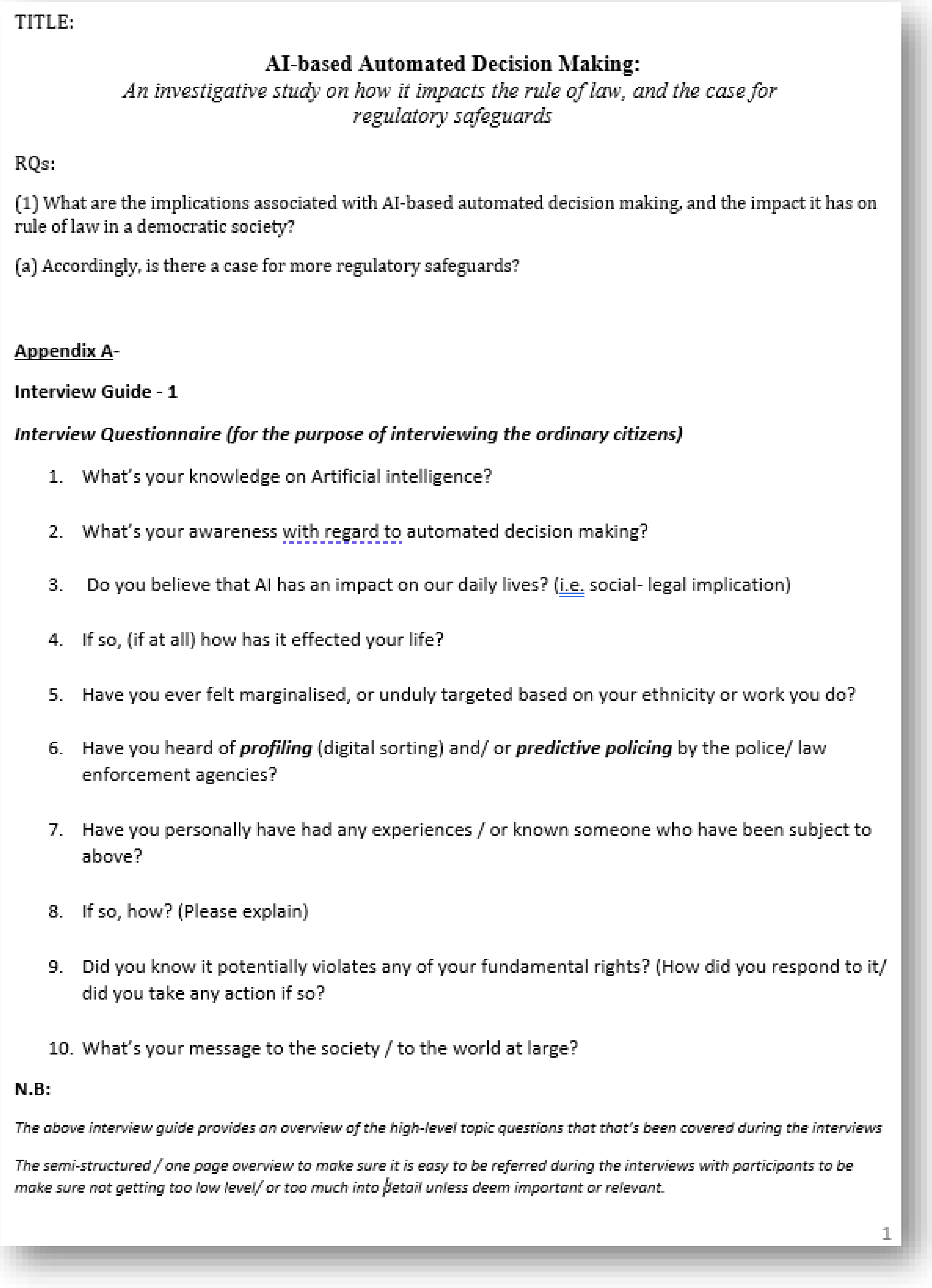
Table 01 - *Overview of the interviewed respondents (including 1 to 1 interviews and the focus group) part of primary data collection*

Category of Interviews/ FG	Date	Interview type/ (code)	Role of the Interviewee	Location
		FG		
Focus Group (FG) (virtual)	16/ 02/ 2022	**A1**	**Parent** of a child victim	Manchester, UK
	16/ 02/ 2022	**A2**	**Parent** of a child victim	Birmingham, UK
	16/ 02/ 2022	**A3**	**Family member** of the victim	Malmö, Sweden
		Interviews (Face-to-face)		
Interview (1-to-1)	02/03/2022	Face-to-face **B1**	**Academic Scholar** / Legal Counsel	Stockholm, Sweden
Interview (1-to-1)	09/03/2022	1-to-1 **C1**	**Industry practitioner** / AI expertise	Boise, ID (USA)
Interview (1-to-1)	09/03/2022	1-to-1 **C2**	**Industry practitioner** / AI & Transformation expertise	Boise, ID (USA)
Interview (1-to-1)	18/04/2022	Face-to-face **C3**	**Industry practitioner** / Transformation expertise	Malmö, Sweden

Table 01: *Overview of the interviewed respondents (part of primary data collection)*

Appendix 03

Sample forms used during the interview process



TITLE:

AI-based Automated Decision Making:
An investigative study on how it impacts the rule of law, and the case for regulatory safeguards

RQs:

(1) What are the implications associated with AI-based automated decision making, and the impact it has on rule of law in a democratic society?

(a) Accordingly, is there a case for more regulatory safeguards?

Appendix A-

Interview Guide - 1

Interview Questionnaire (for the purpose of interviewing the ordinary citizens)

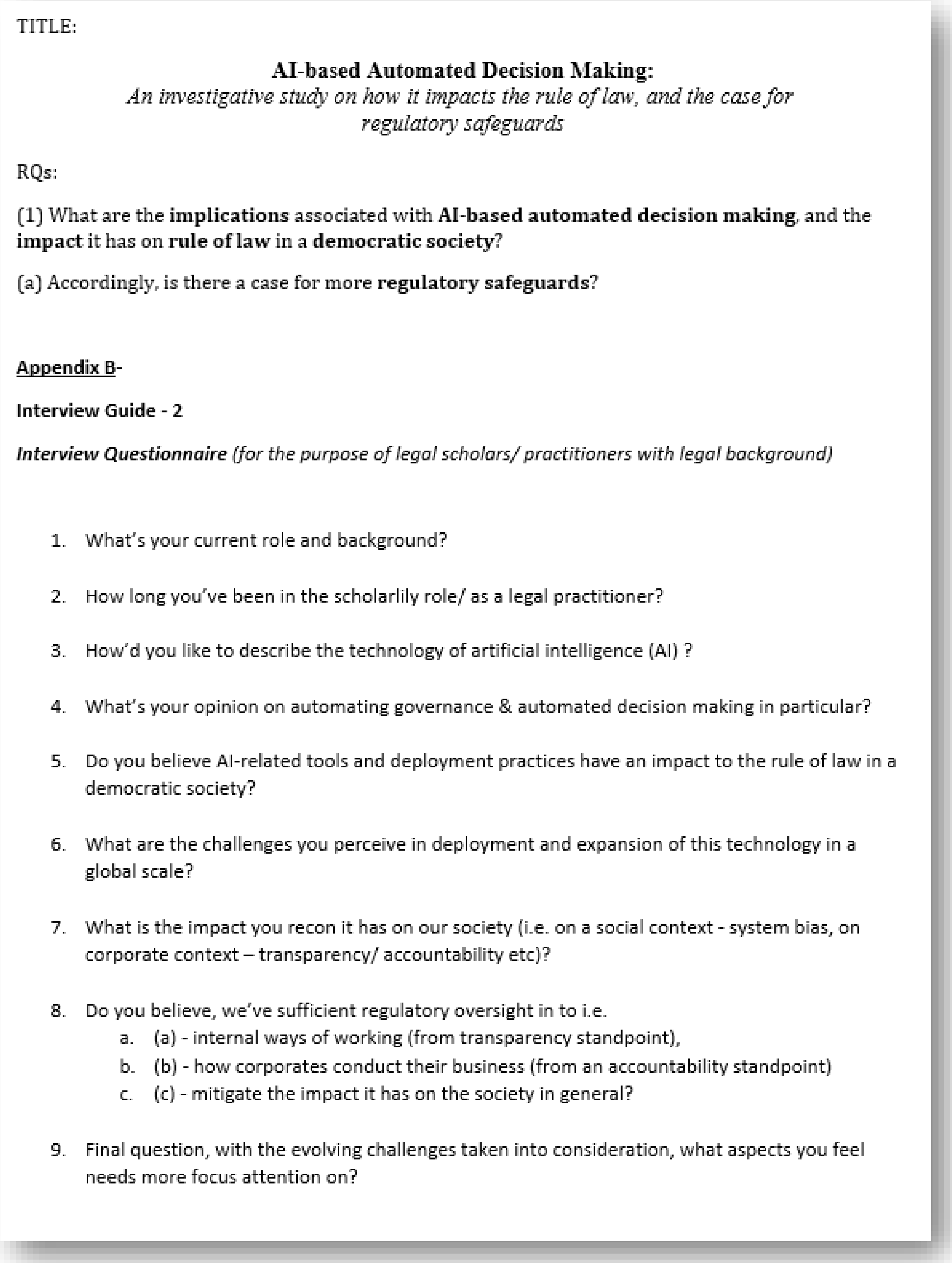
1. What's your knowledge on Artificial intelligence?
2. What's your awareness with regard to automated decision making?
3. Do you believe that AI has an impact on our daily lives? (i.e. social- legal implication)
4. If so, (if at all) how has it effected your life?
5. Have you ever felt marginalised, or unduly targeted based on your ethnicity or work you do?
6. Have you heard of ***profiling*** (digital sorting) and/ or ***predictive policing*** by the police/ law enforcement agencies?
7. Have you personally have had any experiences / or known someone who have been subject to above?
8. If so, how? (Please explain)
9. Did you know it potentially violates any of your fundamental rights? (How did you respond to it/ did you take any action if so?
10. What's your message to the society / to the world at large?

N.B:

The above interview guide provides an overview of the high-level topic questions that that's been covered during the interviews

The semi-structured / one page overview to make sure it is easy to be referred during the interviews with participants to be make sure not getting too low level/ or too much into detail unless deem important or relevant.

1



TITLE:

AI-based Automated Decision Making:
An investigative study on how it impacts the rule of law, and the case for regulatory safeguards

RQs:

(1) What are the **implications** associated with **AI-based automated decision making**, and the **impact** it has on **rule of law** in a **democratic society**?

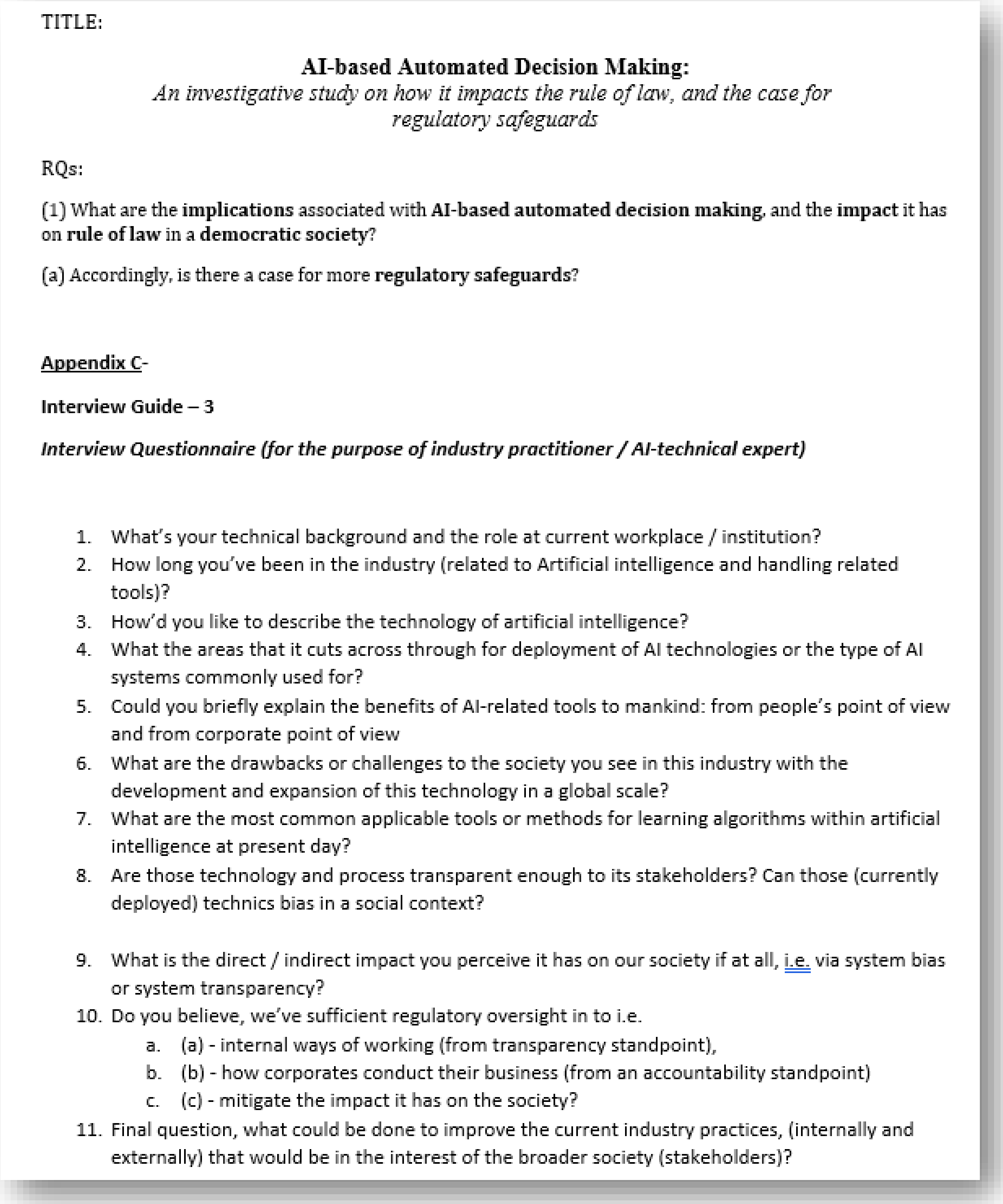
(a) Accordingly, is there a case for more **regulatory safeguards**?

Appendix B-

Interview Guide - 2

Interview Questionnaire *(for the purpose of legal scholars/ practitioners with legal background)*

1. What's your current role and background?
2. How long you've been in the scholarlily role/ as a legal practitioner?
3. How'd you like to describe the technology of artificial intelligence (AI) ?
4. What's your opinion on automating governance & automated decision making in particular?
5. Do you believe AI-related tools and deployment practices have an impact to the rule of law in a democratic society?
6. What are the challenges you perceive in deployment and expansion of this technology in a global scale?
7. What is the impact you recon it has on our society (i.e. on a social context - system bias, on corporate context – transparency/ accountability etc)?
8. Do you believe, we've sufficient regulatory oversight in to i.e.
    a. (a) - internal ways of working (from transparency standpoint),
    b. (b) - how corporates conduct their business (from an accountability standpoint)
    c. (c) - mitigate the impact it has on the society in general?
9. Final question, with the evolving challenges taken into consideration, what aspects you feel needs more focus attention on?



TITLE:

AI-based Automated Decision Making:
An investigative study on how it impacts the rule of law, and the case for regulatory safeguards

RQs:

(1) What are the **implications** associated with **AI-based automated decision making**, and the **impact** it has on **rule of law** in a **democratic society**?

(a) Accordingly, is there a case for more **regulatory safeguards**?

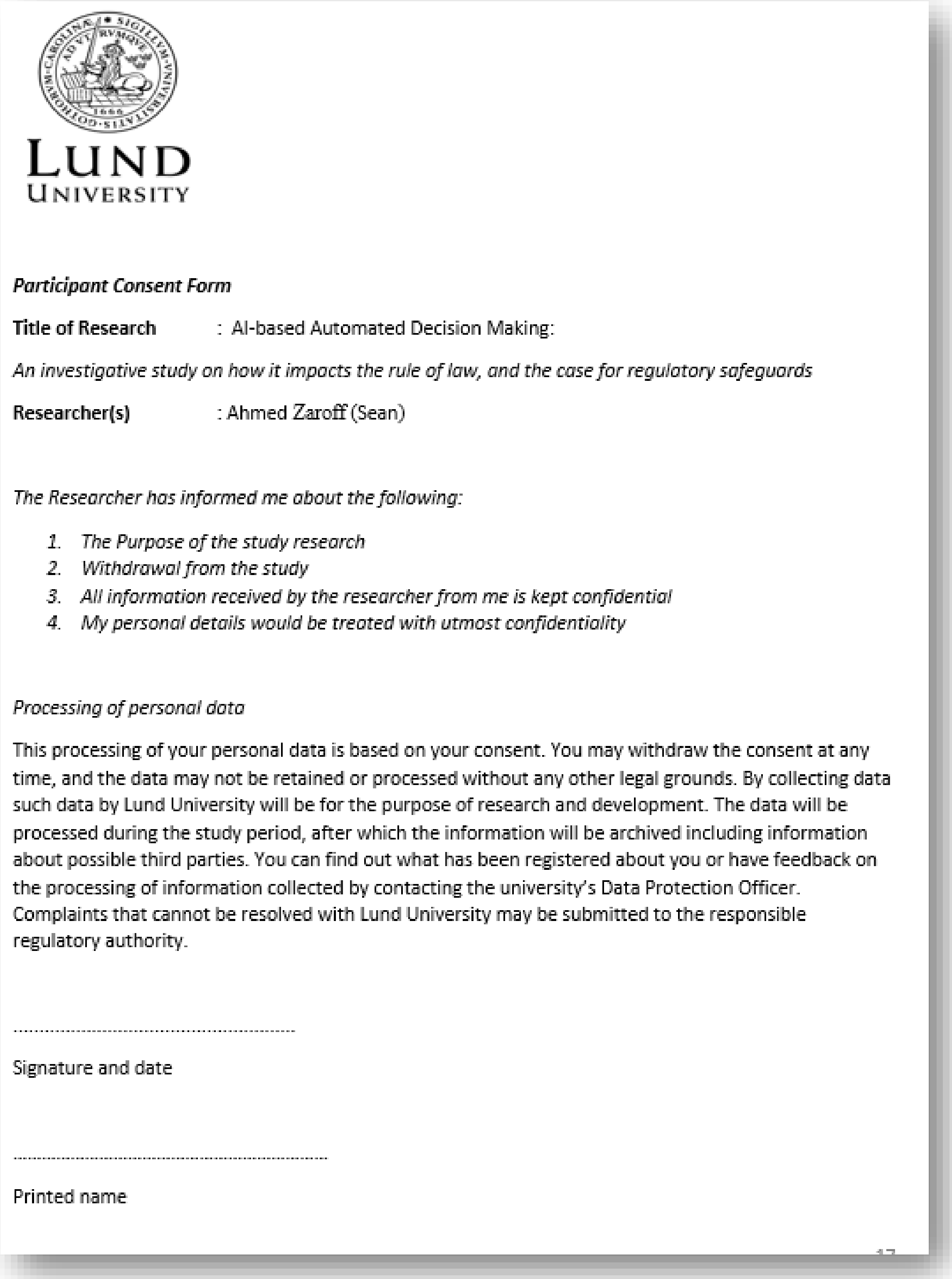
Appendix C-

Interview Guide – 3

Interview Questionnaire (for the purpose of industry practitioner / AI-technical expert)

1. What's your technical background and the role at current workplace / institution?
2. How long you've been in the industry (related to Artificial intelligence and handling related tools)?
3. How'd you like to describe the technology of artificial intelligence?
4. What the areas that it cuts across through for deployment of AI technologies or the type of AI systems commonly used for?
5. Could you briefly explain the benefits of AI-related tools to mankind: from people's point of view and from corporate point of view
6. What are the drawbacks or challenges to the society you see in this industry with the development and expansion of this technology in a global scale?
7. What are the most common applicable tools or methods for learning algorithms within artificial intelligence at present day?
8. Are those technology and process transparent enough to its stakeholders? Can those (currently deployed) technics bias in a social context?

9. What is the direct / indirect impact you perceive it has on our society if at all, i.e. via system bias or system transparency?
10. Do you believe, we've sufficient regulatory oversight in to i.e.
    a. (a) - internal ways of working (from transparency standpoint),
    b. (b) - how corporates conduct their business (from an accountability standpoint)
    c. (c) - mitigate the impact it has on the society?
11. Final question, what could be done to improve the current industry practices, (internally and externally) that would be in the interest of the broader society (stakeholders)?



LUND
UNIVERSITY

Participant Consent Form

Title of Research : AI-based Automated Decision Making:

An investigative study on how it impacts the rule of law, and the case for regulatory safeguards

Researcher(s) : Ahmed Zaroff (Sean)

The Researcher has informed me about the following:

1. *The Purpose of the study research*
2. *Withdrawal from the study*
3. *All information received by the researcher from me is kept confidential*
4. *My personal details would be treated with utmost confidentiality*

Processing of personal data

This processing of your personal data is based on your consent. You may withdraw the consent at any time, and the data may not be retained or processed without any other legal grounds. By collecting data such data by Lund University will be for the purpose of research and development. The data will be processed during the study period, after which the information will be archived including information about possible third parties. You can find out what has been registered about you or have feedback on the processing of information collected by contacting the university's Data Protection Officer. Complaints that cannot be resolved with Lund University may be submitted to the responsible regulatory authority.

....

Signature and date

....

Printed name