# The Making of Trustworthy and Competitive Artificial Intelligence

A Critical Analysis of the Problem Representations of AI in the European Commission's AI Policy

Emilia Kallioinen

Master's Thesis

University of Helsinki

Faculty of Social Sciences

Global Politics and Communication/ Governance, Organizations, and Communication track

April 2022

| Tiedekunta – Fakultet – Faculty<br>Faculty of Social Sciences | | Koulutusohjelma – Utbildningsprogram – Degree Programme<br>Global Politics and Communication |
|---|---|---|
| Tekijä – Författare – Author<br>Emilia Johanna Kallioinen | | |
| Työn nimi – Arbetets titel – Title<br>The Making of Trustworthy and Competitive Artificial Intelligence – A Critical Analysis of the Problem Representations of AI in the European Commission's AI Policy | | |
| Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track<br>Governance, Organizations and Communication | | |
| Työn laji – Arbetets art – Level<br>Master's Thesis | Aika – Datum – Month and year<br>April 2022 | Sivumäärä – Sidoantal – Number of pages<br>68 |

Tiivistelmä – Referat – Abstract

Artificial intelligence (AI), as a constantly developing technology that is difficult to define, strains a society not prepared for its impact. On the other hand, AI represents the future and comes with many opportunities. The European Commission has taken both views into account in its policy for AI, the European approach to AI. The European Commission's AI policy, that introduces a regulation-based approach to AI as the first policy initiative in the world, offers a timely and intriguing topic of study.

This thesis critically examines how AI is represented as a problem in the European Commission's policy over the course of a four-year time frame from 2018 to 2021. It uses a combined set of methods: qualitative content analysis together with Carol Bacchi's WPR approach to inspect five selected European Commission's policy documents. Four of these policy documents are communication papers with an additional white paper. With the help of qualitative content analysis, the main repetitive themes of AI challenges and opportunities are teased out. The WPR approach is used to examine the progression of the AI policy and analyze the problem representations found in policy. Research questions are the following: how has the European Commission's policy on AI come about and how has AI been represented as a policy problem by the European Commission?

The thesis presents the formation of the AI policy by going through policy documents over the period of four years. Additionally, the thesis demonstrates how the Commission's AI policy is one piece of the puzzle that is EU digital politics aiming for technological sovereignty. From the Commission's problem representation of AI, the challenges and opportunities, it is possible to analyze the implicit representations of AI in policy. Although, the policy highlights trustworthiness and competitiveness through its regulatory actions there are other aspects present as well. AI has been represented in policy through eight perspectives, including safety and security, ethical, legal, competitiveness, AI leadership, socioeconomic, ecological, and education. All perspectives rationalize ways for AI to be embraced inside the European Union borders and participate in the shaping of how AI is to be approached. The analysis of each category shows that issues related to safety and security, ethical, legal, competitiveness, and AI leadership seem to stand out whereas socioeconomic, ecological, and education matters are not as strongly stressed. Overall, this thesis has demonstrated how AI has been represented as a problem in the European Commission's policy.

| Avainsanat – Nyckelord – Keywords<br>Artificial intelligence, European Commission, regulation, WPR approach, qualitative content analysis |
|---|
| Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors<br>Tero Erkkilä |
| Säilytyspaikka – Förvaringställe – Where deposited<br><br>Helsingin yliopiston kirjasto, Helsingfors universitets bibliotek, Helsinki University Library |
| Muita tietoja – Övriga uppgifter – Additional information |

# Table of Contents

# 1. INTRODUCTION

Today applications of artificial intelligence (AI) are present in various domains of life such as transport, healthcare, entertainment, security, manufacturing, and agriculture just to name a few. More opportunities are envisioned as AI is a fast-advancing field with huge potential to help solve many problems. On the downside, AI also has potential to harm if not dealt with appropriately. As promising as AI is, it does pose threats along with the opportunities to current and future applications. With AI gaining momentum, the world is now awakening to ponder on its effects – may they be bad or good – and possible actions or solutions to prevent problems and invest in positive outcomes.

This is an opportune time to examine how AI is represented in the European Union as the European Commission has released the first-ever initiative in the world for regulating AI (European Commission, 2021a). The proposal, Artificial Intelligence Act (AIA), stipulates how AI should be deployed inside the European borders and paves the way for EU as a pioneer in matters related to AI and automated decision-making. I am interested in the buildup of this proposal for governing AI.

This thesis will attempt to identify policy problems in the European approach to AI. The European approach to AI is the European Commission's policy on AI. By focusing on policy problems, I will try to unpack the representation of AI constructed by the European Commission. I will also distinguish the challenges and opportunities that EU acknowledges for AI to garner an overview of different perspectives on AI. What kind of problems does the Commission associate with AI? What is the EU approach to the specific problem? What underlying assumptions are proposed through the problem? And how does policy, in fact, shape the problem itself? In this thesis, I seek to raise up those issues that the Commission identifies as important or needy of attention in regards AI. These problem representations are important to review because they can reflect on the overall understanding of AI and the framing of AI in the broader sense. Through policy they may define how AI is approached in the European Union and elsewhere in the future.

## 1.1. Research Topic and Questions

Like stated earlier, this thesis intends to analyze how AI is represented in the European approach to AI. The aim is to recognize and analyze different problem representations of AI

constructed by the European Commission. Additionally, this thesis seeks to form a conception of how the policy has come together.

My research questions are the following:

1. How has the European Commission's policy on AI come about?
2. How has AI been represented as a policy problem by the European Commission?

The analysis of the European approach to AI will be done by assessing selected significant policy documents that have emerged before and at the time of the proposed act to regulate AI. The focus of the thesis is in identifying the issues found in the policy documents that have led to the regulatory proposal. The thesis approaches the problem representations of AI by breaking down the policy, questioning the construction of problems found in policy through the framework "What's the problem represented to be?" (WPR) provided by Carol Bacchi (2009).

As an emerging technology, AI is often approached through the challenges and opportunities associated with it. These challenges and opportunities, or problem representations as Bacchi (2009) would call them, are the objective of this thesis. The way that these problems related to AI are conceptualized and understood have effect in the proceedings taken to acknowledge them. Problem representations evidently shape the policy that Commission is now seeking through its regulatory proposal and other initiatives. The Commission makes decisions and implements policy to solve the problems that it recognizes in AI. The policy thus does not necessarily fit the whole reality of the situation but is rather one view, or representation, of the issue.

There is value in understanding how policy on AI is constructed by the European Commission. It can foster transparency for policy making and contribute to the general perception of AI held by the public. AI policy does not only have possible legal implications, but it can also impact how AI is and will be conceived or treated in the future. This thesis will hopefully offer insight on the range of policy problems that the European Commission recognizes in AI, in addition, to sketching an outline of how the policy has formed.

## 1.2. On Defining Artificial Intelligence

To understand what is prompted when referring to AI, I will dedicate this section to unpack what AI is. I will elaborate on the difficulty of defining it and the implications this may involve when it comes to the general understanding of AI and policy.

Although the birth of AI traces back to the 1950's, artificial intelligence is yet to have a coherent, widely accepted scientific definition (e.g., Monett & Lewis, 2017; Wang 2019). What complicates defining AI is the lack of consensus on a standard definition of intelligence amongst psychologists themselves as well as AI researchers (Legg & Hutter, 2007). In addition to the lack of consensus on the definition of intelligence, the advancements in the field of AI have taken such strides in the years since the 1950's that a single definition has failed to capture the whole essence of what AI is. A phenomenon called the "AI effect" or the "odd paradox" is used to describe how AI constantly renews itself as new technology emerges (McCorduck, 2004 cited in Stone et al., 2016). In short, this means that technology once considered as AI loses its status as intelligent the moment that newer technology appears and replaces it. Thus, the definition that was accurate yesterday is no longer valid today making it difficult to capture the essence of AI. Furthermore, AI has been categorized into different stages based on its abilities of performance: artificial narrow intelligence (ANI), artificial general intelligence (AGI), and artificial super intelligence (ASI) (e.g. Kaplan & Haenlein, 2019). General intelligence is perceived as an intermediate stage between the artificial intelligence as we know it today, which is narrow, and artificial super intelligence that may possibly someday in the future outperform humans (Oliveira, 2017). In addition, AI can be classified by its cognitive, emotional, and social competencies (Kaplan & Haenlein, 2019). The range of categorization and classification of AI may for one part complicate defining AI (Kaplan & Haenlein, 2020) as well as the fact that from a technical point of view AI is not one technology but a variety of techniques and subdisciplines (Stone et al., 2016).

Regardless the difficulty of defining AI and how it is often used as "a collective term for a wide range of technologies or an abstract large-scale phenomenon" (Hagendorff, 2020, p. 111), I will now provide a more expressive definition of AI proposed by Kaplan and Haenlein to elaborate what is meant by it. Kaplan and Haenlein (2019) have defined AI as "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (p. 17). This definition encapsulates the importance of data and machine learning for AI but distinguishes AI as a

broader concept than, for example, just Internet of Things (IoT), big data or machine learning. Kaplan and Haenlein (2019) point out that AI may use "external information obtained through IoT or other big data sources as an input for identifying underlying rules and patterns by relying on approaches from machine learning" (p. 17), but machine learning alone does not "cover a system's ability to perceive data or control, move, and manipulate objects based on learned information be it a robot or another connected device" (Kaplan & Haenlein, 2019, p. 17) like AI does.

The difficulty of defining AI also challenges how AI is approached in other areas. For example, the public perception of AI is quite versatile (Fast & Horvitz, 2017), possibly partly due to the inconsistency in its definition. The inconsistent definition of intelligence also makes it "difficult for policy makers to assess what AI systems will be able to do in the near future, and how the field may get there" (Bhatnagar et al, 2018, p. 118). Similar observation is raised by Kaplan and Haenlein (2019) as they ponder what consequences a broad definition as opposed to a narrow definition would have for legal AI purposes. For this reason, Bhatnagar et al. (2018) have argued for an atlas of intelligence that provides the "the tools to properly evaluate, compare and classify AI systems" (p. 117). Bhatnagar's et al. vision may become more and more essential as the field continues to develop and have more concrete societal repercussions.

# 2. RELEVANT RESEARCH: THE CURRENT STATE OF AI GOVERNANCE RESEARCH

In this chapter, I will present existing literature on AI governance to understand some of the backdrop against which the European Commission is acting by conducting their AI policy. The review of literature will consist of research from the perspective of AI governance ranging from ethics to governance as well as a short section on the influence of public perception of AI on governance.

## 2.1. Ethics versus Governance

With the rise of AI on account of increased accessibility to big data, governance related steps have been taken to confront this development. The often-adopted approach to ensure that AI remains safe has centered on ethics where principles and values to guide the use of AI have been formed (e.g., Floridi & Cowls, 2019; Whittlestone et al., 2019). Efforts of "soft law" have been put forward by actors ranging from public to private: companies, non-profit organizations, research institutions, and public sector organizations (Jobin et al., 2019), including the Ethics Guidelines for Trustworthy AI by the European Commission (AI HLEG, 2019b). This trend of forming ethical guidelines to guide AI has been noticed in academia as a substantial amount of the existing research in AI policy has seemed to focus on AI ethics (e.g., Floridi et al., 2018; Floridi & Cowls, 2019). Less attention has been given to AI governance as the frame to overcome concerns and controversies related to AI as opposed to ethics (Ulnicane et al., 2020). But this is changing.

A study surveying peer reviewed research articles on the international debate on AI regulation between the years 2009 and 2019 found that most, up to 94%, of these articles were published after the year 2015 (Almeida, Santos, & Farias, 2020). The study reveals that ethics and risks have remained continuous topics in these articles throughout the years, but the topic of AI regulation became prominent in 2016 (Almeida, Santos, & Farias, 2020). These findings don't only demonstrate that AI has tickled interest in the recent years but also that AI governance, more specifically regulation of AI, has raised attention only in the past few years.

### 2.1.1. Ethics Alone Won't Save AI

An emerging view in literature seems to be that ethical considerations in the form of guidelines and principles alone are not sufficient to ensure fair development and deployment of AI as concerns over the misuse or lack of sufficiency and effectiveness over ethics have been consistently brought up. Mittelstadt (2019) states that "without a fundamental shift in regulation, translating principles into practice will remain a competitive, not cooperative, process" (p. 505), suggesting that ethical principles can be easily exploited at the cost of ethical or trustworthy AI without proper regulatory measures. This kind of exploitation could be visible in how actors with personal interests in AI development or innovation may try to promote certain principles that suit them the best at the cost of others, hiding these selfish intentions behind AI ethics. Floridi (2021b) has called this *digital ethics shopping* and provided the definition:

> the malpractice of choosing, adapting, or revising ("mixing and matching") ethical principles, guidelines, codes, frameworks, or other similar standards (especially but not only in the ethics of AI), from a variety of available offers, in order to retrofit some pre-existing behaviours (choices, processes, strategies, etc.), and hence justify them a posteriori, instead of implementing or improving new behaviours by benchmarking them against public, ethical standards. (Floridi, 2021b, p. 82-83)

Another term, quite close by, used to display ethical actions with little ethical commitment is *ethics bluewashing* that is the "malpractice of making unsubstantiated or misleading claims about, or implementing superficial measures in favour of, the ethical values and benefits of digital processes, products, services, or other solutions in order to appear more digitally ethical than one is" (Floridi, 2021b, p. 83). Parallel observations of ethics washing are made by McMillan and Brown (2019).

Resseguier and Rodrigues (2020) also question the effectiveness of AI ethics of today. Similarly, they fear of "its potential for misuse as a replacement for regulation" or "using ethics to prevent the implementation of legal regulation that is actually necessary" (Resseguier & Rodrigues, 2020, p. 2). *Digital ethics lobbying*, as Floridi (2021b) has called it, refers to "the malpractice of exploiting digital ethics to delay, revise, replace, or avoid good and necessary legislation (or its enforcement) about the design, development, and deployment of digital processes, products, services, or other solutions" (p. 85). Resseguier and Rodrigues (2020) do acknowledge AI ethics to have a place in addressing "cognitive and perceptive

inertia that hinders our capacity to see what is different from before or in different contexts, cultures or situations and what, as a result, calls for a change in behaviour (regulation included)" (p. 3), but view that AI ethics are not being used for this. For ethics to be effective, they call for close investigation of the ethics in use to make sure of their purpose so that it is in line with the held norms and values (Resseguier & Rodrigues, 2020). They propose that "without a continuous process of questioning what is or may be obvious, of digging behind what seems to be settled, of keeping alive this interrogation, ethics is rendered ineffective (Resseguier & Rodrigues, 2020, p. 3).

Hagendorff (2020) brings about the notion of ethics as a powerless mechanism to implement its "own normative claims" (p. 99). He also argues that due to this weakness to impose real influence ethics may appear as appealing to those who prefer self-governance over legislation (Hagendorff, 2020). One of Hagendorff's critiques of the available ethical guidelines of today are the vague, abstract technical explanations. For example, AI is used as "a collective term for a wide range of technologies or an abstract large-scale phenomenon" (Hagendorff, 2020, p. 111) rather than carefully defined accordingly to the context that it is used in. This results in ethics operating "at a maximum distance from the practices it actually seeks to govern" (Hagendorff, 2020, p. 112) discouraging compliance amongst AI developers and the whole industry.

### 2.1.2. Ethical Frameworks

Floridi and Cowls (2019) have coined the term "principle proliferation" (p. 2) to address the issue of the abundance of emerging ethical principles "for the adoption of socially beneficial AI" (p. 2). According to Floridi (2021), this trend of forming ethical principles can result in confusion on what are good principles or unnecessary pressure to form guidelines for the sake of appearance. Floridi and Cowls (2019) critically view the high volume of principles as incordinate and perplexing, but still believe in the potential of ethical guidelines "for future efforts to create laws, rules, technical standards, and best practices for ethical AI in a wide range of contexts" (p. 2). Thus, they have come up with an overarching framework with five central principles to answer to the problem of proliferation. Their proposed core principles for ethical AI are beneficence, non-maleficence, autonomy, justice, and explicability (Floridi & Cowls, 2019). In another study investigating the global landscape of AI ethics by reviewing different ethical guidelines around AI, the following principles stood up: transparency, justice and fairness, non-maleficence, responsibility, and privacy (Jobin et al., 2019).

Whereas critique towards ethical guidelines alone forming the base for the governance of AI have surfaced, AI ethics seem to still hold a place in the overall governance of AI. As can be seen from the events taking place in the European Union, calls for additional measures have been put to action. Ethics are not sufficient or effective enough on their own to watch over the field of AI, but they do form the base for implementing stricter policy.

### 2.1.3. Towards AI Governance Frameworks

This growing critical stance towards the sufficiency of AI ethics has not gone unnoticed in academia. For example, Ulnicane et al. (2020) have taken an interest in AI governance and positioned it at the center of their research. In their paper, they introduce governance as an alternative frame to ethics in addressing the controversies related to AI presented in different policy documents. Their goal is to shift focus from ethics and give attention to the study of AI governance that is more underdeveloped perspective.

Although AI governance is still somewhat of an underdeveloped field of research (e.g. Taeihagh, 2021), some theoretical frameworks for the regulation of AI have already been proposed. Gasser and Almeida (2017) have introduced a model for AI governance that consists of three interacting layers that sit between society and AI systems. The foundational layer is the technical layer followed by an ethical layer to address ethical concerns. On the top, there is the social and legal layer to assign appropriate regulation. Rahwan (2018) has also proposed a conceptual framework for approaching the regulation of AI and algorithmic systems that takes on the human-in-the-loop (HITL) thinking but broadens the idea of an individual human to the society. By doing this and introducing a social contract framework, Rahwan (2018) not only considers the watch over algorithmic performance as a technical matter, but also a social. This he calls the society-in-the-loop approach (Rahwan, 2018).

Wirtz et al. (2020) have critiqued both of these aforementioned models for the lack of concrete information on how to implement regulation in practice. According to Wirtz et al. (2020), Rahwan has disregarded to consider the government responsibilities in his proposed framework. Gasser and Almeida also fail to provide details on who is to assign and oversee regulation, nor do they provide a theoretical foundation for their model (Wirtz et al., 2020).

Against this backdrop, Wirtz et al. (2020) have proposed their own layered AI governance framework building on insights from AI challenges and governance literature along with regulation theory. Taeihagh (2021) has stated that "to enhance the benefits of AI while

minimising the adverse risks they pose, governments worldwide need to understand better the scope and depth of the risks posed" (p. 138). The framework for AI governance presented by Wirtz et al. (2020) acknowledges and uses the risks, threats, or challenges introduced by AI as the base for the needed policy actions, also factoring the possible benefits of AI for choosing the right form of policy action. Their model differs from Rahwan's and Gasser and Almeida's in providing "a detailed explanation for a regulatory process, in which measures of regulation are developed, evaluated and enacted" (Wirtz et al., 2020, p. 826).

The model begins with AI applications/services and technology layer to distinguish between the different kinds of functions of the AI system in question. These functions can vary from data acquisition to data processing to data embedment. Each of the function comes with own specific challenges. Based on these functions, the second layer defines the challenges of that particular AI system in use. This layer is divided into three categories: AI society, AI ethics, and AI law and regulation. In the course of this layer, the distinctive challenges are identified for further proceedings. Next layer is the AI regulation process layer. Regulation process begins with framing the problem where "stakeholders interested in the regulation of certain challenges come together to formulate a common understanding of the problem, defining the objective of the regulatory action they want to enact" (Wirtz et al., 2020, p. 823). Following the framing step, comes the assessment of risks, benefits, and costs. Thereafter, the risks, benefits, and costs must be evaluated accordingly to provide the necessary information for the last step, which is the risk management. In the risk management stage, the decisions regarding regulation are made and finally regulation is implemented. Implementation of the regulation is followed by monitoring of its success and possible side effects. The regulatory actions, formed at the policy layer, may be directed at more technical challenges, for example in the form of "industry standards or guidelines to promote the disclosure of AI and data use" (Wirtz et al., 2020, p. 825), or they may touch upon ethical concerns by setting certain moral principles for AI technology. The most time-consuming form of regulation is the implementation of new norms and laws with an aim to solve social or legal challenges. The final layer, collaborative governance layer, represents those involved in the regulatory process. These actors can range from representatives from governmental institutions to private organizations and NGOs, and can form agencies, committees, or foundations.

### 2.1.4.  Public Perception of AI

Another interesting area of research around AI is the public perception of AI. There is a handful of research done on how AI is perceived (e.g. Araujo et al., 2020; Cui & Wu, 2021; Fast & Horvitz, 2017; Neri & Cozman, 2020). These public perceptions about AI should not be disregarded as insignificant. As Mannes (2020) points out, "how people feel about AI will shape how it is adopted and deployed" (p. 63). On similar lines, it's been said that "understanding public concerns about AI is important, as these concerns can translate into regulatory activity with potentially serious repercussions" (Stone et al., 2016, cited in Fast & Horvitz, 2017, p. 963). The reception of AI from the perspective of the public should thus be considered when governing AI.

### 2.1.5.  Summarizing

I have attempted to present some of the existing literature from the perspective of AI governance. The aim of this chapter was to establish the current state of AI governance research and shed light on its progression from ethical guidelines to governance frameworks. Although the discipline here seems to still be quite young, the relation between research in AI ethics and AI governance is important to distinguish as it shows the development of the governance for AI. Ethics have dominated the field of research to some extent, but the trend is gradually changing. The introduced conceptual AI governance models can provide valuable insight into the planning of policy structures. Another option to resort to is to look towards previous governance frameworks in technology, for example the internet (Almeida & Gasseida, 2017). There may be a lot more to adopt from the emergences of previous technologies governance, policy, and ethics wise (Ulnicane et al., 2021). While assessing AI governance structures, one should also consider the way that AI is perceived in the society.

This academic background presented in the chapter is valuable to keep in mind as this thesis focuses on assessing policy that leads to a regulatory proposal on AI. I believe that the European approach to AI will also offer fertile grounds for future research on the intersection of AI ethics and AI governance as EU is on the frontline attempting to take concrete steps to regulate AI in a human-centric way. The Commission's regulatory proposal on AI could be seen as a real-life example of an AI governance model. More theoretical governance frameworks may arise following the release of the Commission's regulatory proposal.

# 3. THE WPR APPROACH

In the previous chapter, I introduced different policy models through AI ethics and AI governance. These models suggest ways forward for the governance of AI. The emphasis of this thesis is in the progression of the European AI policy and the way that AI has been framed in policy. An article published in March 2022 shares a similar interest in the analysis of the policy evolution as it looks at the political drivers and the policy process of AI regulation and governance in the EU (Justo-Hanani, 2022). Contrary to this study, though, my thesis takes a different theoretical approach on the topic. To find answers to my research question concerning the representations of AI in the European Commission's policy, I will turn to Carol Bacchi's (1999) conceptualization of policy problems. Bacchi (2009) has introduced a poststructuralist, both theoretical and methodological approach called the "What's the Problem Represented to be?", or the WPR approach, to help explain the underlying presuppositions behind policies or policy proposals. This approach is applied to the context of my thesis as it offers a critical tool to assess policy proposals.

The WPR approach is situated in the tradition of critical policy studies in social sciences. Central to critical policy studies are naturally policy processes. Critical policy studies are interested in the knowledge used in policy making: "both the knowledge used to shape policy and the kinds of knowledge and assumptions that guide the implementation of policy decisions" (Fischer et al., 2015, p. 1). Critical policy studies reject the positivist thinking of knowledge and seek to understand how the larger context of knowledge affects policy processes (Fischer et al., 2015). There is a strong sense of furthering democratic processes associated with critical policy studies as the orientation moves away from technocratic models of democracy (Fischer et al., 2015).

The tradition emerged during the 1960's "as an academic response to the social and political turmoil" (Fischer et al., 2015, p. 2). The political tensions of the time period in the Western societies, involving of events such as "the civil rights struggle, the War on Poverty, the Vietnam War, the threat of nuclear holocaust, student unrest, the emergence of the women's movement, health and safety problems, recognition of the environmental crisis, and more", were not adequately met in social science research in the opinion of many (Fischer et al., 2015, p. 2). Against this backdrop, more critical analysis on public policy began to surface. The theoretical approaches that have emerged, and keep emerging, within critical policy studies can be distinguished from another (Fischer et al., 2015). Fischer et al. (2015) have

named three approaches that have especially stood out within the field: interpretive, critical, and poststructuralist. Bacchi's framework can be seen as a continuation of the tradition of critical policy studies, representing the poststructuralist strand. The WPR framework, first introduced in 1999 (Bacchi, 1999), draws inspiration especially from the philosopher Michel Foucault on concepts such as "problematization, governmentality, subjectification and contestation" (Bacchi, 2010a). Bacchi's interest has focused on critically assessing issues related to women's rights and gender equality, although her policy analysis tool can be applied to many kinds of policy, including the context of EU and AI technologies. Similar to my topic, the WPR approach has previously been applied to the analysis of the problematizations in EU's AI ethics (Koulu, 2020) and problem representations of risk in the GDPR (Padden & Öjehag-Pettersson, 2021).

The core of the WPR approach is to question and interrogate the way we are governed (Bacchi, 2009). Its "goal is to discover deep-seated ontological and epistemological commitments within policies ('solutions') that are probably hidden to policy makers and planners" (Bacchi, 2010b, p. 63). Bacchi's (2009) approach defies the premise that policy is in the first place implemented with an intention to address a problem. Whereas conventional policy seeks to fix or solve an issue or a problem from the outside, the WPR approach views that the policy itself is constituting to the production of the problem that the policy supposedly is solving (Bacchi, 2009). Thus, problems are "endogenous" and "created within" the policy (Bacchi, 2009, p. x) while polices "give shape to problems" (Bacchi, 2009, p. 1). A problem in Bacchi's understanding "refers simply to the kind of change implied in a particular policy proposal" (Bacchi, 2009, p. xi). The objectives of Bacchi's approach are the interpretations or representations of problems in policy proposals.

Bacchi (1999) states that "how we perceive or think about something will affect what we think ought to be done about it" (p. 1). Those involved in policy processes have great power in defining what may be considered a problem and what may not be considered a problem by simply imposing policies as policies are implicitly involved in defining problems (Bacchi, 2009). Either policy makers name problems associated with the policy in question, or they imply these problems unobtrusively through policy proposals. For example, by regulating AI in a certain manner, a specific problem is insinuated by this action. Bacchi acknowledges the relation between the policy maker's perception of a situation and how that is reflected in the actions taken to address it. But what about those problems that are left outside of the perceptions of policy makers? The WPR approach suggests that the problems that policy

seeks to solve are only one representation of the situation, and the approach is interested in uncovering alternative problem representations that are left unexamined (Bacchi, 2012). Additionally, Bacchi (1999) encourages reflection on those issues that remain "unaddressed or undiscussed because of the ways certain 'problems' are represented" on a more general level (p. 2).

To critically assess problem representations in policies or policy proposals, Bacchi (2009) has presented a set of six questions to break down the process of problem production in policy making:

1. What's the "problem" (e.g. of "problem gamblers", "drug use/abuse", domestic violence, global warning, health inequalities, terrorism, etc.) represented to be in a specific policy?
2. What presuppositions or assumptions underlie this representation of the "problem"?
3. How has this representation of the "problem" come about?
4. What is left unproblematic in this problem representation? Where are the silences? Can the "problem" be thought about differently?
5. What effects are produced by this representation of the "problem"?
6. How/where has this representation of the "problem" been produced, disseminated, and defended? How could it be questioned, disrupted, and replaced? (Bacchi, 2009, p. 2)

The intention of my thesis is to analyze the policy problems related to AI in the European Commission's documents that have contributed to the European approach to AI. My view is that Bacchi's approach suits this intention well. The WPR approach offers a fresh and critical angle to analyze the formation of AI policy questioning how the policy has come about, what it is that it aims to do, and what problem representations are insinuated by it. As AI is often approached through the challenges and opportunities affiliated with it, it seems fitting to use a critical policy analysis that places problematizations at the center. Another viable road that could have been taken to guide the theoretical foundation of this thesis is discourse theory.

My research questions are strongly inspired by Bacchi's question framework. My first research question "How has the European Commission's policy on AI come about?" is directly drawn from Bacchi's third question on the formation of the representation of the problem in question. Bacchi's (2009) third question asks "How has this representation of the "problem" come about?" (p. 2). The question offers insight on "the specific developments

and decisions that contribute to the formation of identified problem representations" (Bacchi, 2009, p. 10). The research question looks into, in this case, the background of the European AI policy examining its origins and evolution and "highlighting the conditions that allow a particular problem representation to take shape and to assume dominance" (Bacchi, 2009, p. 11).

My second research question "How has AI been represented as a policy problem by the European Commission?" is inspired by Bacchi's questions one and two. The first aim of my second research question is to distinguish "implied problem representations in specific policies or policy proposals" (Bacchi, 2009, p. 4). This is supported by Bacchi's (2009) first question, "What's the "problem" represented to be in a specific policy?" (p. 2). After identifying the problem representations related to AI, a closer examination is navigated towards the *conceptual logics* behind the identified representations with the help of Bacchi's (2009) second question, "What presuppositions or assumptions underlie this representation of the "problem"?" (p. 2). According to Bacchi (2009), "conceptual logic refers to the meanings that must be in place for a particular problem representation to cohere or to make sense" (p. 5). One such meaning, as proposed by Bacchi (2009), is a *binary* or a *dichotomy* which the problem representation may be founded on. In binary settings, problems are presented with two competing sides. According to Bacchi (2009), "what is on one side of a binary is considered to be excluded from the other side" (p. 7). Additionally, one side is often "considered to be more important or more valued than the other side" (Bacchi, 2009, p. 7). The binary thinking is apparent in Commission's AI policy in, for example, the way that it balances between market increase and ethical considerations. In addition to binaries, *key concepts* and *categories* may be behind the presuppositions or assumptions on problem representations (Bacchi, 2009).

The remaining three questions from Bacchi's six questions will be left outside of this thesis. Question four's aim is to inspect possible shortcomings of the problem representations by asking "What is left unproblematic in this problem representation? Where are the silences? Can the "problem" be thought about differently?" (Bacchi, 2009, p. 2). This question is designed to discover the silences, the problems that are unrepresented in policy, if there are any. Although this question would have been an interesting one to ask, it will be left out due to space constrains and the angle of this thesis. Question five, "What effects are produced by this representation of the 'problem'?" (Bacchi, 2009, p. 2), is asking for the impacts of the policy problem which are not available for evaluation based on the selected database.

Question six asks "How/where has this representation of the 'problem' been produced, disseminated, and defended? How could it be questioned, disrupted, and replaced?" (Bacchi, 2009, p. 2), which is also beyond the scope of the research data. The data selected consists only of documents produced by the European Commission so analysis beyond the Commission's views is not possible.

# 4. METHODS AND DATA

In this chapter, I will present the research methods and the data selected in this thesis to study the European approach to AI and the policy problems found in the European approach to AI. I will also explicate how my research is executed in practice. The research method to conduct this thesis is a combination of qualitative content analysis and Bacchi's WPR approach. The first phase of analysis has been done through the method of qualitative content analysis. In this phase, I have drawn the main repetitive themes of AI challenges and opportunities from the data and organized them into their respective categories. With the second phase of analysis, the results of the first phase were further examined and developed with the question framework from Bacchi's (2009) WPR approach. These results are presented in the sixth chapter and provide an answer to my second research question. Along with analyzing the representation of policy problems, the WPR approach was also used to examine the evolution of the AI policy. In the fifth chapter I have systematically presented how the European approach to AI has come about and answered my first research question.

## 4.1. Methods

Building on the ideas of a German psychologist Mayring, Drisko and Maschi (2015) offer a definition to qualitative content analysis as "a set of techniques for the systematic analysis of texts of many kinds addressing not only manifest content but also the themes and core ideas found in texts as primary content" (p.85). In the first phase of the analysis, the selected data consisting of Commission's policy documents has first been reviewed using this technique of qualitative content analysis to identify and categorize AI challenges and opportunities presented in the documents. The scope of analysis in this thesis is any implication of a challenge or an opportunity related to AI. Challenge is used as a generic word for all negative implications of AI such as risks or threats. Opportunity, on the other hand, refers to the possibilities or benefits linked to AI by the Commission. I have not been interested in individual, specific words but rather meanings insinuated within the text. These meanings are found within sentences, phrases, or even whole paragraphs. After having identified the challenges and opportunities from the data and analyzed the problem representations, I have categorized them (Table 4 at the end of chapter six). Sub-categorization is used to show the variety of views that are put forward.

The reasons why qualitative content analysis suits this thesis well are in the flexibility for interpretation of context and latent meanings as well as the formal aspects of the content (Drisko & Maschi, 2015), not just the manifest content. Based on this understanding, qualitative content analysis gives freedom to analyze the data comprehensively. Another viable option for an analysis method producing similar results to qualitative content analysis would have been thematic analysis. Regardless of this equally apt option, I decided to proceed with qualitative content analysis because Drisko and Maschi (2015) suggest that "what researchers vaguely label as 'thematic analysis' may be most similar to contemporary qualitative content analysis" (p. 83).

One limitation involved in using qualitative content analysis to spot challenges and opportunities related to AI from the data is that the processing of the text is manually done. There is a risk that the analysis of the texts will be subjective due to bias. There is also the chance that some challenges or opportunities have been overlooked. To ensure that this risk is reduced to the minimum, I have attempted to be as transparent as possible.

With the first phase of analysis which consisted of categorizing the AI challenges and opportunities, I have been able to showcase the main policy problems evident in the data. I have also broken each category down in order to distinguish the variety of aspects under the main problem. Identifying these policy problems has provided an answer to Bacchi's (2009) first question "What's the 'problem' represented to be in a specific policy?" (p. 2) and conveniently also started the process of answering my second research question. After identifying and characterizing policy problems in the data, I have proceeded to provide answers to Bacchi's (2009) second question used in this thesis, "What presuppositions or assumptions underlie this representation of the "problem"?" (Bacchi, 2009, p. 2). Bacchi's second question has been applied to critically analyze the problem representations with the intention to interpret the formation of the policy problems further. By applying Bacchi's second question to my analysis, I have aspired to reveal the still latent meanings behind the problem representations identified earlier. Bacchi's two first questions make up my second research question which asks, "How has AI been represented as a policy problem by the European Commission?" I answer to my second research question in sixth chapter of this thesis.

My first research question, "How has the European Commission's policy on AI come about?", has been drawn from Bacchi's (2009) third question, "How has this representation

of the 'problem' come about?" (p. 2) and is answered in chapter five as the emergence of the European approach to AI is explained in detail. The chapter contextualizes the setting for the analysis of the AI problem representations explained previously.

As mentioned earlier, questions four, five and six from Bacchi's (2009) question framework have been discarded in this thesis due to space constrains. The data is also unfit to answer to all of Bacchi's questions. Especially questions five and six simply cannot be answered due to the limitations of the research data.

## 4.2. Data

Over the course of a four-year timeframe, ever since the EU member states and Norway signed the Declaration of Cooperation on AI (European Commission, 2018a), the European Commission has released multiple documents related to their strategy on AI. For my research data, I have carefully selected the significant policy documents that have been involved in the development of the European approach to AI between years 2018 and 2021. The documents have been chosen to correspond to the progression of Commission's policy making on AI. These documents outline the actions planned for EU in regards AI over the four-year time period. Four of these five documents constitute of Commission's communication papers which neatly summarize main messages from the Commission regarding AI at the time of their release. A communication from the Commission is a policy paper that outlines the possible problems in a specific policy field or evaluates certain policy in effect.[1] It often sums up the main policy updates in a compact manner. In addition to the communications from the Commission, I have included a white paper ordered by the Commission. A white paper, on the other hand, is a report that often presents a way forward for the issue at hand.[2] All the documents are published by the European Commission. The documents are all accessible electronically. The chosen documents overlap periods of the Juncker and the von der Leyen administrations.

In the practice of the WPR research method, the selection of data is considered to be "fairly open-ended" (Bacchi, 2009, p. 20) implying that it is up to the author of the study to make the imperative decisions in choosing the suitable texts. In addition to legislation and government

---

[1] Communication. (n.d.). EU Monitor. Retrieved April 15, 2022 from
https://www.eumonitor.eu/9353000/1/j9vvik7m1c3gyxp/vh7dptp45uyn
[2] Policy measures of the European Union. (January 8, 2022). Wikipedia. Retrieved April 15, 2022 from
https://en.wikipedia.org/wiki/Policy_measures_of_the_European_Union#White_Paper

reports, the data could consist of a range of publications on the topic like, for example, from the media (Bacchi, 2009). Bacchi (2009) rightly reminds that "choosing policies to examine is itself an interpretive exercise" (p. 20) sending the study to a certain direction. In this thesis, I have opted to select policy papers published by the Commission of mainly the same caliber. Like stated earlier, this includes communication papers and one white paper. Both document types fit the research approach I have chosen for this thesis. Therefore, I have included all four communication papers published by the Commission over the period of 2018–2021 on AI policy in the data. Furthermore, the one white paper that was published before the release of the actual regulatory proposal is an excellent addition to the repertoire of the communication papers. I believe one strength of my data to be in the consistency of the data and another in the systematic run-through the selected data offers over the development of Commission's AI policy.

Table 1: *Selection of Research Data*

| Policy document | Release date |
|---|---|
| Communication: Artificial Intelligence for Europe | April 25, 2018 |
| Communication: Coordinated Plan on Artificial Intelligence | December 7, 2018 |
| Communication: Building Trust in Human-Centric Artificial Intelligence | April 8, 2019 |
| White Paper: On Artificial Intelligence - A European approach to excellence and trust | February 19, 2020 |
| Communication: Fostering a European approach to Artificial Intelligence | April 21, 2021 |

The first document is a communication from the Commission, Artificial Intelligence for Europe (European Commission, 2018b), that suggests a way forward for EU on handling AI. This document is the starting point for the data and the first policy document to surface on AI after the Declaration of Cooperation on AI (European Commission, 2018a) between member states. It is the European strategy to take up AI. The AI strategy is followed by a communication from the Commission titled Coordinated Plan on Artificial Intelligence (European Commission, 2018c) that outlines the results of meetings between member states,

Norway, Switzerland and the Commission during June and November 2018 to identify actions to take to build on the joint AI strategy. The communication is a supporting document of the actual Coordinated Plan (European Commission, 2018d) and "highlights the main objectives and initiatives of the plan" (European Commission, 2018c, p. 2). Following the Coordinated Plan on AI is the communication on Building Trust in Human-Centric Artificial Intelligence (European Commission, 2019) which presents the requirements for trustworthy AI following the ethics guidelines (AI HLEG, 2019b). Next document in the data is the White Paper on Artificial Intelligence - A European approach to excellence and trust (European Commission, 2020a) that suggests policy options for AI in the EU in line with the European strategy for data (European Commission, 2020c) and the Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics (European Commission, 2020b). Last but not least, the selected data sample includes the Communication: Fostering a European approach to artificial intelligence (European Commission, 2021b) that is released as the supplementary document to the regulatory proposal (European Commission, 2021a), the updated Coordinated Plan (European Commission, 2021c) and the Impact of Assessment document (European Commission, 2021d).

This thesis is interested in the buildup and the formation of policy that culminates as the regulatory proposal. The regulatory proposal itself is intentionally left outside of the selected data. This decision is justified for several reasons. First, the communication paper released at the time of the proposal is included in the data and successfully recapitulates the main points raised from the proposal. Thus, in this context, there is no need for closer examination of the actual proposal. Secondly, the statements that have surfaced already in the prior years to the proposal are seen as the primary contributors to the composition of the legislation and thus preferred. The proposal itself is not at the center of this research data, but rather the factors that have contributed to the formation of this specific proposal. The inclusion of the regulatory proposal would shift the focus of the thesis strongly to the proposal, which is not the main topic of this thesis. The challenges and opportunities of AI which are the center of this thesis are likely to have been acknowledged in preceding policy documents already. Thirdly, the technicality of the regulatory proposal advocates for it to be left out of the scope of the analysis of this thesis. My skills, as the author of this thesis, may not be sufficient to interpret such a technical document on the development and deployment of a complex, scientific matter such as AI. All this being said, the regulatory proposal is a major factor to

keep in mind in the analysis of the AI policy and is not completely left out as the data does include the communication document published at time of the release of the proposal. This ensures that the necessary information from the regulatory proposal will then be acknowledged in the appropriate matter.

The same reasoning used for the exclusion of the regulatory proposal can be applied to the exclusion of other documents such as the Coordinated Plan on AI (European Commission, 2018d) or the updated version of the same document three years later, the Coordinated Plan on Artificial Intelligence 2021 Review (European Commission, 2021c). The Coordinated Plan of 2018 is included in the communication paper, Coordinated Plan on Artificial Intelligence (European Commission, 2018c), and the second version of the plan is in the communication paper, Fostering a European approach to Artificial Intelligence (European Commission, 2021b). Other similar documents with contribution to the policy but not necessarily enough importance to be included in the data individually are documents published by the Commission's appointed High-Level Expert Group on AI (AI HLEG, 2019a; AI HLEG, 2019b; AI HLEG, 2019c; AI HLEG, 2020a; AI HLEG, 2020b), the Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics (European Commission, 2020b), and the Impact Assessment of the Regulation on AI (European Commission, 2021d). For the scope of this thesis, the communication papers that touch on the content proposed in these documents serve satisfactorily.

# 5. PROGRESSION OF EUROPEAN AI POLICY

In this chapter, I will provide an answer to the first research question "How has the European Commission's policy on AI come about?" I aim to explain the premises that my analysis begins from and go over the development of the European approach to AI. I will begin by covering the works that have been bubbling in the European Commission in regards AI, systemically present the documents that have contributed to the policy process, and then move onto the larger picture of EU digital policy and technological sovereignty. This chapter will shed light on the progression of the AI policy towards the Artificial Intelligence Act (AIA). To contextualize the events leading up to the release of the AI regulatory proposal, I will use documents published by the European Commission as well as statements by the Commission's president, Ursula von der Leyen. Some of these documents are also included in my primary research data. At the end of this chapter, I will provide the EU definitions for AI.

## 5.1. The European Approach to AI

During the Juncker administration, while discussing EU legislative priorities for 2018–2019 in December 2017, the Commission together with the Council and the Parliament agreed on working on "ensuring a high level of data protection, digital rights and ethical standards while capturing the benefits and avoiding the risks of developments in artificial intelligence and robotics" (European Commission, 2017).

Following this, the work towards a legal framework for AI officially began in April 2018 when EU member states and Norway signed the Declaration of Cooperation on AI (European Commission, 2018a) with an intention to upgrade Europe's competitiveness in the field of AI and form a unified front to address challenges brought by technological advancements in AI. Soon after the signing of the declaration, Romania, Greece, Cyprus, and Croatia also joined the initiative.[3] With the signing of this declaration, policy on AI began forming.

Since the election and formation of the von der Leyen administration in 2019[4], AI has continued to be treated as an important matter from two perspectives. The president-elect

---

[3] EU Member States sign up to cooperate on Artificial Intelligence. (April 19, 2018). Shaping Europe's Digital Future, European Commission. Retrieved January 27, 2022 from https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence
[4] The Commissioners. (n.d). European Commission. Retrieved February 9, 2022 from https://ec.europa.eu/commission/commissioners/2019-2024_en

Ursula von der Leyen's speech on November 27, 2019, suggested that "we [EU] will automate work that is wearisome for us humans: carrying heavy loads, performing repetitive tasks in factories or in offices" (von der Leyen, 2019a, p. 8). This statement contains a strong implication that speaks out to the opportunistic side of AI. On the other hand, the speech also brought up the need for regulation to ensure that AI will not bypass the human:

> With the General Data Protection Regulation we set the pattern for the world. We have to do the same with artificial intelligence. Because in Europe we start with the human being. It is not about damming up the flow of data. It is about making rules that define how to handle data responsibly. For us the protection of a person's digital identity is the overriding priority. (von der Leyen, 2019a, p. 9)

In her political guidelines for the European Commission of 2019–2024, von der Leyen envisions a "Europe fit for the digital age" (von der Leyen, 2019b, p. 13-14). She wants "Europe to strive for more by grasping the opportunities from the digital age within safe and ethical boundaries" (von der Leyen, 2019b, p. 13). As is apparent from von der Leyen's statements on AI presented here, the political discourse on AI seems to balance between the great opportunities envisioned and the need for rules to ensure safety and fairness. This two-parted approach to AI is proven to steer the policy as will be explicated in the later part of the analysis in chapter six.

The regulatory proposal, Artificial Intelligence Act (AIA), is the product of a line of preparatory work done by the Commission between the years 2018–2021. The AIA, published on April 21, 2021, serves as a first ever proposal for a legal framework for AI. The regulatory proposal offers Commission's latest view on AI and suggests a risk-based approach to regulating AI (European Commission, 2021a). The risk-based approach separates four levels of risk in AI systems: unacceptable risk, high risk, limited risk, and minimal or no risk. Based on the level of the risk involved with the AI system, the framework determines the strictness of the requirements for the use of the technology. In the case of the AI system having unacceptable risk, the technology is automatically banned. For high-risk AI, there are strict requirements that need to be met. The regulatory proposal has stipulated a list of high-risk AI technology. With limited risk in an AI system, the requirements are lessened. Minimal or no risk AI systems do not face any obligations. It is estimated that the regulation could commence in a transitional period in the second half of the year 2022 and become applicable to operators earliest in the year 2024 (European Commission, 2021a).

The AIA is not the first time that European Union is showing leadership in taking up technological matters. In 2016, EU introduced the General Data Protection Regulation (GDPR) to ensure data protection and privacy of EU citizens (Regulation 2016/679). In von der Leyen's (2019a) words earlier, it is stated that EU should set the pattern for AI as was done with the GDPR.

The AIA has received interest by the research community along with other actors, such as non-profit organizations like Algorithm Watch[5]. The professor of Philosophy and Ethics of Information at the University of Oxford, Luciano Floridi[6], comments in his analysis of the proposed legislation that "on the whole, it [AIA] is a good starting point to ensure that the development of AI in the EU is ethically sound, legally acceptable, socially equitable, and environmentally sustainable, with a vision of AI that seeks to support the economy, society, and the environment" (Floridi, 2021a, p. 216). On the contrary, the proposal has sparked critique, especially regarding its risk-based approach to regulation. One worry seems to be that fundamental human rights are disregarded with an approach that favors innovation and growth through its regulatory loopholes (Hidvegi, Leufer, & Massé, 2021).

Along with the regulatory proposal in April 2021, the Commission published other supporting documents: a communication on Fostering a European Approach to Artificial Intelligence (European Commission, 2021b), an updated Coordinated Plan on AI (European Commission, 2021c), and an Impact Assessment of the Regulation on AI (European Commission, 2021d). Out of these documents released with the proposal, the communication paper is included in the data selected for this thesis as it ties together the main messages from the regulatory proposal and the supporting documents.

The regulatory proposal and the supporting documents published in spring 2021 are the culmination of the policy that has been under the works in the past few years ever since the Declaration of Cooperation on AI (European Commission, 2018a). Although some recognition of the regulatory proposal is in place, this thesis is more interested in the overall development of the policy on AI over the past years. After the decision to cooperate on finding the means to govern AI in the EU, the Commission has released a handful of documents that have contributed to the formation of the European approach to AI. All of

---

[5] Draft AI Act: EU needs to live up to its own ambitions in terms of governance and enforcement (August 2021). Algorithm Watch. Retrieved October 5, 2021 from https://algorithmwatch.org/en/eu-ai-act-consultation-submission-2021/

[6] University of Oxford (n.d.). Professor Luciano Floridi. Retrieved January 24, 2022 from https://www.oii.ox.ac.uk/people/profiles/luciano-floridi/

these documents together demonstrate how the Commission's regulatory proposal for AI has come about. In Table 2, I have demonstrated the timeline of the emergence of the documents involved in shaping the European approach to AI.

Table 2: *Development of the Commission's AI policy*

| Time of release | Policy document |
|---|---|
| December 2017 | • Joint Declaration on the EU's legislative priorities for 2018–19 |
| April 2018 | • Declaration of Cooperation on Artificial Intelligence<br>• Communication: Artificial Intelligence for Europe |
| June 2018 | • Set up of the HLEG on AI and European AI Alliance |
| December 2018 | • Coordinated Plan on Artificial Intelligence<br>• Communication: Coordinated Plan on Artificial Intelligence |
| April 2019 | • Communication: Building Trust in Human-Centric Artificial Intelligence<br>• HLEG: Ethics Guidelines for Trustworthy Artificial Intelligence<br>• HLEG: A Definition of AI |
| June 2019 | • HLEG: Policy and Investment Recommendations for Trustworthy AI |
| February 2020 | • White Paper: On Artificial Intelligence – A European approach to excellence and trust<br>• Report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics<br>• Communication: A European strategy for data |
| July 2020 | • Public consultation on the AI White Paper<br>• HLEG: Assessment List for Trustworthy AI (ALTAI)<br>• HLEG: Sectoral Considerations on Policy and Investment Recommendations for Trustworthy AI |
| April 2021 | • Regulatory proposal: AIA |

| | <ul><li>Communication: Fostering a European approach to Artificial Intelligence</li><li>Coordinated Plan on Artificial Intelligence 2021 Review</li><li>Impact Assessment of the Regulation on Artificial Intelligence</li></ul> |
| --- | --- |

Source: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

Only 15 days after the Declaration of Cooperation in April 2018 (European Commission, 2018a), the communication paper on Artificial Intelligence for Europe kicked everything off as it offered the base for a unified strategy towards AI (European Commission, 2018b). In the aftermath of the strategic alignment, an independent High-Level Expert Group on AI (AI HLEG) was set up to advise on policy development.[7] A multi-stakeholder forum, AI Alliance, was also opened to offer feedback to the AI HLEG.[8] Towards the end of the year 2018, a Coordinated Plan on AI was released to propose joint actions for member states on AI (European Commission, 2018c; European Commission, 2018d). In April 2019, the AI HLEG (2019b) put forward their Ethics Guidelines for Trustworthy Artificial Intelligence along with a communication from the Commission on Building Trust in Human-Centric Artificial Intelligence (European Commission, 2019). At the same time, the AI HLEG (2019a) presented its conception of the definition of AI given by the Commission. What followed quite soon was AI HLEG's (2019c) Policy and Investment Recommendations for Trustworthy AI. In February 2020, the Commission (2020a) published a white paper on AI for suggestions on policy options for the regulation of AI. A report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics (European Commission, 2020b) coupled with the Communication on a European strategy for data (European Commission, 2020c) were also released at that time. Other supporting documents put together by the AI HLEG were the Assessment List for Trustworthy AI (ALTAI) (2020a) and the Sectoral Considerations on Policy and Investment Recommendations for Trustworthy AI (2020b). A public consultation on the white paper consisting of citizen and stakeholder views on AI was also compiled (European Commission, 2020h).

---

[7] High-level expert group on artificial intelligence. (September 27, 2021). Shaping Europe's digital future, European Commission. Retrieved September 29, 2021 from https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

[8] The European AI Alliance. (September 27, 2021). Shaping Europe's digital future, European Commission. Retrieved February 9, 2022 from https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance

The scale of documents presented here is versatile and shows the range of actors involved in policy making. The contribution of the HLEG's consultation to the formation of policy is indisputable, but it is difficult and not in the interest of this thesis to estimate the depth of its influence. The same applies to the public consultation of the white paper and its impact on policy. One could also ponder on the effects of related policy documents such as the report on the safety and liability implications (European Commission, 2020b) as well as the European strategy for data (European Commission, 2020c) Overall, it is not the aim of the WPR approach nor this thesis to evaluate who has contributed to the policy and how effectively, but rather how the policy comes to view the processed issues. Under these circumstances, it is vital to understand the emergence of the policy and to acknowledge the contributing documents. Ultimately, the documents below (Figure 1) provide the efficient means for analysis of the actual policy problems in the practice of the WPR approach as they sum up the progression of the policy over the course of the four-year time frame.

Figure 1: *Primary data*
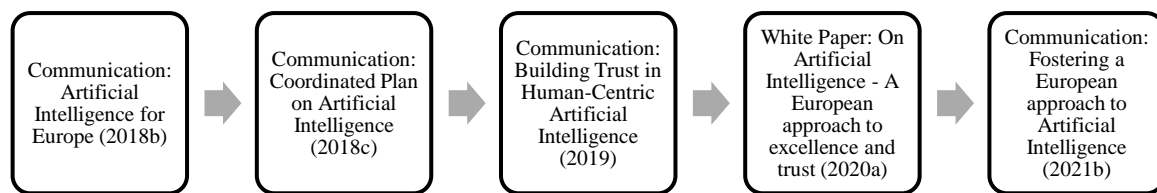


## 5.2. European Technological Sovereignty

The Commission's approach to AI and their proposal for a legal framework on AI has by no means evolved in a vacuum. There are multiple on-going initiatives and projects with emphasis on advancement of digital services. For example, the Commission's Digital Compass that sets the way for the Digital Decade.[9] The Digital Compass and the Digital Decade refer to Commission's vision, targets, and avenues for a successful digital transformation of Europe by 2030. This initiative builds on the Commission's digital strategy Shaping the Europe's Digital Future announced in February of 2020 with updated views on

---

[9] Europe's Digital Decade: digital targets for 2030. (n.d.). European Commission. Retrieved February 14, 2022 from https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en

how COVID-19 pandemic has influenced the understanding of digitalization.[10] In addition, the Digital Europe Program (DIGITAL) is an EU funding program that focuses on digital transition of businesses, citizens, and public administrations.[11] It provides strategic funding in five areas with one of them being AI. Other initiatives related to the field of AI are the EU Cybersecurity Strategy (European Commission, 2020g), the Digital Services Act (2020e), the Digital Markets Act (2020f), and the Data Governance Act (2020d). AI is just one portion of this digital whole that is under way in the EU. This backdrop behind the more specific AI policy is important to bear in mind when one searches for an answer to how the AI policy has formed.

The European quest for controlled AI seems to be continuum of a wider trajectory related to digital policy in the EU that is aiming for technological sovereignty. Technological sovereignty, or digital sovereignty, which are here used as synonyms, is a topic of interest that has repeatedly appeared in von der Leyen's messages (e.g. von der Leyen, 2019b; von der Leyen, 2019c). In her political guidelines, von der Leyen (2019b) states that "it may be too late to replicate hyperscalers, but it is not too late to achieve technological sovereignty in some critical technology areas" (p. 13) initiating that technological sovereignty is on the table for the EU. In like manner, von der Leyen (2019c) asks Thierry Breton, the Commissioner for Internal Market, in his mission letter to enhance Europe's technological sovereignty, including AI. The demand for European technological sovereignty may only have been strengthened in the fight against the COVID-19 pandemic (e.g., Bauer & Erixon, 2020; Darnis, 2020). A communication paper on AI (European Commission, 2021b) reaffirms this by stating that "AI has demonstrated its potential by contributing to the fight against COVID-19, helping to predict the geographical spread of the disease, diagnose the infection through computed tomography scans and develop the first vaccines and drugs against the virus" (p.1).

Floridi (2019) describes *digital sovereignty* simply as "control of the digital" including "the control of data, software (e.g. AI), standards and protocols (e.g. 5G, domain names), processes (e.g. cloud computing), hardware (e.g. mobile phones), services (e.g. social media, e-commerce), and infrastructures (e.g. cables, satellites, smart cities)" (p. 370-371). Floridi

---

[10] Europe's Digital Decade: Commission sets the course towards a digitally empowered Europe by 2030. (March 9, 2021). European Commission. Press release. Retrieved September 14, 2021 from https://ec.europa.eu/commission/presscorner/detail/en/IP_21_983

[11] The Digital Europe Programme. (September 22, 2021). Shaping Europe's Digital Future, European Commission. Retrieved September 28, 2021 from https://digital-strategy.ec.europa.eu/en/activities/digital-programme

(2019) states that digital sovereignty is "an epochal struggle" giving it high importance. He reminds that "the most visible clash is between companies and states" (Floridi, 2019, p. 371). By this, he refers to the power dynamic between companies and states (Floridi, 2019). Development, deployment, and maintenance of the digital is usually in the hands of the companies whereas states have the power to steer development through the exercise of different regulatory actions (Floridi, 2019). In a nutshell, companies decide "the nature and speed of change" and the states govern "the direction of change" (Floridi, 2019, p. 371). With its AI policy, the European Commission may aspire to gain some power over the field and manage how it will be shaped. This goal over the field of AI could also be seen as supporting the vision of European technological sovereignty.

Bauer and Erixon (2020) have classified four factors that have contributed to the acquirement of technological sovereignty in Europe. These include culture, control, competitiveness, and cybersecurity, which they summarize as follows:

> 1. Culture: the "cultural" approach to technology sovereignty starts from the assumption that Europe is different from other parts in the world in our defense of values and market regulations – manifested for instance in data protection rights. At the heart of this view is the perception that, in particular, digital regulation presents a fundamental choice between individual rights and business freedom, and that Europe has made its choice to protect human values and rights over business freedom.

> 2. Control: there is a strand in the debate that takes a command-and-control view of technological sovereignty, arguing that the EU or individual Member States need to have the policy instruments to control the outcomes of the digital economy in general and how citizens and companies use modern digital services.

> 3. Competitiveness: another viewpoint collects different thoughts and considerations around industrial competitiveness – the future capacity of European multinational enterprises to compete on world markets and fears about declining European influence vis-à-vis other standard-setting powers.

> 4. Cybersecurity: finally, there is a growing demand for new policies to protect personal and business data, and to have at disposal all the tools and technologies necessary to protect digital integrity and digital resilience. (Bauer & Erixon, 2020, p.8)

The cultural factor is advocated in the Commission's AI policy by placing high value on human-centricity through regulation all the while the Commission is setting the EU apart from others by doing this. For control, it is in the Commission's interest to stipulate an environment where the development and deployment of AI happens on European terms through policy and regulation. It is also in the Commission's interest to ensure its competitiveness against, for example, USA and China in the field of AI technology. In acquiring regulatory practices in AI, the Commission has also underlined the importance of, for instance, data privacy and cybersecurity.

## 5.3. The EU Definitions of AI

Because the focus of this thesis is on AI in the European context, it is valuable to know how AI is defined by the Commission. Next, I will provide the definitions for AI used by the European Commission at different stages.

Before the regulatory proposal was released, the appointed independent High-Level Expert Group on AI (AI HLEG, 2019a) had put out a document that addressed the definition of AI. Their aim was to expand on the previously suggested definition by the European Commission's communication on AI:

> Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.

> AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications). (European Commission, 2018b)

Although the expert group does develop the definition they do provide a disclaimer stating that their "description and definition of AI capabilities and research areas is a very crude oversimplification of the state of the art" and that "the intent of this document is not to precisely and comprehensively define all AI techniques and capabilities, but to describe summarily the joint understanding of this discipline that the High-Level Expert Group is using in its deliverables" (AI HLEG, 2019a).

The expert group's proposal for a more defined definition of AI is:

> Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.

> As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems). (AI HLEG, 2019a)

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

In the Annex I, are listed these techniques and approaches:

(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;

(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;

(c) Statistical approaches, Bayesian estimation, search and optimization methods. (European Commission, 2021a)

Due to the fast-changing nature of AI, the regulatory proposal's definition of an AI system "aims to be as technology neutral and future proof as possible, taking into account the fast technological and market developments related to AI" (European Commission, 2021a). This technology-neutral and future-proof definition of AI systems attempts to also "cover

techniques and approaches which are not yet known or developed" (European Commission, 2021b, p. 6). With a future-proof definition for AI, the regulatory framework requires continuous reevaluation.

The variety of definitions and the development of the definitions over such short period of time shows that AI is a difficult subject to deal with. The technical and changing nature of the discussed subject complicates defining it which is an important step in approaching the subject via regulation.

# 6. REPRESENTATIONS OF AI

In the previous chapter on the emergence of the European approach to AI, I responded to my first research question on how the policy on AI has formed over the years by introducing documents involved and the environment they have been released in. I also briefly discussed the larger phenomenon of technological sovereignty that AI policy taking place in the EU is a fraction of. In this chapter, I will continue with my analysis of the policy and the representation of AI while answering to my second research question "How has AI been represented as a policy problem by the European Commission?" This chapter will focus on the five selected data documents as defined in my methods and data chapter (Table 1) and again in the previous chapter (Figure 1).

The European Commission has separated two main objectives in its policy making. On one hand, it highlights how it seeks to dynamically to enhance the competitiveness of AI within Europe whereas, on the other, it wants to improve the trustworthiness of AI for European citizens. Both objectives are believed to be enhanced through AI regulation. These objectives are repeated in the data over the course of four years of policy making. A third, less emphasized objective is preparing the society for changes to come (European Commission, 2018b). The socioeconomic objective seems to become less important because as the policy develops the "three-pronged approach to boost the EU's technological and industrial capacity and AI uptake across the economy, prepare for socio-economic changes, and ensure an appropriate ethical and legal framework" (European Commission, 2019, p. 1) reforms into "a regulatory and investment oriented approach with the twin objective of promoting the uptake of AI and of addressing the risks associated with certain uses of this new technology" (European Commission, 2020a, p. 1).

The Commission's view is that the two main objectives work hand in hand and complement one another as "Europe's approach to Artificial Intelligence shows how economic competitiveness and societal trust must start from the same fundamental values and mutually reinforce each other" (European Commission, 2019, p. 10). It is seen that competitiveness embarks from consumer trust and safety of the technology: "the further development and promotion of . . . safety standards and support in EU and international standardisation organisations will help enable European businesses to benefit from a competitive advantage, and increase consumer trust" (European Commission, 2018b). The twin objectives have been preserved in the latest communication paper on the proposed regulation to AI: "The

Commission's proposal therefore pursues the twin objectives of addressing the risks associated with specific AI applications in a proportionate manner and of promoting the uptake of AI" (European Commission, 2021b, p. 6). The overall goal of the regulatory proposal on AI is to combine "greater safety and fundamental rights protection while supporting innovation, enabling trust without preventing innovation" (European Commission, 2021b, p. 7).

Even though, the Commission has strongly communicated these two perspectives in its policy, I have attempted to also bring attention to other perspectives present in the research data to comprehensively analyze the representation of AI by the European Commission in this thesis. The perspectives I have formed have been identified through different challenges and opportunities that the Commission has expressed of AI in their policy documents. The challenges and opportunities have helped sketch the way for how AI is represented in the data. By categorizing these AI challenges and opportunities found in the data, I have identified the matters that the policy seeks to address or solve and answered to Bacchi's first question. Bacchi's second question has helped me to dive deeper into each category and identify the implicit policy problems suggested under each perspective.

## 6.1. AI Problem Representations

While, the Commission has acclaimed competitiveness and trust as its policy objectives, many other aspects are also displayed in the documents. The challenges and opportunities related to AI found in the data present a range of issues which I have organized into categories and sub-categories to ease the processing of them. Inspiration for using categorization has been drawn from earlier research oriented towards uncovering and distinguishing the AI challenges and opportunities involved. A particular multidisciplinary study outlines the main crosscutting themes in AI research including challenges as one theme amongst decision-making, application, and big data (Dwivedi et al., 2019). These challenges have been divided into categories of social; economic; data; organizational and managerial; technological and technology implementation; political, legal, and policy; as well as ethical (Dwivedi et al., 2019). In a similar manner to Dwidevi et al. (2019) organizing the multidisciplinary collection of challenges found in AI research into categories, Kaplan and Haenlein (2020) have identified six key dilemmas of AI: politics, economics, society, technology, environment, and law. These categorizations help understand the vast volume of aspects to consider with AI. In this fashion, I have formed eight categories with their

respective sub-categories to showcase the width of matters Commission considers with AI (Table 4 at the end of this chapter). I have come up with the categories through careful assessment of the research data by letting the data speak for itself. The categories used in this thesis consist of the following: safety and security, ethical, legal, competitiveness, AI leadership, socioeconomic, ecological, and education.

It is worthy to mention that many of the challenges or opportunities overlap with one another. The technical nature of AI (e.g. AI's ability to learn) is quite frankly a challenge to all categories as the way AI behaves may evolve and change how it is to be approached. On top of that, for example, the effects of AI in the employment landscape are highly relevant to both socioeconomic aspects and the EU's competitiveness. Subsequently, the changing work field may affect the safety and security or ethical risks related to AI when the work force starts to lack the necessary skills needed to understand or upkeep AI whereas increasing education on AI can ease this effect. Next, I will dive into each category in turn, present the challenges and opportunities involved, and analyze the problem representations suggested by the policy from that perspective.

### 6.1.1. Safety and Security

Without safe and secure AI technology, there really is no use for AI. One of the main threats to safety and security for the European citizens acknowledged by the Commission is intrusion of privacy. Ensuring digital privacy and protecting personal data during the use of AI-enabled products and services is considered important as "AI increases the possibilities to track and analyse the daily habits of people" (European Commission, 2020a, p. 11). Remote biometric identification, which refers to "facial recognition tools to check passers-by in public spaces" (European Commission, 2021b, p. 7), is also acknowledged as an intrusive surveillance risk. In the regulatory proposal, the use of remote biometric identification is prohibited under certain circumstances (European Commission, 2021a).

> The real-time use for law enforcement purposes would in principle be prohibited in publicly accessible spaces, unless when exceptionally authorised by law. Any authorisation is subject to specific safeguards. In addition, all AI systems intended to be used for remote biometric identification of natural persons must undergo an ex ante conformity assessment procedure by a notified body to check compliance with the requirements for high-risk AI systems, and will be subject to stricter logging and human oversight requirements. (European Commission, 2021b, p. 7)

Another aspect brought to attention in the research data is the use of AI "in breach of EU data protection and other rules . . . by employers to observe how their employees behave" (European Commission, 2020a, p. 11). The Commission acknowledges a risk that "AI may also be used to retrace and de-anonymise data about persons, creating new personal data protection risks even in respect to datasets that per se do not include personal data" (European Commission, 2020a, p. 11).

A common worry amongst citizens from the perspective of safety and security is that "AI can have unintended effects or even be used for malicious purpose" (European Commission, 2020a, p. 9). Material safety and security threats can surface from application of AI in weapons systems or technical flaws by AI. There is the possibility that AI can be used for weaponization, although this aspect has only been raised once in the communication paper on the Coordinated Plan on AI in relation to international security (2018b). It is simply stated that "the Union will continue to stress that international law, including International Humanitarian Law and Human Rights Law, applies fully to all weapons systems, including autonomous weapons systems, and that States remain responsible and accountable for their development and use in armed conflict" (European Commission, 2018c, p. 8). The research data suggests that technical hick-ups in the products and services provided by AI may impose a more prominent threat to autonomous weapons systems. It's worth noting that "these risks can be caused by flaws in the design of the AI technology, be related to problems with the availability and quality of data or to other problems stemming from machine learning" (European Commission, 2020a, p. 12). This threat may materialize in a situation such as an autonomous car wrongly identifying an object on the road and causing an accident with serious consequences (European Commission, 2020a, p. 12).

In the communication on Fostering a European approach to Artificial Intelligence it is also stated that AI systems that "distort a person's behaviour through subliminal techniques or by exploiting specific vulnerabilities in ways that cause or are likely to cause physical or psychological harm" (2021b, p. 7) are to be prohibited.

According to the Commission, AI technology does not only pose safety and security challenges but also opportunities. The benefits of AI are envisioned to enhance the safety of transportation (e.g. European Commission, 2018b) or help "law enforcement agencies to fight crime more efficiently" (European Commission, 2019, p. 1). AI may also help in "anticipating cybersecurity threats" (European Commission, 2018b) or "anticipating natural

disasters" (European Commission, 2018c, p. 1). AI technology's possibilities are also acknowledged in helping build resilience to future shocks as "European companies will dispose of sufficient expertise to rapidly apply AI to new challenges" (European Commission, 2021b, p. 3).

Regarding safety and security of AI, the Commission seems to raise up the concern of AI affecting the privacy of its citizens or causing material hazards in the society (European Commission, 2020a; European Commission, 2021b). By raising these concerns, the Commission confirms that AI is an unpredictable technology that may pose threats that the EU is not equipped to handle. For example, using autonomous cars means exposing people to technological errors in traffic that may have serious repercussions. Nevertheless, on a different note, the Commission highlights how AI may enhance the safety of transportation (European Commission, 2018b) or enable law enforcement tools to more effective crime prevention (European Commission, 2019). With these remarks, the Commission suggests that the current transportation system or law enforcement could be improved with the help of AI. The message from the Commission is inconsistent and a little confusing: on one hand, it is advocating for the benefits of AI, whereas on the other, it is critical towards its application. The same issue is thus presented with two competing outcomes. This can be seen as an example of a binary positioning of a problem representation (Bacchi, 2009). To lower the risk of safety and security threats and embrace the benefits of AI technology, the Commission has come to propose its regulatory actions. The regulatory proposal is a mechanism in policy designed to address the safety and security risks raised with the development and deployment of AI technology.

### 6.1.2. Ethical

Ethics is one of the issues that the Commission addresses in its AI policy. As a disruptive technology, the Commission acknowledges that AI introduces new situations and scenarios that demand attention and caution that have not been needed before. The Commission recognizes a specific need for ethics along with regulatory actions: "AI applications should not only be consistent with the law, but also adhere to ethical principles and ensure that their implementations avoid unintended harm" (European Commission, 2019, p. 2). In this way, ethics play a crucial role in the governance of AI.

In the face of a new emerging technology, the Commission acknowledges that AI can disrespect its EU fundamental rights unless dealt with.

The use of AI can affect the values on which the EU is founded and lead to breaches of fundamental rights, including the rights to freedom of expression, freedom of assembly, human dignity, nondiscrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation, as applicable in certain domains, protection of personal data and private life, or the right to an effective judicial remedy and a fair trial, as well as consumer protection. (European Commission, 2020a, p. 11).

Breach of these rights can occur through use of biased data leading to discrimination (e.g. European Commission, 2020a) or lack of accountability and transparency (e.g. European Commission, 2019) as AI has the functionality to make decisions autonomously without human help. The Commission acknowledges that "the specific characteristics of many AI technologies, including opacity ('black box-effect'), complexity, unpredictability and partially autonomous behaviour, may make it hard to verify compliance with, and may hamper the effective enforcement of, rules of existing EU law meant to protect fundamental rights" (European Commission, 2020a). It is implied that situations where decisions made by algorithms are not traceable should be avoided. Another interesting notion to consider is that depending on the context where AI is applied, the magnitude of ethical risks may vary (European Commission, 2019). The characteristics of AI such as opaque decision-making, complexity, unpredictability, autonomy, and the machine's ability to learn on its own along with the context of application make AI a force that needs to be dealt with in an appropriate manner to ensure that it is safe and to be trusted. The Commission has adopted the ethics guidelines drafted by the AI HLEG (European Commission, 2019). According to the guidelines, the trustworthiness of an AI system should consider seven requirements: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; societal and environmental well-being; and accountability (European Commission, 2019).

All this being said, it is also the Commission's view that the strong EU fundamental values and rights will provide the needed ground for which AI policy will be built on and ethical guidelines applied to as "it is vital that European AI is grounded in our values and fundamental rights such as human dignity and privacy protection" (European Commission, 2020a, p. 2).

Ensuring ethical development and deployment of AI is at the center of the AI policy. Although the policy making aims for a regulatory framework, it does not rule out the importance of ethics guidelines. Leaving ethics out of the equation on developing and deploying AI would be quite backwards. Instead, the Commission states that "there is a need for ethics guidelines that build on the existing regulatory framework and that should be applied by developers, suppliers and users of AI in the internal market, establishing an ethical level playing field across all Member States" (European Commission, 2019, p. 2). Ethics could thus be seen as the building blocks for the regulatory framework, ensuring that AI is developed "in a way that puts people at its centre and is thus worthy of the public's trust" (European Commission, 2019, p. 2). For this reason, the Commission's ethical guidelines serve a purpose in the policy and are not used for ethics bluewashing (Floridi, 2021b). The inclusion of the ethical perspective also corresponds with the AI governance models established earlier in the chapter on relevant research (Gasser & Almeida, 2017; Rahwan, 2018; Wirtz et al., 2020). Ethics guidelines could also enhance public acceptance of a new technology and impact the way AI is welcomed to a society (Mannes, 2020).

This analysis is strictly speculative of how the Commission seeks to benefit from ethics guidelines as part of their AI policy. It should be noted that this thesis does not evaluate the practices of the ethical guidelines, but only the policy documents that rationalize the use of the ethical guidelines in question. The effectiveness of the Commission's ethics guidelines in the development and deployment of AI in support of the EU fundamental values is an entirely different question.

### 6.1.3. Legal

The Commission seems to strongly strive to be the one in the driver's seat over the European AI industry from a legal perspective. As a solution to many issues that the EU faces with AI, the Commission has laid down its plan to strengthen the existing laws on AI by proposing regulation specifically for AI. Although, regulation is intended to build trust in AI, it does not come without its difficulties.

The characteristics of AI mentioned in the previous section, such as opacity, complexity, unpredictability, and autonomy, complicate legal matters on AI. The Commission remarks that AI is a difficult subject to control as "certain specific features of AI technologies (e.g. opacity) can make the application and enforcement of such legislation more challenging and generate high risks for which a tailored regulatory response is needed" (European

Commission, 2021b, p. 6). A one size fits all solution is thus unlikely for AI from a legal perspective. The risk-based approach in the regulatory proposal could to some extent be justified by this remark. There are also limitations in the scope of existing EU legislations given that "the use of AI in products and services can give rise to risks that EU legislation currently does not explicitly address" (European Commission, 2020a, p. 14). The Commission (2020a) points out that "these risks may be present at the time of placing products on the market or arise as a result of software updates or self-learning when the product is being used" (p. 14). The governing of AI technology is thus very difficult due to the unpredictable nature of the technology.

A significant current legal deficit is the absence of an EU-wide legislation over AI technologies that "may . . . reduce overall levels of safety and undermine the competitiveness of European companies" (European Commission, 2020a, p. 12). Closely related to this is the divergence of national legal initiatives between member states on AI (e.g. European Commission, 2020a) resulting in legal uncertainty and lack of common standardization. For this reason and from the perspective of the Commission, the enforcement of the regulation and the application of the Commission's AI policy is anticipated.

Even though AI technology creates legal challenges that demand action, the Commission is not completely unprepared as "the EU has a strong and balanced regulatory framework to build on, which can set the global standard for a sustainable approach to this technology" (European Commission, 2018b). With this, the Commission is referring to, for example, its product liability framework and the General Data Protection Regulation that was applied in 2018 (Regulation 2016/679). Both acts are applicable to AI as well. Additionally, the Free Flow of Non-Personal Data Regulation (Regulation 2018/1807) has been seen as a great enabler for unlocking data and removing barriers for non-personal data (e.g. European Commission, 2018c). With this foundation, the regulatory framework has not been completely created from scratch and should easily fall in place into the digital politics of the EU.

The legal perspective is closely linked to the safety and security and the ethical perspectives. A legislative system is naturally one of the most effective ways to govern. With a legal foundation, the Commission strives to ensure AI to be safe, secure, and ethical. It seems that the Commission is concerned with the lack of an EU-wide legislation and member states' independent initiatives to take on AI, worrying about the dispersion of approaches towards AI

inside the Union. If a unified approach to AI is missing, then the EU has a weaker position in legally addressing ethical or safety risks that may emerge and endanger consumers or citizens using AI applications. Furthermore, the encompassing regulatory framework would preempt ethical or safety risks from arising in the first place. That said, another concern here appears to some extent stem from the fear of hurting EU's single market and competitiveness through a shattered legal system. This theme will be unfolded in the upcoming section on competitiveness.

### 6.1.4. Competitiveness

EU's competitiveness in the field of AI is one of the Commission's policy's cornerstones along with ensuring safe and ethical AI through regulation. Competitiveness appears to compose of varied aspects including investment, single market, legislation, attractiveness, uptake of AI, and data access.

The fact that USA and China are ahead of the EU in the field of AI (e.g. Castro & McLaughlin, 2021; Probst et al., 2018) has seemed to influence how the Commission strategizes its investment goals. In the research data, the Commission benchmarks USA and China (European Commission, 2018b). It becomes clear that Europe is currently lacking in investments which may be factoring in its incapability to challenge USA or China in the market. In 2016, China had privately invested in 6.5-9.7 billion euros in AI and USA 12.1-18.6 billion euros whereas Europe only 2.4-3.2 billion euros (European Commission, 2018b). Since then, the Commission has set the goal of reaching at least EUR 20 billion in investment by 2020 and from there on EUR 20 billion per year over the course of the next decade (e.g. European Commission, 2018b). In the footsteps of its competition, the Commission places pressure on stepping up its private investments in AI (European Commission, 2018b), demonstrating the Commission's urge to reach to its competition's level and beyond. The Commission stresses that "national efforts need to be increased" (European Commission, 2018c, p. 3) and that it is "essential to make sure that the private sector is fully involved in setting the research and innovation agenda and provides the necessary level of co-investment" (European Commission, 2020a, p. 7). The private sector along with the EU member states play a crucial role in reaching the goal of EUR 20 billion each year. The European Union itself will invest EUR 1 billion per year from Horizon Europe and the Digital Europe programs (European Commission, 2021b, p. 1-2).

One of the Commission's worries regarding AI is the fragmented single market. Like already established in the legal section, this fragmentation may be caused by the lack of "a common European approach to AI" and differing national AI initiatives taken by member states (European Commission, 2020a, p. 2). The Commission (2020a) states that "if the EU fails to provide an EU-wide approach, there is a real risk of fragmentation in the internal market, which would undermine the objectives of trust, legal certainty and market uptake" (p. 10). A functioning internal single market is thus strongly seen as relating to the competitiveness of the European AI industry. A disjointed approach on AI, or any other subject for that matter, within the Union can be seen as jeopardizing the internal single market that aims for a "harmonized regulatory environment", enables easy-flowing trade, and strengthens the European marketplace as a whole (Bradford, 2020, p. 7). The dispersed market environment that the members states' individual strategies on AI causes is not ideal for the EU. Instead, the Commission seeks unity between its members. In fact, the EU has gone as far as to adopt its member states' stringent regulatory standards to its policy when they have risked injuring the harmonization of the single market (Bradford, 2020). In the white paper, the Commission voices its support for upward regulatory convergence when it comes to regulating AI (European Commission, 2020a) verifying this argument.

Common standards laid down by EU regulation enhance the functioning of the single market. In the long run, forming a unified front on AI within the EU is implicitly represented as delivering a leverage for EU against its competitors whoever they may be. It seems that the Commission has timed its interference in European AI smartly because in 2018, when the Commission began pushing its policy, only a handful of EU member states had a national strategy for AI (European Commission, 2018b; European Commission 2018c). This has allowed a good internal position for the Commission to coordinate its AI policy and act as a pioneer on the regulatory front globally.

Labor migration in the tech industry is directly impacting the competitiveness of the EU. Brain-drain is recognized as a problem in Europe as "talented researchers and promising start-ups frequently receive interesting offers from abroad" (European Commission, 2018c, p. 5). More specialists are needed in the information and communications field (European Commission, 2018c). As skilled workers leave to go work elsewhere, loses the EU its valuable labor force. Brain-drain may send the message to the outside that EU is not an attractive place from the technological perspective. The overall attractiveness of EU including its image and reputation as a potential employee are thus on the line. The blue card,

which offers non-EU citizens the opportunity to come work in Europe, is also one example of EU attempting to revive the field (European Commission, 2018c).

An impacting factor in brain-drain might be the state of research and development in Europe. The Commission (2018c) points out that "Europe must be able to train, attract and retain talent of this kind, and encourage entrepreneurship, diversity and gender balance" (p. 5). Thus, effort is directed towards the research and innovation network. Calls for "a lighthouse centre of research, innovation and expertise" (European Commission, 2020a, p. 6) are made as "Europe cannot afford to maintain the current fragmented landscape of centres of competence with none reaching the scale necessary to compete with the leading institutes globally" (European Commission, 2020a, p. 6).

Another way to improve the image of EU as a technological hub appears to be encouraging public administrations and small and medium sized enterprises (SMEs) in Europe to take up AI in their operations. The Commission emphasizes that there is room for improvement in the "uptake of AI across the EU economy and public administration" (European Commission, 2020a, p. 5). Areas of public interest should "rapidly begin to deploy products and services that rely on AI in their activities" (European Commission, 2020a, p. 8) and SMEs should be supported in taking up AI as they might lack the necessary skills (European Commission, 2020a; European Commission, 2018c). The Commission seems to think that by fully immersing Europe in AI, it will become more attractive. The Commission emphasizes that the "European industry cannot miss the train" when it comes to the uptake of AI (European Commission, 2018b), suggesting that this will depict a progressive Europe.

Although Europe's weak position in consumer applications and online platforms is blocking some of its access to large amounts of data, the Commission is putting its hopes on the ever-growing volume of data in the world and the public and industrial data of which EU holds large volumes of (European Commission, 2020a). The Commission is also encouraging companies to open up their privately held data for re-use while continuing to do the same with public sector data (European Commission, 2018b). As AI feeds off data, access to large amounts of it is an important factor linked to the level of competitiveness. The Commission point-blankly states that "Europe's current and future sustainable economic growth and societal wellbeing increasingly draws on value created by data" (European Commission, 2020a, p. 1). It could be that Europe's weak access to consumer data (European Commission, 2020a) prevents it from reaching a certain level in the field of AI. The European data strategy

has addressed this matter comprehensively (European Commission, 2020c). While the data strategy is not included in my research data, it implicitly influences policy on AI and vice versa.

Opportunities for AI related to competitiveness are acknowledged in multiple sectors, for example, agriculture, healthcare, transport, robotics, energy, and financial services (European Commission, 2020a), as well as in "improving the efficiency of production systems through predictive maintenance" (European Commission, 2020a, p. 1). For post COVID-19 economic growth, expectations are put on AI to help recovery through the Recovery and Resilience Facility (RRF) investments (European Commission, 2021b). For seizing opportunities of AI, the Commission is also seeing potential in the blue card in "allowing high-skilled non-EU citizens to work and live in the EU" (European Commission, 2018c, p. 5).

### 6.1.5. AI Leadership

The Commission recognizes an opportunity to pursue leadership in paving a regulatory way and setting global standards for AI. The Commission (2020a) expresses how "Europe is well positioned to exercise global leadership in building alliances around shared values and promoting the ethical use of AI" (p. 8). Its mission seems to have become to spread the message of human-centric AI to the rest of the world as it is "convinced that international cooperation on AI matters must be based on an approach that promotes the respect of fundamental rights, including human dignity, pluralism, inclusion, nondiscrimination and protection of privacy and personal data and it will strive to export its values across the world" (European Commission, 2020a, p. 9). The Commission sees that "The EU can make a unique contribution to the worldwide debate on AI based on its values and fundamental rights" (European Commission, 2018b). It also highlights its position for showing example in "developing and using AI for good and promoting a human-centric approach and ethics-by-design principles" (European Commission, 2018c, p. 8).

The Commission's motivation for EU to become a global leader in the field of AI is strong. The Commission notes that at the moment "AI regulation is in infancy" (European Commission, 2021b, p. 4) and the Commission plans for the EU to take advantage of this. It sums that:

> Europe can combine its technological and industrial strengths with a high-quality
> digital infrastructure and a regulatory framework based on its fundamental values to

become a global leader in innovation in the data economy and its applications as set out in the European data strategy. (European Commission, 2020a, p. 2)

The Commission has recognized the absence of regulation in AI industry and fittingly to its values started to pursue a leading role here. I say fitting because this agenda entails similarities to, for example, the agenda of the GDPR. By emphasizing human-centricity and regulation in AI, Commission seems to seek a fresh competitive edge against its competitors, while promoting European fundamental values and all that it stands for. The quest for the leading regulator position over the AI industry is beneficial for Europe's image and offers the EU a significant footing over the industry. AI regulation may just be the EU's opening to competition with, for example, USA and China. As Floridi (2021a) has been arguing: "the challenge is no longer digital innovation but the governance of the digital" (p. 220). The actions of the Commission are well positioned with this argument. And what became clear from the literature review over AI governance, the absence of regulation is concrete. The Commission seems to have hit this window of opportunity. Having missed this window, the future of EU in matters related to AI could be very different. Now it has firmed itself a familiar position of a responsible regulator that differs from emphasizing market freedom like the USA or placing the state above all like China, gaining itself credibility as a legitimate operator in AI.

In the final communication paper, the Commission states that "the stakes are high for the EU to spearhead the development of new ambitious global norms, AI-related international standardisation initiatives and cooperation frameworks, in line with the rules-based multilateral system and the values it upholds" (European Commission, 2021b, p. 4). Although it is not yet possible to evaluate the impacts of the EU taking a global leader position, I would suspect a domino effect following. A phenomenon called *the Brussels Effect* is "the EU's unilateral ability to regulate the global marketplace" (Bradford, 2020, p. 1). By assigning regulatory practices to AI technology inside EU borders, the Commission will not just incentivize member states but third countries and external market actors to comply to its rules if they plan on continuing their businesses within and with the EU. This means that European standards would not just concern European operators but anyone who seeks to operate in the EU providing the EU with power of setting norms for the AI industry.

Bradford (2020) separates the EU's internal motive of pursuing a single market from the external motive for setting norms globally. In the research data, the Commission seems to

also distinguish these as separate entities as it makes a point of ensuring "consistency between the EU's external actions and its internal policies" (European Commission, 2021b, p. 4). According to Bradford (2020), the norm setting agenda has transpired as kind of a side product of the initial pursuit of the internal single market that has been designed to enhance the EU's competitiveness. The norm setting agenda of the Commission is visible in its policy on AI by the inclusion of a separate section dedicated for international aspects in the white paper (European Commission, 2020a). The HLEG also heard non-EU actors in the process of forming the ethical principles (European Commission, 2020a).

Along with the benefits of the single market for the competitiveness of the EU in AI industry, the ability to set a global norm is valuable because with it the EU can prove its influence as a legitimate, global economic power (Bradford, 2020). The ability to set the practices for a global technological industry would provide an advantageous environment for the EU to operate and possibly raise EU's rank in the global AI race. Failing to do this, although it is not directly expressed in policy, it could be assumed that the Commission fears of the EU being left behind. The strong global position in AI would also strengthen the EU's endeavor for technological sovereignty.

### 6.1.6. Socioeconomic

Socioeconomic changes are inevitable as AI grows. For instance, AI is expected to have big impact on the labor market by causing transformation and disappearance of certain jobs (e.g. European Commission, 2018b). The Commission (2018c) expresses that "technological changes will modify the skills required of workers, meaning that potentially very large numbers of workers will need to upskill" (p. 5). For example, the impact of AI on lower skilled jobs may cause inequality "if not addressed early and proactively" (European Commission, 2018b). On the other side of the coin, though, are the new job profiles that AI brings (e.g. European Commission, 2018b), but the advantage of this may not be fulfilled unless the work force is trained sufficiently to take on AI. It is suggested that:

> [P]olicy-makers will develop strategies to deal with employment changes in order to ensure inclusiveness, as the pace with which some jobs will disappear and others appear is likely to accelerate, while business models and the way tasks or jobs are performed will change. This may make it necessary to modify current labour market and social protection arrangements to support transitions in the labour market. (European Commission, 2018c, p. 6)

Another socioeconomic challenge that may affect the citizens' attitude to the changing labor market is distrust in AI technology. It is important to ensure "that workers are given the chance to adapt and to have access to new opportunities" so that they accept AI (European Commission, 2018b).

Overall, the acceptance and vast application of AI requires trust in the technology and "to further strengthen trust, people also need to understand how the technology works, hence the importance of research into the explainability of AI systems" (European Commission, 2018b). Public awareness on AI could be increased by benchmarking "the technical capabilities of AI components and systems to give a realistic understanding of where the technology stands" (European Commission, 2018b). Information of "the AI system's capabilities and limitations, in particular the purpose for which the systems are intended, the conditions under which they can be expected to function as intended and the expected level of accuracy in achieving the specified purpose" is specifically important for the transparency of AI applications considered high-risk (European Commission, 2020a, p. 20). Trust is also related to the predictability, responsibility, verifiability, fundamental rights, and ethical rules of the technology (European Commission, 2018c). To enhance trust, it should always be clearly communicated when one is interacting with an AI system instead of a human (European Commission, 2020a). It is evident that the ethical aspect is very much present in the construction of trustworthiness of AI.

On top of the new job profiles that AI can bring to the table, it is expected to ease workload through automation by freeing workers from "repetitive, strenuous and even dangerous tasks (for example cleaning unsafe or difficult to access locations such as industrial pipes)" (European Commission, 2018b). It is also envisioned that AI will "help summarise large amounts of data, provide more accurate information and suggest decisions, including using AI to assist doctors with diagnosis" (European Commission, 2018b). For the aging population and persons with disabilities "AI can provide new solutions to support more people to participate and remain in the labour market" (European Commission, 2018b).

Other socioeconomic benefits that AI may bring to the table are in healthcare as AI can help "making diagnosis more precise" and "enabling better prevention of diseases" (European Commission, 2020a, p. 1). In the data documents, the Commission (e.g. European Commission, 2018c) repeatedly brings up a project in which a database consisting of health images is being developed to help the diagnosis of cancer. AI has also potential "in informing

citizens and enhancing citizens' engagement initiatives" (European Commission, 2021b, p. 3).

Socioeconomically, the Commission begins from the assumption that AI modifies the lives of each and every European citizen. One of the policy initiatives from the earlier policy documents was the preparation for socioeconomic changes (European Commission, 2018b). Although, its emphasis lessened in the two recent policy documents, the white paper (European Commission, 2020a) and the communication paper on the proposal (European Commission, 2021b), socioeconomic considerations have remained important. It is also possible that socioeconomic changes related to AI were moved to be handled elsewhere as the policy developed and trustworthiness and competitiveness through regulation were placed as its cornerstones. It is also difficult to prepare for a situation that correlates to technological development that is somewhat unpredictable. Bottom line is that the Commission acknowledges that AI is the future, and the EU needs to start preparing for the changes it will bring as best as it can. For some this may mean adaptation in work life (European Commission, 2018c), but for everyone it means coming to terms with an emerging technology and its uptake in society. To guide the uptake of AI, the Commission has included societal wellbeing as an item in the ethics guidelines stating that:

> "Furthermore, the impact of AI systems should be considered not only from an individual perspective, but also from the perspective of society as a whole. The use of AI systems should be given careful consideration particularly in situations relating to the democratic process, including opinion-formation, political decision-making or electoral contexts. Moreover, AI's social impact should be considered. While AI systems can be used to enhance social skills, they can equally contribute to their deterioration." (European Commission, 2019, p. 6)

The Commission acknowledges that the transformation to a more digital Europe requires vast social adaptation to the changes AI technologies bring as well building citizens' trust in AI. Even so, the policy documents prove that there is a strong will to seize AI to fullest. Transformations in the labor market may affect competitiveness through employment negatively but also positively. The Commission appears to embrace the socioeconomic pros of AI as overriding the cons, willing to do the work to incorporate new technology to the European ways even when this may mean that some people may be temporarily unemployed (European Commission, 2018b). By presenting evidence of all the ways that AI can be of

service in the future in the policy, for example helping diagnose cancer (European Commission, 2018c), the Commission is verifying AI's societal importance and value.

### 6.1.7. Ecological

The ecological perspective considers AI from a sustainable angle. The Commission seems to agree that the development and deployment of AI should begin from the assumption that it will not burden the environment more. Possible challenges to consider from this perspective have to do with "the environmental impact of AI systems" and the need for AI to be "duly considered throughout their lifecycle and across the entire supply chain, e.g. as regards resource usage for the training of algorithms and the storage of data" (European Commission, 2020a, p. 2).

Opportunities from this perspective are more abundant and offer hope. AI is presented to help make sustainable choices, find solutions, and fight against climate change. Its potential is in "contributing to climate change mitigation and adaptation" (European Commission, 2020a, p. 1) and helping decrease pollution (European Commission, 2021b). It can also "reduce energy consumption" (European Commission, 2019, p. 1), "enable farmers to use water and natural resources more efficiently" (European Commission, 2019, p. 1), "predict environmental and climate change" (European Commission, 2019, p. 1), and "contribute to the European Green Deal objectives, helping industry, companies, public authorities and citizens make more sustainable choices" (European Commission, 2021b, p. 3).

Although the ecological aspect has been included as one part of analysis, it needs to be highlighted that the Commission's emphasis on environment and sustainability is quite weak in the research data. Nevertheless, having included the environmental perspective in the policy suggests that the Commission does consider the environment when it comes to the development and deployment of AI. This shows that the Commission acknowledges one of the biggest threats of our time, the climate crisis. The environmental aspect has also been included as an item in the ethics guidelines drafted by the HLEG (European Commission, 2019). The Commission has stated that:

> "For AI to be trustworthy, its impact on the environment and other sentient beings should be taken into account. Ideally, all humans, including future generations, should benefit from biodiversity and a habitable environment. Sustainability and ecological responsibility of AI systems should hence be encouraged. The same applies to AI

solutions addressing areas of global concern, such as for instance the UN Sustainable Development Goals." (European Commission, 2019, p. 6).

The inclusion of both the environment and the societal wellbeing in the ethics guidelines indicates promise in the Commission's trustworthy policy objective.

### 6.1.8. Education

In matters related to education, the Commission turns its head towards the areas that demand improvement for the successful integration of AI in Europe. Attention is directed towards life-long learning and improvement of citizens' digital skills (European Commission, 2018b; European Commission, 2018c). Attaining basic digital skills while investing in areas that AI cannot replace such as critical thinking, creativity, and management is encouraged (European Commission, 2018b). Underneath these policy actions, the Commission may be implying that AI will not fully flourish unless all EU citizens are on broad with the changes it will bring.

A gender balance is raised as a matter that demands attention in education in the field of AI (European Commission, 2018b) insinuating that the number of minority groups studying and working in the technical field needs to increase. The Commission (2020a) suggests that "particular efforts should be undertaken to increase the number of women trained and employed in this area" (p. 6) in addition to people from diverse backgrounds (European Commission, 2018b). To achieve "non-discriminatory and inclusive" AI "more women and people of diverse backgrounds, including people with disabilities, need to be involved in the development of AI, starting from inclusive AI education and training" (European Commission, 2018b). Although the idea behind achieving non-discriminatory and inclusive AI by suggesting that groups of people that are underrepresented in the field of AI simply need to be involved is novel, one could criticize the Commission's simplified approach to a problem that may rather be larger.

Additionally, the Commission reminds that "ethics in the development and use of new technologies should also be featured in programmes and courses" (European Commission, 2018b). The demand for ethics to be incorporated to the AI programs and courses hints that AI has yet to reach the sufficient level of consideration that prepares students to view AI from various perspectives. Overall, the education aspect is connected to the socioeconomic wellbeing of the European citizens in preparing them for the changes that AI brings to the

labor force and society as well as EU's competitiveness in educating AI professionals equipped with necessary skills.

To meet the challenges of AI and materialize the opportunities of AI in education, the renewed Digital Education Action Plan takes AI into account (European Commission, 2020i). In the white paper, the Commission (2020a) mentions how the updated Digital Education Action Plan "will help make better use of data and AI-based technologies such as learning and predictive analytics with the aim to improve education and training systems and make them fit for the digital age" (p. 6). Increasing awareness of AI so that citizens are equipped with the necessary information on dealing with AI is also one of the action points in the plan (European Commission, 2020a). The Commission also sees an opportunity in AI providing "enhanced opportunities for persons with disabilities and older persons to better education" (European Commission, 2021b, p. 1).

## 6.2. AI Problem Representations Categorized

In the preceding sections I have answered to my second research question "How has AI been represented as a policy problem by the European Commission?" I have carefully reviewed the selected policy documents in the hopes of identifying all problem representations of AI to analyze the presumptions and motives behind them. Next, I have analyzed each category (safety and security, ethical, legal, competitiveness, AI leadership, socioeconomic, ecological, and education) in turn to answer how the problem has been represented in policy to better make sense of how it may have contributed the representation of AI. For demonstration, I have summarized all eight categories in a table (Table 4).

Table 4: *AI Problem Representations*

|  | **Challenges** | **Opportunities** |
|---|---|---|
| **Safety and security** | Intrusion of privacy<br><br>Material risks (weaponization, technical errors, AI's ability to learn) | Safer transport<br><br>Crime prevention and cybersecurity<br><br>Anticipation of natural disasters |

| | Exploitation of human vulnerabilities to distort behavior | Preparation for future shocks |
|---|---|---|
| **Ethical** | Disrespect of EU fundamental rights<br><br>Bias and discrimination<br><br>Lack of transparency and accountability | Strong EU fundamental values and rights to build on |
| **Legal** | Difficulty of regulating AI due to the nature of the technology<br><br>Lack of EU wide legislation<br><br>Divergent national legal AI initiatives | Strong existing regulatory framework to build on |
| **Competitiveness** | Investment deficiency<br><br>Fragmented single market<br><br>Brain-drain<br><br>R&D<br>SMEs<br><br>Lack of data and access to data | Potential of public and industrial data<br><br>Sectoral advantages<br><br>RRF<br><br>Blue card |
| **AI leadership** | - | Global leadership in AI through regulation and ethics |
| **Socioeconomic** | Transformations in labor market | Ease the life of workers<br><br>Health |

| | General distrust in AI | Engagement in democratic processes |
|---|---|---|
| **Ecological** | Environmental impact of AI | Sustainability<br><br>Climate change mitigation |
| **Education** | Lack of digital skills<br><br>Gender imbalance | AI-based technologies in education |

The European Commission strives to dissolve the safety and security and ethical challenges related to AI by establishing AI policy that has a regulatory mechanism. These include issues such as privacy, technical errors, distortion, accountability, discrimination, and so on. On the other hand, there are opportunities that AI can bring to the table by enhancing safety and security conditions. Nonetheless, the rapidly advancing technological nature of AI will simply not be overlooked inside the EU anymore as regulatory actions have been underlined. The ethical guidelines that bring a human-centric perspective to AI have been planned to effortlessly be applied on top of the fundamental values that already exist in the EU. The ethical guidelines also play a role in establishing the foundation for AI regulation. From the legal perspective that is built on ethical AI, the Commission aims to strengthen existing legislations with the proposed AI regulation, although, regulating AI does not come without its challenges as the governed matter is more fluid rather than static. The difficulty of defining AI, never mind regulating it, demonstrates this. With the legal act, the Commission also wants to harmonize the EU member states in development and deployment of AI to form a unified front. Divergent national AI initiatives are seen as a problem that should not exist. Positively, as with the ethics, the Commission does not have to start completely afresh with the implementation of AI regulation as it can build its regulation on top of the already existing frameworks.

Improving the EU's competitiveness in AI is held in high regard in the research data. Deliberation is especially directed towards investments and the functions of the European internal market, but attractiveness of the European AI industry, uptake of AI in Europe, and data access are also considered. Closely related to the functionality of the internal single

market is the legal aspect that would establish a harmonized marketplace. Having a fragmented single market may harm the industry's standardization, legal certainty, and trust never mind cause a weakened state of competitiveness. Thus, the Commission pushes for regulation with the aim of avoiding the fragmentation of the internal market. Attention is also directed to making the EU an attractive center for technological innovation. Additionally, the Commission plans on supporting, for example, small and medium sized enterprises (SMEs) on incorporating AI to their businesses to enhance the uptake of AI in the EU. Some of the competitive opportunities that the Commission recognizes include data potentialities and advantages of AI in various sectors. From acquiring a strong internal position in the AI industry, the Commission strives to also become a global leader in AI by becoming a messenger of human-centric, ethical AI and setting and spreading regulatory practices from the EU to the world. By acquiring this, the Commission aims to reenforce the EU's position in the global AI race whole strengthening its level of competitiveness and attractiveness.

Socioeconomic emphasis has lessened as the policy has evolved. Still, the Commission has a strong will to conform to the uptake of AI as AI is believed to bring great societal improvements such as relieving workload through automation or advancing health. The transforming labor market may also impact competitiveness through employment. To help the progress of a society adapting to AI, the Commission pays attention to finding ways to ease AI's strain on labor market and build trust in AI. The ecological perspective does not play a major part in the policy documents, but the environmental impact of AI and the capabilities of AI in the environmental crisis are expressed. The development and deployment of AI should not burden the environment more. Finally, the education perspective contemplates on how to successfully integrate AI to society through education and training and fix the vulnerabilities in AI education. The interest in education may imply that Europe as it stands is not fully equipped to welcome the benefits of AI. Awareness and education in AI are needed.

Overall, it could be summed up that in its policy on AI the European Commission has over eight different categories argued for the uptake and regulation of the AI technology. It can be concluded that competitiveness is highly regarded along with safety and security, ethical, legal, and AI leadership matters. Less value is placed on socioeconomic, ecological, and education issues throughout the data documents, although each still serves its purpose. It is noteworthy to mention that although the categories are separately displayed here, in reality the lines are more blurred and the categories may overlap. By rationalizing for utilization and

regulation of AI in its policy through the different problem representation, the Commission has come to contribute to how AI is thought about (Bacchi, 2009). It will remain to be seen how powerful the European approach to AI will be in how it shapes the governance of AI.

# 7. CONCLUSION

This thesis has examined the progression of the European Commission's AI policy and analyzed the representations of AI found in the European approach to AI. I have approached this topic from the theoretical and methodological framework offered by Carol Bacchi (2009), the "What's the Problem Represented to be?" approach. The aim of this thesis has been to critically assess the problem representations of AI found in the European Commission's AI policy and make sense of what the Commission is aiming for with its policy processes in the larger scheme. Understanding the preceding stages to policy proposals and the underlying motives to why something is governed in a specific way is a valuable information that can help understand the policy decisions made. This thesis has attempted to illuminate the motives behind AI policy. What kinds of representation of AI challenges and opportunities are involved and how do these shape policy? Only by identifying these problematizations, can we fairly review the policy in question. Before proceeding to analysis of the problem representations, the thesis looks into how the European Commission's AI policy has developed.

My first research question, "How has the European Commission's policy on AI come about?", has been answered in chapter five of the thesis. This research question has shed light on the progression of the policy. I presented the formation of the European approach to AI by going through significant policy documents involved in the process over the period of four years, 2018-2021. I also demonstrated how the Commission's AI policy falls into place in the larger scheme of digital politics in the EU that aim for European technological sovereignty. The policy on AI can be seen as one piece of the EU's quest for technological sovereignty. The number of definitions established by the EU for AI over the years indicates the difficulty of governing such technology.

The second research question, "How has AI been represented as a policy problem by the European Commission?", has been answered in chapter six. I have analyzed the implicit representations of AI in policy from the European Commission's representation of AI challenges and opportunities. I have argued that the Commission approaches AI from eight different perspectives including safety and security, ethical, legal, competitiveness, AI leadership, socioeconomic, ecological, and education. The first five perspectives have seemed to weigh heavier in the policy as opposed to the three latter, although each has seemed to serve its purpose. I have demonstrated that the Commission sees that current and emerging

AI technologies should be approached with considerations to safety and security and ethics. To support safety and security and ethical considerations, the Commission seems to stress a functioning legal framework for regulation. Ethical guidelines could be considered as the building blocks for the operability of the regulatory framework whereas the legal framework ensures the safety and security of AI. Ethics, although having received some critique academically (Floridi, 2021b), hold a place in the European approach to AI.

The legal aspect is expressed to have a strong relation to AI competitiveness through the harmonization of the European Union and setting the preconditions for smooth trade within the internal market. In fact, both internal and external aspects for the introduction of a regulatory framework seem to be in valued interest in the Commission's AI policy. The EU seeks a strong regulatory position internally for purposes within the EU but also on global front for purposes outside of the EU. The global leadership position would support European AI competitiveness and give the EU an advantageous stand on spreading the message of human-centric AI and the values and fundamental rights that the EU is built on. Competitively, the Commission also wants to invest on AI through private and national investments, attractiveness, uptake of AI, and data access.

Although the Commission seems to emphasize fostering trustworthiness and competitiveness through regulation in its policy, there are also other perspectives that remain significant. Socioeconomically, transformations in the labor market and the public's trust in the technology should be addressed. Lack of attention to the socioeconomic transitions that AI brings may result in a weakened state of competitiveness through employment or refusal of the technology if there is no trust in it. On the other hand, useful advantages for the whole society are acknowledged in AI. Ecologically, the development and deployment of AI should take place in a way that will in no way put more strain on the environment but may help find more sustainable solutions to help mitigate the climate crisis. Education perspective has influence on building society's trust in AI through educating the society on AI as well as improving the EU's competencies in the AI industry through training.

This thesis has sought to provide insight on the European Commission's AI policy through examining the way that AI has been represented as a problem in the policy. It has demonstrated the AI policy problems that the European Commission has emphasized. By presenting AI through these policy problems, the European Commission is contributing to how AI is conceived. The strength of this thesis lies in the timeliness of the topic. The

European Commission's proposal for regulating AI that came out in April 2021 offered a good opening to take a closer look at how AI has been represented in the policy as the EU stands out as a pioneer in AI regulation. AI technology will inevitably only generalize, and with its AI policy the European Commission is now establishing the role that the EU will take in this. That being said, this thesis is by far not the only study interested in the topic (e.g. Justo-Hanani, 2022) and the subject will most likely attract interest increasingly.

Every study has its limitations and this one is no exception. This thesis focused on assessing a total of five policy documents that consisted of four communication papers and one white paper, leaving out, for example, the coordinated plan on AI (European Commission 2018d; European Commission 2021c) or the regulatory proposal (European Commission, 2021a). Although, the selection of the research data was reasonably justified, this thesis could have gained depth from expanding the research data and including other relevant papers in the analysis. This may have resulted in a more comprehensive and telling analysis. Bacchi's framework proved to be a fitting tool for critical inspection of a subject that is surrounded by both potential and risks, but I cannot help but wonder if the analysis could have provided more had all the WPR framework's questions been applied. Of course, this would have also meant that the selection of the data used in this study would have had to been expanded. With analyzing the silences and effects produced by the European Commission's representation of AI or how and where the representations of AI had been produced, disseminated, and defended (Bacchi, 2009), I could have dug deeper to the representations carried out by the policy.

Although the volume of the research data was moderate and ensured comprehensive review of the data, it is possible that there may be other impediments to the scope of rigor in the analysis. With qualitative research methods, there is the risk that I have missed out on some noteworthy points no matter how reasonable the size of the data and systematic the data assessment has been, especially with such a dimensional topic like AI. Misinterpretations may also have taken place while implementing the chosen methods to the data. It is possible that the questions offered by Bacchi have not been applied to their full potential due to subconscious bias. One researcher's interpretation of the data may differ from another one's no matter how rigorously one follows the qualitative methods.

Lastly, this thesis takes the European Commission's regulatory proposal into account when it is necessary but does not directly examine it. Closer assessment of the European

Commission's regulation framework, for example against governance AI models (Gasser & Almeida, 2017; Rahwan, 2018; Wirtz et al., 2020), would provide an interesting area for further research. The EU's capabilities of setting norms with the AI regulation could also be measured after the regulation comes to force. In fact, an analysis of the overall effectiveness of the European Commission's AI policy would in time provide an intriguing subject for study. To take advantage of the results of this particular study, the representations of policy problems outlined in this thesis could be compared to the representation of AI in the policy of a member state or a non-EU state. Overall, I believe that the usefulness of this thesis is in providing a categorization of possible representations of problems to look out for in policies on AI.

# 8. LIST OF REFERENCES

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, *35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Bacchi. (1999). *Women, Policy and Politics: the construction of policy problems*. London: SAGE.

Bacchi. (2009). *Analysing Policy: What's the problem represented to be?* Frenchs Forest, N.S.W: Pearson.

Bacchi, C. (2010a). Foucault, Policy and Rule: Challenging the Problem-Solving Paradigm. Institut for Historie, Internationale Studier og Samfundsforhold, Aalborg Universitet. FREIA's tekstserie No. 74 https://doi.org/10.5278/freia.33190049

Bacchi, C. (2010b). Poststructuralism, discourse and problematization: Implications for gender mainstreaming. *Kvinder, køn & forskning*, (4).

Bacchi, C. (2012). Introducing the "What's the Problem Represented to be?" Approach. In A. Bletsas & C.  Beasley (Eds.), *Engaging with Carol Bacchi: Strategic interventions and exchanges*, (p. 21- 24). The University of Adelaide Press.

Bauer, M., & Erixon, F. (2020). *Europe's quest for technology sovereignty: Opportunities and pitfalls*. European Centre for International Political Economy.

Bhatnagar, S., Alexandrova, A., Avin, S., Cave, S., Cheke, L., Crosby, M., ... & Hernández-Orallo, J. (2017, November). Mapping intelligence: Requirements and possibilities. In *3rd Conference on Philosophy and Theory of Artificial Intelligence* (p. 117-135). Springer, Cham.

Bradford. (2020). *The Brussels Effect: How the European Union Rules the World*. New York, NY: Oxford University Press.

Carriço, G. (2018). The EU and artificial intelligence: A human-centred perspective. *European View, 17*(1), 29–36. https://doi.org/10.1177/1781685818764821

Castro, D., & McLaughlin, M. (2021). Who is winning the AI race: China, the EU, or the United States? 2021 update. Center for Data Innovation. Retrieved March 16, 2022 from https://www2.datainnovation.org/2021-china-eu-us-ai.pdf

Cui, D., & Wu, F. (2021). The influence of media use on public perceptions of artificial intelligence in China: Evidence from an online survey. *Information Development, 37*(1), 45–57. https://doi.org/10.1177/0266666919893411

Darnis, J. P. (2020). A COVID-19 moment for technological sovereignty in Europe? *ARIES*, *20*, 29. https://www.iai.it/sites/default/files/iaicom2029.pdf

Drisko, J. & Maschi, T. (2015). *Content analysis*. New York: Oxford University Press.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice, and policy. *International Journal of Information Management 57*, 2021, 101994.

European Commission. (2017). Joint Declaration on the EU's legislative priorities for 2018-19. Retrieved March 2, 2022 from https://ec.europa.eu/info/sites/default/files/joint-declaration-eu-legislative-priorities-2018-19_en.pdf

European Commission. (2018a). Declaration of Cooperation on Artificial Intelligence. Retrieved January 24, 2020 from https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence

European Commission. (2018b). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Empty: Artificial Intelligence for Europe. Retrieved January 25, 2022 from https://digital-strategy.ec.europa.eu/en/library/communication-artificial-intelligence-europe

European Commission. (2018c). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Coordinated Plan on Artificial Intelligence. Retrieved January 25, 2022 from https://knowledge4policy.ec.europa.eu/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en

European Commission. (2018d). Annexes to the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Coordinated Plan on Artificial

Intelligence. Retrieved February 9, 2022 from
https://knowledge4policy.ec.europa.eu/publication/coordinated-plan-artificial-
intelligence-com2018-795-final_en

European Commission. (2019). Communication from the Commission to the European
Parliament, the European Council, the European Economic and Social Committee and
the Committee of the Regions: Building Trust in Human-Centric Artificial
Intelligence. Retrieved January 25, 2022 from
https://ec.europa.eu/jrc/communities/sites/default/files/ec_ai_ethics_communication_
8_april_2019.pdf

European Commission. (2020a). White Paper on Artificial Intelligence – A European
approach to excellence and trust. Retrieved January 25, 2022 from
https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-
approach-excellence-and-trust_en

European Commission. (2020b). Report from the Commission to the European Parliament,
the Council and the European Economic and Social Committee: Report on the safety
and liability implications of Artificial Intelligence, the Internet of Things and robotics.
Retrieved January 25, 2022 from https://ec.europa.eu/info/sites/default/files/report-
safety-liability-artificial-intelligence-feb2020_en.pdf

European Commission. (2020c). Communication from the Commission to the European
Parliament, the Council, the European Economic and Social Committee and the
Committee of the Regions: A European strategy for data. Retrieved February 9, 2022
from https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-
age/european-data-strategy_en

European Commission. (2020d). Proposal for a Proposal for a Regulation of the European
Parliament and of the Council on European data governance (Data Governance Act).
Retrieved February 14, 2022 from https://eur-lex.europa.eu/legal-
content/EN/TXT/?uri=CELEX%3A52020PC0767

European Commission. (2020e). Proposal for a Regulation of the European Parliament and of
the Council on a Single Market for Digital Services (Digital Services Act) and
amending Directive 2000/31/EC. Retrieved February 14, 2022 from https://digital-

    strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-
    council-single-market-digital-services-digital-services

European Commission. (2020f). Proposal for a Proposal for a Regulation of the European
    Parliament and of the Council on contestable and fair markets in the digital sector
    (Digital Markets Act). Retrieved February 14, 2022 from
    https://ec.europa.eu/info/sites/default/files/proposal-regulation-single-market-digital-
    services-digital-services-act_en.pdf

European Commission. (2020g). Joint Communication to the European Parliament and the
    Council: The EU's Cybersecurity Strategy for the Digital Decade. Retrieved February
    14, 2022 from https://digital-strategy.ec.europa.eu/en/library/eus-cybersecurity-
    strategy-digital-decade-0

European Commission. (2020h). Public consultation on the AI White Paper. Retrieved March
    2, 2022 from https://digital-strategy.ec.europa.eu/en/white-paper-artificial-
    intelligence-public-consultation-towards-european-approach-excellence-and

European Commission. (2020i). Communication from the Commission to the European
    Parliament, the Council, the European Economic and Social Committee and the
    Committee of the Regions: Digital Education Action Plan 2021-2027. Retrieved
    March 15, 2022 from https://eur-lex.europa.eu/legal-
    content/EN/TXT/?uri=CELEX%3A52020DC0624

European Commission. (2021a). Proposal for a Regulation of the European Parliament and
    the Council: Laying Down Harmonised Rules on Artificial Intelligence (Artificial
    Intelligence Act) and Amending Certain Union Legislative Acts. Retrieved from
    https://eur-lex.europa.eu/legal-
    content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

European Commission. (2021b). Communication from the Commission to the European
    Parliament, the Council, the European Economic and Social Committee and the
    Committee of the Regions: Fostering a European approach to Artificial Intelligence.
    Retrieved January 25, 2022 from https://digital-
    strategy.ec.europa.eu/en/library/communication-fostering-european-approach-
    artificial-intelligence

European Commission. (2021c). Annexes to the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Fostering a European approach to Artificial Intelligence: Fostering a European approach to Artificial Intelligence, Coordinated Plan on Artificial Intelligence 2021 Review. Retrieved January 25, 2022 from https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review

European Commission. (2021d). Impact Assessment of the Regulation on Artificial Intelligence. Retrieved February 9, 2022 from https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-intelligence

Fast, E., & Horvitz, E. (2017, February). Long-term trends in the public perception of artificial intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31, No. 1).

Fischer, Torgerson, D., Durnová, A., & Orsini, M. (2015). *Handbook of critical policy studies*. Cheltenham, United Kingdom: Edward Elgar Publishing.

Floridi, L. (2020). The fight for digital sovereignty: What it is, and why it matters, especially for the EU. *Philosophy & Technology*, *33*(3), 369-378.

Floridi, L. (2021a). The European Legislation on AI: a Brief Analysis of its Philosophical Approach. *Philosophy & Technology*, 1-8.

Floridi L. (2021b) Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. In L. Floridi (Ed.) *Ethics, Governance, and Policies in Artificial Intelligence*. Philosophical Studies Series, vol 144. Springer, Cham. https://doi.org/10.1007/978-3-030-81907-1_6

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689-707.

Floridi, L., & Cowls, J. (2021). A unified framework of five principles for AI in society. In *Ethics, Governance, and Policies in Artificial Intelligence* (p. 5-17). Springer, Cham.

Gasser, U., & Almeida, V. A. (2017). A layered model for AI governance. *IEEE Internet Computing*, *21*(6), 58-62.

Regulation 2016/679. Protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). European Parliament, Council of the European Union. Retrieved February 7, 2022 from https://eur-lex.europa.eu/eli/reg/2016/679/oj

Regulation 2018/1807. Framework for the free flow of non-personal data in the European Union. European Parliament, Council of the European Union. Retrieved February 7, 2022 from https://eur-lex.europa.eu/eli/reg/2018/1807/oj

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99-120.

Hidvegi, F., Leufer, D., & Massé, E. (February 17, 2021). Accessnow. Blogpost. Retrieved October 4, 2021 from https://www.accessnow.org/eu-regulation-ai-risk-based-approach/

High-Level Expert Group on Artificial Intelligence (AI HLEG). (2019a). A Definition of AI: Main Capabilities and Disciplines. Shaping Europe's digital future, European Commission. Retrieved February 9, 2022 from https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

High-Level Expert Group on Artificial Intelligence (AI HLEG). (2019b). Ethics Guidelines for Trustworthy AI. Shaping Europe's digital future, European Commission. Retrieved February 9, 2022 from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

High-Level Expert Group on Artificial Intelligence (AI HLEG). (2019c). Policy and Investment Recommendations for Trustworthy AI. Shaping Europe's digital future, European Commission. Retrieved February 9, 2022 from https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence

High-Level Expert Group on Artificial Intelligence (AI HLEG). (2020a). Assessment List for Trustworthy AI (ALTAI). Shaping Europe's digital future, European Commission. Retrieved February 9, 2022 from https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

High-Level Expert Group on Artificial Intelligence (AI HLEG). (2020b). Sectoral
Considerations on Policy and Investment Recommendations for Trustworthy AI.
European AI Alliance, European Commission. Retrieved February 9, 2022 from
https://futurium.ec.europa.eu/en/european-ai-alliance/document/ai-hleg-sectoral-
considerations-policy-and-investment-recommendations-trustworthy-ai

Jobin, Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature
Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Justo-Hanani. (2022). The politics of Artificial Intelligence regulation and governance reform
in the European Union. *Policy Sciences*, *55*(1), 137–159.
https://doi.org/10.1007/s11077-022-09452-8

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On
the interpretations, illustrations, and implications of artificial intelligence. *Business
Horizons, 62*(1), 15-25.

Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and
opportunities of artificial intelligence. *Business Horizons, 63*(1), 37-50.

Koulu, R. (2020). Human control over automation: EU Policy and AI Ethics. *European
Journal of Legal Studies, 12*(1), 9-46.

Larsson, S. (2020). On the governance of artificial intelligence through ethics guidelines.
*Asian Journal of Law and Society, 7*(3), 437-451.

Legg, S., & Hutter, M. (2007). A collection of definitions of intelligence. In B. Goertzel, & P.
Wang (Eds.), *Advances in Artificial General Intelligence: Concepts, Architectures
and Algorithms. Proceedings of the AGI Workshop 2006* (p. 17–24). IOS Press.

Mannes, A. (2020). Governance, Risk, and Artificial Intelligence. *The AI Magazine, 41*(1),
61–69. https://doi.org/10.1609/aimag.v41i1.5200

McMillan, D., & Brown, B. (2019). Against ethical AI guidelines and self interest. *ACM
International Conference Proceeding Series*, 19–21.
https://doi.org/10.1145/3363384.3363393

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine
Intelligence, 1* (11), 501–507

Monett, D., & Lewis, C. W. (2017, November). Getting clarity by defining artificial intelligence—a survey. In *3rd Conference on Philosophy and Theory of Artificial Intelligence* (p. 212-214). Springer, Cham.

Neri, H., & Cozman, F. (2020). The role of experts in the public perception of risk of artificial intelligence. *AI & Society, 35*(3), 663–673. https://doi.org/10.1007/s00146-019-00924-9

Oliveira, E. (2017). Beneficial AI: the next battlefield. *Journal of Innovation Management, 5*(4), 6-17.

Padden, M., & Öjehag-Pettersson, A. (2021). Protected how? Problem representations of risk in the General Data Protection Regulation (GDPR). *Critical Policy Studies*, *15*(4), 486-503.

Probst, L., Pedersen, B., Lefebvre, V., & Dakkak-Arnoux, L. (2018). Digital Transformation Monitor: USA-China-EU plans for AI: where do we stand? European Commission. Retrieved March 16, 2022 from https://ati.ec.europa.eu/sites/default/files/2020-07/USA-China-EU%20plans%20for%20AI%20-%20where%20do%20we%20stand%20%28v5%29.pdf

Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology, 20*(1), 5-14.

Resseguier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society, 7*(2), 2053951720942541.

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., & Teller, A. (2016). *Artificial intelligence and life in 2030* (Study Panel Report 2015-2016). https://ai100.stanford.edu/2016-report

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, 1-21.

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W. G. (2020). Framing governance for a contested emerging technology: insights from AI policy. *Policy and Society, 40*(2), 158-177.

Ulnicane, I., Eke, D. O., Knight, W., Ogoh, G., & Stahl, B. C. (2021). Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies. *Interdisciplinary Science Reviews, 46*(1-2), 71-93.

von der Leyen, U. (2019a). *Speech in the European Parliament Plenary Session*. European Commission. Retrieved January 24, 2022 from https://ec.europa.eu/info/sites/default/files/president-elect-speech-original_en.pdf

von der Leyen, U. (2019b). A Union that Strives for More. My agenda for Europe, Political Guidelines for the Next European Commission 2019–2024. European Commission. Retrieved January 24, 2022 from https://ec.europa.eu/info/sites/default/files/political-guidelines-next-commission_en.pdf

von der Leyen, U. (2019c). Mission letter – Thierry Breton, Commissioner for Internal Market. European Commission. Retrieved January 24, 2022 from https://ec.europa.eu/commission/commissioners/sites/default/files/commissioner_mission_letters/president-elect_von_der_leyens_mission_letter_to_thierry_breton.pdf

Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence, 10*(2), 1-37.

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., ... & Schwartz, O. (2018). *AI now report 2018* (p. 1-62). New York: AI Now Institute at New York University.

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: towards a focus on tensions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (p. 195-200).

Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *International Journal of Public Administration*, *43*(9), 818-829.